MPEG-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue?

Raphaël Troncy[1], Òscar Celma[2], Suzanne Little[3] and Roberto García[4], and Chrisa Tsinaraki[5]

[1] CWI Amsterdam, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
raphael.troncy@cwi.nl
[2] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
[3] Institute of Computer Vision and Applied Computer Sciences, Leipzig, Germany
[4] Universitat de Lleida, Lleida, Spain
[5] Technical University of Crete (TUC/MUSIC), Crete, Greece

Abstract. MPEG-7 can be used to create complex and comprehensive metadata descriptions of multimedia content. Since MPEG-7 is defined in terms of an XML schema, the semantics of its elements have no formal grounding. In addition, certain features can be described in multiple ways. In order to make MPEG-7 interoperable with domain-specific ontologies, the semantics of the MPEG-7 descriptors also need to be expressed formally in an ontology. This article describes four independent approaches to build a multimedia ontology based on the MPEG-7 standard and discusses the similarities and differences between them.

1 Introduction

MPEG-7, formally named *Multimedia Content Description Interface* [15], is an ISO/IEC standard developed by the Moving Picture Experts Group (MPEG) for the structural and semantic description of multimedia content. MPEG-7 standardizes *tools* or ways to define multimedia *Descriptors* (Ds), *Description Schemes* (DSs) and the relationships between them. The descriptors correspond either to the data features themselves, generally low-level features such as visual (e.g. texture, camera motion) and audio (e.g. spectrum, harmonicity), or semantic objects (e.g. places, actors, events, objects). Ideally, most low-level descriptors would be extracted automatically, whereas human annotation would be required for producing high-level descriptors. The description schemes are used for grouping the descriptors into more abstract description entities. These tools as well as their relationships are represented using the *Description Definition Language* (DDL), the core part of MPEG-7. After a requirement specification phase, the W3C XML Schema recommendation[6] has been adopted as the most appropriate syntax for the MPEG-7 DDL.

[6] http://www.w3.org/XML/Schema

The flexibility of MPEG-7 is therefore based on allowing descriptions to be associated with arbitrary multimedia segments, at any level of granularity, using different levels of abstraction. The downside of the breadth targeted by MPEG-7 is its complexity and its ambiguity. Hence, MPEG-7 XML Schemas define 1182 elements, 417 attributes and 377 complex types which make the standard difficult to manage. Moreover, the use of XML Schema implies that a great part of the semantics remains implicit. For example, very different syntactic variations may be used in multimedia descriptions with the same intended semantics, while remaining valid MPEG-7 descriptions. Given that the standard does not provide a formal semantics for these descriptions, this syntax variability causes serious interoperability issues for multimedia processing and exchange [16, 17, 22]. The profiles introduced by MPEG-7 and their possible formalization [21] concern, by definition, only a subset of the whole standard.

For alleviating the lack of formal semantics in MPEG-7, four multimedia ontologies represented in OWL and covering the whole standard have been proposed [1, 4, 9, 24]. In this paper, the proposers of these four ontologies compare and discuss these four modeling approaches. The four MPEG-7 based ontologies are presented in section 2. We use then a common example, the semantic description of regions of a still image, in order to compare the resulting RDF descriptions (section 3). We discuss the main differences of these ontologies using three criteria: *i)* the way the multimedia ontology is linked with domain semantics, *ii)* the MPEG-7 coverage of the multimedia ontology, and *iii)* the scalability and modeling rationale of the conceptualization (section 4). We present other related multimedia ontologies in section 5 and give a conclusion in section 6.

2 MPEG-7 based Multimedia Ontologies

From 2001 until the present time, four main ontologies that formalize the MPEG-7 standard using Semantic Web languages have been proposed. In the following, we describe these four ontologies, and the main characteristics as well as the context in which they have been developed are summarized in the Table 1.

2.1 Hunter's MPEG-7 ontology

In 2001, Hunter proposed an initial manual translation of MPEG-7 into RDFS (and then into DAML+OIL) and provided a rationale for its use within the Semantic Web [9]. This multimedia ontology was translated into OWL, extended and harmonized using the ABC upper ontology [14] for applications in the digital libraries [10, 11] and eResearch fields [12].

The current version is an OWL Full ontology containing classes defining the media types (Audio, AudioVisual, Image, Multimedia, Video) and the decompositions from the MPEG-7 Multimedia Description Schemes (MDS) part [15]. The descriptors for recording information about the production and creation, usage,

	Hunter	**DS-MIRF**	**Rhizomik**	**COMM**
Foundations	ABC	none	none	DOLCE
Complexity	OWL-Full[a]	OWL-DL[b]	OWL-DL[c]	OWL-DL[d]
Coverage	MDS+Visual	MDS+CS	All	MDS+Visual
Reference	[9]	[24]	[4]	[1]
Applications	Digital Libraries, e-Research	Digital Libraries, e-Learning	Digital Rights Management, e-Business	Multimedia Analysis and Annotations

Table 1. Summary of the different MPEG-7 based Multimedia Ontologies.

[a] `http://metadata.net/mpeg7/`
[b] `http://www.music.tuc.gr/ontologies/MPEG703.zip`
[c] `http://rhizomik.net/ontologies/mpeg7ontos`
[d] `http://multimedia.semanticweb.org/COMM/`

structure and the media features are also defined. The ontology can be viewed in Protégé[7] and has been validated using the WonderWeb OWL Validator[8].

This ontology has usually been applied to describe the decomposition of images and their visual descriptors for use in larger semantic frameworks. Harmonizing through an upper ontology, such as ABC, enables queries for abstract concepts such as subclasses of *events* or *agents* to return media objects or segments of media objects. While the ontology has most often been applied in conjunction with the ABC upper model, it is independent of that ontology and can also be harmonized with other upper ontologies such as SUMO [18] or DOLCE [3].

2.2 DS-MIRF ontology

In 2004, Tsinaraki *et al.* have proposed the DS-MIRF ontology that fully captures in OWL DL the semantics of the MPEG-7 MDS and the Classification Schemes. The ontology can be visualized with GraphOnto or Protege and has been validated and classified with the WonderWeb OWL Validator. The ontology has been integrated with OWL domain ontologies for soccer and Formula 1 [25] in order to demonstrate how domain knowledge can be systematically integrated in the general-purpose constructs of MPEG-7. This ontological infrastructure has been utilized in several applications, including audiovisual digital libraries and e-learning.

The DS-MIRF ontology has been conceptualized manually, according to the methodology outlined in [24]. The XML Schema simple datatypes defined in MPEG-7 are stored in a separate XML Schema to be imported in the DS-MIRF ontology. The naming of the XML elements are generally kept in the `rdf:ID`s of the corresponding OWL entities, except when two different XML Schema constructs have the same names. The mapping between the original names of the

[7] `http://protege.stanford.edu/`
[8] `http://www.mygrid.org.uk/OWL/Validator`

MPEG-7 descriptors and the `rdf:ID`s of the corresponding OWL entities is represented in an OWL DL mapping ontology. Therefore, this ontology will represent, for example, that the `Name` element of the MPEG-7 type `TermUseType` is represented by the `TermName` object property, while the `Name` element of the MPEG-7 type `PlaceType` is represented by the `Name` object property in the DS-MIRF ontology. The mapping ontology also captures the semantics of the XML Schemas that cannot be mapped to OWL constructs such as the sequence element order or the default values of the attributes. Hence, it is possible to return to an original MPEG-7 description from the RDF metadata using this mapping ontology. This process has been partially implemented in GraphOnto [19], for the OWL entities that represent the `SemanticBaseType` and its descendants.

The generalization of this approach has led to the development of a transformation model for capturing the semantics of any XML Schema in an OWL DL ontology [23]. The original XML Schema is converted into a main OWL DL ontology while a OWL DL mapping ontology keeps trace of the constructs mapped in order to allow circular conversions.

2.3 Rhizomik ontology

In 2005, Garcia and Celma have presented the Rhizomik approach that consists in mapping XML Schema constructs to OWL constructs following a generic XML Schema to OWL together with an XML to RDF conversion [4]. Applied to the MPEG-7 schemas, the resulting ontology covers the whole standard as well as the Classification Schemes and TV Anytime[9]. It can be visualized with Protege or Swoop[10] and has been validated and classified using the Wonderweb OWL Validator and Pellet.

The Rhizomik ontology was originally expressed in OWL Full, since 23 properties must be modeled using an `rdf:Property` because they have both a data type and object type range, i.e. the corresponding elements are both defined as containers of complex types and simple types. An OWL DL version of the ontology has been produced, solving this problem by creating two different properties (`owl:DatatypeProperty` and `owl:ObjectProperty`) for each of them. This change is also incorporated into the XML2RDF step in order to map the affected input XML elements to the appropriate OWL property (object or datatype) depending on the kind of content of the input XML element.

The main contribution of this approach is that it benefits from the great amount of metadata that has been already produced by the XML community. Moreover, it is implemented in the ReDeFer project[11], which allows to automatically map input XML Schemas to OWL ontologies and, XML data based on them to RDF metadata following the resulting ontologies. This approach has been used with other large XML Schemas in the Digital Rights Management domain, such as MPEG-21 and ODRL [6], or in the E-Business domain [5].

[9] `http://www.tv-anytime.org`
[10] `http://code.google.com/p/swoop`
[11] http://rhizomik.net/redefer

2.4 COMM ontology

In 2007, Arndt *et al.* have proposed COMM, the *Core Ontology of MultiMedia* for annotation. Based on early work [20, 13], COMM has been designed manually by re-engineering completely MPEG-7 according to the intended semantics of the written standard. The foundational ontology DOLCE serves as the basis of COMM. More precisely, the Description and Situation (D&S) and Ontology of Information Objects (OIO) patterns are extended into various multimedia patterns that formalize the MPEG-7 concepts. The use of a upper-level ontology 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.

COMM covers the most important part of MPEG-7 that is commonly used for describing the structure and the content of multimedia documents. Current investigations show 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 other multimedia patterns.

COMM is an OWL DL ontology that can be viewed using Protege. Its consistency has been validated using Fact++-v1.1.5. Other reasoners failed to classify it due to the enormous amount of DL axioms that are present in DOLCE. The presented OWL DL version of the core module is just an approximation of the intended semantics of COMM since the use of OWL 1.1 (e.g. qualified cardinality restrictions for number restrictions of MPEG-7 low-level descriptors) and even more expressive logic formalisms are required for capturing its complete semantics[12].

3 Comparing the MPEG-7 Ontologies

To compare the four MPEG-7 based ontologies described above, we used a task to annotate the famous "Big Three" picture, taken at the Yalta (Crimea) Conference, showing the heads of government of the United States, the United Kingdom, and the Soviet Union during World War II (figure 1).

The description could be obtained either manually or automatically from an annotation tool. It could also be the result of an automatic conversion from an MPEG-7 description. The annotation should contain the media identification and locator, define the still region `SR1` of the image, and provide the semantics of the region using `http://en.wikipedia.org/wiki/Churchill` for identifying the resource `Winston Churchill`. We provide below the RDF descriptions generated by each ontology using the N3 syntax.

[12] The reification schema of DOLCE D&S is even not completely expressible in OWL 1.1

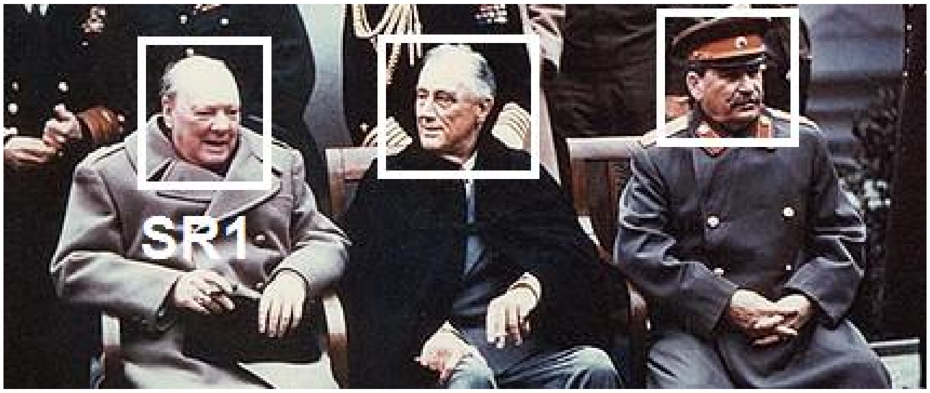

Fig. 1. The “Big Three” at the Yalta Conference (Image adapted from Wikipedia), `http://en.wikipedia.org/wiki/Yalta_Conference`

3.1 Hunter’s MPEG-7 Ontology

Figure 2 shows an RDF description created manually and compliant with Hunter’s MPEG-7 ontology. It focuses mainly on the decomposition of the image into `StillRegion`s and the annotation of the segment with some simple semantics. A basic `VisualDescriptor` (`DominantColor`) is also included to show how the output from analysis tools can be included. These statements could also be generated via an (semi-)automatic segmentation and analysis algorithm or through an interactive user annotation interface. In this instance, domain semantics are attached through the use of the `depicts` relation. Alternatively, an upper ontology could be used to create triples such as `mpeg7:Media mpeg7:depicts abc:Event` and hence to a domain specific ontology describing, for example, political history.

```
@prefix mpeg7: <http://metadata.net/mpeg7/mpeg7.owl> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

:image mpeg7:MediaLocator <http://en.wikipedia.org/wiki/Image:Yalta_Conference.jpg> .
:image rdf:type mpeg7:image .
:image mpeg7:depicts "The Big Three at the Yalta Conference" .
:image mpeg7:spatial_decomposition :SR1 .
:SR1 a mpeg7:StillRegion .
:SR1 mpeg7:SpatialMask :mask01 .
:mask01 a mpeg7:Polygon .
:mask01 mpeg7:dim (2 5) .
:mask01 mpeg7:Coords (5 25 10 20 15 15 10 10 5 15) .
:SR1 mpeg7:DominantColor "rgb(255,255,225)" .
:SR1 mpeg7:depicts "Churchill" .
```

Fig. 2. RDF description of Figure 1 compliant with Hunter’s MPEG-7 ontology

3.2 DS-MIRF ontology

Figure 3 shows an RDF description created manually and compliant with the DS-MIRF ontology. The description consists in the image title and the description of the spatial decomposition. The region `SR1` showing Winston Churchill is described in details using the region bounding box coordinates, the region title and some material relevant with the content.

```
@prefix MDS: <http://127.0.0.1:8080/ontologies/MPEG703/MDS#>
@prefix Visual: <http://127.0.0.1:8080/ontologies/MPEG703/Visual#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

MDS:Mpeg7 rdf:ID "Yalta" MDS:Description [MDS:ContentEntityType;
  MDS:MultimediaContent [MDS:ImageType; rdf:about "#IMG1"] ].

MDS:ImageType rdf:ID "IMG1"; MDS:Image [MDS:StillRegionType;
  MDS:MediaLocator [MDS:MediaLocatorType; MDS:MediaUri
  "http://en.wikipedia.org/wiki/Image:Yalta_Conference.jpg"]
  MDS:CreationInformation [MDS:CreationInformationType;
    MDS:Creation [MDS:CreationType; MDS:Title [MDS:TitleType;
      contentString "The Big Three at the Yalta Conference"] ] ]
    MDS:SpatialDecomposition [MDS:StillRegionSpatialDecompositionType;
      MDS:StillRegion [MDS:StillRegionType; rdf:about "#SR1"] ] ].

MDS:StillRegion [MDS:StillRegionType; rdf:ID "SR1"
  MDS:CreationInformation [MDS:CreationInformationType;
    MDS:Creation [MDS:CreationType;
      MDS:Title [MDS:TitleType; contentString "Winston Churchill"] ]
    MDS:RelatedMaterial [MDS:RelatedMaterialType;
      MDS:MediaLocator [MDS:MediaLocatorType;
        MDS:MediaUri "http://en.wikipedia.org/wiki/Churchill"] ] ]
  MDS:SpatialMask [MDS:SpatialMaskType;
    MDS:SubRegion [Visual:RegionLocatorType;
      Visual:Polygon[Visual:PolygonType;
        Visual:Coords[MDS:IntegerMatrixType; MDS:dim "2 5"
          MDS:contentIntegerVector "5 25 10 20 15 15 10 10 5 15"] ] ] ] ].
```

Fig. 3. RDF description of Figure 1 compliant with the DS-MIRF ontology

3.3 Rhizomik Ontology

Figure 4 shows an RDF description generated automatically from an input MPEG-7 XML description and compliant with the Rhizomik ontology. The first step in the XML to RDF mapping process is to model the XML elements tree using the corresponding OWL object and datatype properties from the Rhizomik ontology, i.e. those properties produced in the XSD to OWL mapping for the corresponding elements. The result is a representation for the XML tree based on OWL properties. Then, the resulting RDF graph is enriched with type relations to the OWL Classes that correspond to the XML complex types involved.

Figure 5 shows an HTML rendering of this RDF description. This rendering is based on the ReDeFer RDF2HTML tool[13] which produces a more user-friendly view of the RDF metadata and allows the user to interact with it.

[13] `http://rhizomik.net/redefer`

```
@prefix : <http://rhizomik.net/ontologies/2005/03/Mpeg7-2001.owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<urn:example#Yalta> :Description [ a :ContentEntityType;
  :MultimediaContent [ a :ImageType;
    :Image <urn:example#IMG1> ] ] .

<urn:example#IMG1> a :StillRegionType;
  :CreationInformation [ :Creation  [
    :Title "The Big Three at the Yalta Conference" ] ];
  :MediaLocator [
    :MediaUri "http://en.wikipedia.org/wiki/Image:Yalta_Conference.jpg" ];
  :SpatialDecomposition  [
    :StillRegion <urn:example#SR1> ] .

<urn:example#SR1> a :SegmentType;
  :Semantic <http://en.wikipedia.org/wiki/Churchill>;
  :TextAnnotation  [ :KeywordAnnotation  [
    :Keyword "Churchill", "Winston" ] ] ;
  :SpatialMask [ :SubRegion [ :Polygon [ :Coords  [
    :dim "2 5";
      rdf:value "5 25 10 20 15 15 10 10 5 15" ] ] ] ] .

<http://en.wikipedia.org/wiki/Churchill>
 :Label  [
   :Name "Winston Churchill" ].
```

Fig. 4. RDF description of Figure 1 compliant with the Rhizomik ontology

3.4 COMM Ontology

Figure 6 shows an RDF description created manually and compliant with the COMM ontology. The description starts with the traditional housekeeping and the declaration of the various namespaces. COMM has been designed in various modules for representing the *visual* and *audio* parts or the *datatypes* used. The descriptors used for localizing parts of media assets are formalized in the *localization* module. Finally, the COMM multimedia patterns use extensively the properties defined in the DOLCE D&S ontology.

The decomposition of Figure 1 into the still region `SR1` (the bounding box of Churchill's face) is represented by an `image-data` instance which plays a `still-region-role`. It is located by a `digital-data` instance which expresses the `region-locator-descriptor` defined by a `bounding-box`. Due to the semantic annotation pattern, one can annotate the still region by connecting it with the URI `http://en.wikipedia.org/wiki/Winston_Churchill`, an instance of a `foaf:Person`.

4 Discussion

The four RDF descriptions detailed above differ substantially, even though they aim to represent the same semantics and are all based on MPEG-7. In this section, we discuss the main differences of the underlying ontologies using three criteria: *i)* the way the multimedia ontology is linked with domain semantics, *ii)* the MPEG-7 coverage of the multimedia ontology, and *iii)* the scalability and modeling rationale of the conceptualization.

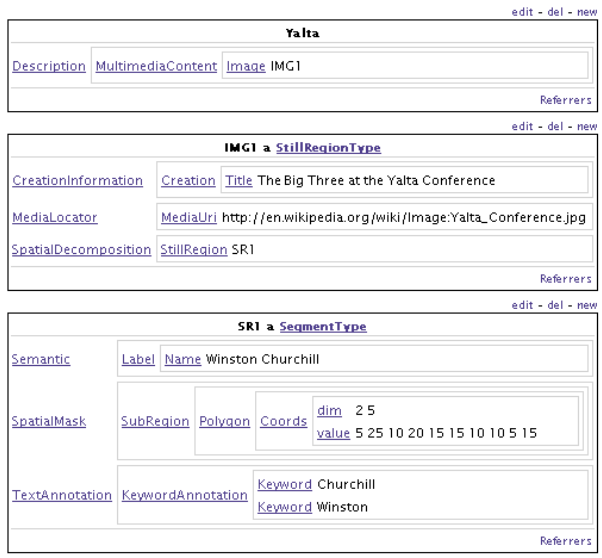

Fig. 5. HTML view automatically generated by the ReDeFer tool

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dns: <http://multimedia.semanticweb.org/COMM/extended-dns-very-lite.owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix core: <http://multimedia.semanticweb.org/COMM/core.owl#> .
@prefix visual: <http://multimedia.semanticweb.org/COMM/visual.owl#> .
@prefix loc: <http://multimedia.semanticweb.org/COMM/localization.owl#> .
@prefix data: <http://multimedia.semanticweb.org/COMM/datatype.owl#> .

<#SR1> a core:image-data;
  dns:realized-by [
    http://en.wikipedia.org/wiki/Image:Yalta_Conference.jpg a core:media ];
  dns:plays [
    visual:still-region-role [
      dns:requires [
        loc:spatial-mask-role [
          dns:played-by [
            loc:region-locator-descriptor [
              dns:defines [
                loc:bounding-box [
                  data:has-rectangle "5 25 10 20 15 15 10 10 5 15"^^xsd:string ] ] ] ] ] ] ] ];
  dns:setting [
    core:semantic-annotation [
      dns:satisfies [
        dns:method [
          dns:defines [
            core:semantic-label-role [
              dns:played-by http://en.wikipedia.org/wiki/Churchill ] ];
          dns:defines [
            core:annotated-data-role [
              dns:played-by <#SR1> ] ] ] ];
      dns:setting-for [
        http://en.wikipedia.org/wiki/Churchill a foaf:Person;
          foaf:name "Winston Churchill" ] ] ].
```

Fig. 6. RDF description of Figure 1 compliant with the COMM ontology

4.1 Integration with domain semantics

The link between a multimedia ontology and any domain ontologies is crucial. In our example, we observe that a more complete description could include information about "Churchill" (a person, a British Prime Minister, etc.) and about the event. In addition, details about the provenance of the image (e.g. date taken, photographer, camera used) could also be linked to complete the description. The statements contained in the descriptions above, in conjunction with any of the four underlying ontologies presented in this paper, can then be used to answer queries such as "*find all images depicting Churchill*" or "*find all media depicting British Prime Ministers*". Furthermore, subjective queries such as "*find images with a 'bright' segment in them*", where 'bright' is defined as `mpeg7:DominantColor` greater than rgb(220,220,220), are also possible.

Hunter's MPEG-7 and COMM ontologies both use an upper ontology approach to relate with other ontologies (ABC and DOLCE). Hunter's ontology uses either semantic relations from MPEG-7, such as `depicts`, or defines external properties that use an MPEG-7 class, such as `mpeg7:Multimedia`, as the domain or range. In COMM, the link with existing vocabularies is made within a specific pattern: the *Semantic Annotation Pattern*, reifing the DOLCE Ontology of Information Object (OIO) pattern. Consequently, any domain specific ontology goes under the `dolce:Particular` or `owl:Thing` class.

The DS-MIRF ontology integrates domain knowledge by sub-classing one of the MPEG-7 `SemanticBaseType`: places, events, agents, etc. Furthermore, it fully captures the semantics of the various MPEG-7 relationships represented as instances of the `RelationType`. According to the standard, the value of the these properties must come from some particular classification schemes: `RelationBaseCS`, `TemporalRelationCS`, `SpatialRelationCS`, `GraphRelationCS` and `SemanticRelationCS`. A typed relationship ontology extending DS-MIRF has been defined for capturing all these relationships.

4.2 Coverage of a multimedia ontology

The four multimedia ontologies discussed here cover partially or totally MPEG-7 (see Table 1). They also extend sometimes the standard. For example, Hunter's MPEG-7 ontology has been extended for the description of scientific mixed-media data. Common terms used in signal processing and image analysis for describing detailed low-level features such as eccentricity, major axis length, lightest color, etc. are lacking in the MPEG-7 visual descriptors. These extra visual feature descriptors have been introduced as sub-properties of the the visual descriptor and color properties, using the namespace `mpeg7x` to keep these extensions independent of the core MPEG-7 descriptors [7].

The modeling approach of COMM 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 (e.g. manual annotation, output of an analysis algorithm), 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[14].

4.3 Modeling decisions and scalability

An important modeling decision for each of the four ontologies is how much they are tied to the MPEG-7 XML Schema. These decisions impact upon the ability of the ontology to support descriptions generated automatically and directly from MPEG-7 XML output and on the complexity of the resulting RDF. Therefore the modeling choices also affect the scalability of the systems using these ontologies and their ability to handle large media data sets and cope with reasoning over very large quantities of triples.

Both the DS-MIRF and the Rhizomik ontologies are based on a systematic one-to-one mapping from the MPEG-7 descriptors to equivalent OWL entities. For the DS-MIRF ontology, the mapping has been carried out manually while for the Rhizomik ontology, it has been automated using an XSL transformation and it is complemented with an XML to RDF mapping. This has been a key motivator for the Rhizomik ontology and the ReDeFer tool where the objective is to provide an intermediate step before going to a more complete multimedia ontology, such as COMM.

The advantage of the one-to-one mapping is that the transformation of the RDF descriptions back to MPEG-7 descriptions may be automated later on. In addition, this approach enables the exploitation of legacy data and allows existing tools that output MPEG-7 descriptions to be integrated into a semantic framework. The main drawback of this approach is that it does not guarantee that the intended semantics of MPEG-7 is fully captured and formalized. On the contrary, the syntactic interoperability and conceptual ambiguity problems such as the various ways of expressing a semantic annotation remain.

The COMM ontology avoids doing a one-to-one mapping for solving these ambiguities that come from the XML Schemas, while an MPEG-7-to-COMM converter is still available for re-using legacy metadata. A direct translation from an MPEG-7 XML description using Hunter's ontology is possible. However, in practice, the multimedia semantics captured by the ontology have instead been used to link with domain semantics. Therefore rather than translating MPEG-7 XML descriptions into RDF, this ontology has been used to define semantic statements about a media object and to relate these statements to the domain semantics. This results in a smaller number of triples (see Table 2).

	Hunter	**DS-MIRF**	**Rhizomik**	**COMM**
Number of RDF triples	11	20	27	19

Table 2. Raw number of RDF triples (without inference) of each image description.

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

5 Related work

The MPEG-7 based ontologies discussed here aim to provide richer semantics and better frameworks for multimedia description and exchange than can be addressed by these standards. Related efforts to develop multimedia ontologies include the following. The Visual Descriptor Ontology[15] (VDO) [2] is based on the MPEG-7 Visual part and used for image and video analysis. Hollink *et al.* have proposed a visual ontology by extending Wordnet with multimedia semantics from Hunter's ontology, specifically for use within the museums and art domain [8]. Vembu *et al.* have developed an MPEG-7 based ontology and applied it to annotating football (soccer) videos [26]. Similar to the approach used in Hunter's ontology and in COMM, this ontology uses the decomposition and visual components of MPEG-7 and captures high-level domain semantics in domain specific ontologies.

6 Conclusion

In this paper, we have described the four ontologies based on the MPEG-7 standard developed for the purpose of formally expressing multimedia semantics and facilitating interoperability with domain ontologies. We used an image annotation scenario and we produced RDF descriptions for demonstrating some of the differences and similarities between these ontologies and how they integrate domain semantics.

While developing each of these ontologies, different modeling decisions have been made and different approaches to integration were chosen. These choices have been based upon the particular implementation and intended use of the ontologies. While the resulting ontologies are different, they still share the core multimedia semantics, since they are based upon the MPEG-7 standard, and can therefore be seen as complementary solutions to the central issue. Descriptions built using these ontologies offer many advantages over MPEG-7 data structured using only the MPEG-7 XML Schema.

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. Suzanne Little's work has been partially conducted during the tenure of a MUSCLE Internal Fellowship under contract FP6-507752. The DS-MIRF ontology was partially funded in the scope of the DELOS II Network of Excellence in Digital Libraries under contract FP6-026059. The Rhizomik ontology has been partially funded by the Spanish research project S5T (Scalable Semantic personalized Search of Spoken and written contents on the Semantic Web) TIN2005-06885.

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

References

1. Richard Arndt, Raphaël Troncy, Steffen Staab, Lynda Hardman, and Miroslav Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *6th International Semantic Web Conference (ISWC)*, 2007.
2. Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis Avrithis, Siegfried Handschuh, Yiannis Kompatsiaris, Steffen Staab, and Michael G. Strintzis. Semantic Annotation of Images and Videos for Multimedia Analysis. In *2nd European Semantic Web Conference (ESWC)*, pages 592–607, 2005.
3. Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening Ontologies with DOLCE. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, pages 166–181, 2002.
4. Roberto Garcia and Oscar Celma. Semantic Integration and Retrieval of Multimedia Metadata. In *5th International Workshop on Knowledge Markup and Semantic Annotation*, pages 69–80, 2005.
5. Roberto Garcia and Rosa Gil. Facilitating Business Interoperability from the Semantic Web. In *10th International Conference on Business Information Systems (BIS)*, pages 220–232, 2007.
6. Roberto Garcia, Rosa Gil, and Jaime Delgado. A Web Ontologies Framework for Digital Rights Management. *Journal of Artificial Intelligence and Law*, 15:137–154, 2007.
7. Laura Hollink, Suzanne Little, and Jane Hunter. Evaluating the Application of Semantic Inferencing Rules to Image Annotation. In *3rd International Conference on Knowledge Capture (K-CAP)*, 2005.
8. Laura Hollink, Marcel Worring, and A.Th. Schreiber. Building a Visual Ontology for Video Retrieval. In *13th ACM International Conference on Multimedia*, 2005.
9. Jane Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *1st International Semantic Web Working Symposium (ISWC)*, pages 261–281, 2001.
10. Jane Hunter. Combining the CIDOC/CRM and MPEG-7 to Describe Multimedia in Museums. In *Museums on the Web (MW)*, 2002.
11. Jane Hunter. Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):49–58, 2003.
12. Jane Hunter and Suzanne Little. A Framework to Enable the Semantic Inferencing and Querying of Multimedia Content. *International Journal of Web Engineering and Technology – Special Issue on the Semantic Web*, 2(2/3):264–286, 2005.
13. Antoine Isaac and Raphaël Troncy. Designing and Using an Audio-Visual Description Core Ontology. In *Workshop on Core Ontologies in Ontology Engineering*, 2004.
14. Carl Lagoze and Jane Hunter. The ABC Ontology and Model (v3.0). *Journal of Digital Information*, 2(2), 2001.
15. MPEG-7. Multimedia Content Description Interface. Standard No. ISO/IEC 15938, 2001.
16. Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005.

17. Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4), 2004.
18. Adam Pease, Ian Niles, and John Li. The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, 2002.
19. Panagiotis Polydoros, Chrisa Tsinaraki, and Stavros Christodoulakis. GraphOnto: OWL-based ontology management and multimedia annotation in the DS-MIRF framework. *Journal of Digital Information Management (JDIM)*, 4(4):214–219, 2006.
20. Raphaël Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *2nd International Semantic Web Conference (ISWC'03)*, pages 566–581, Sanibel Island, Florida, USA, 2003.
21. Raphaël Troncy, Werner Bailer, Michael Hausenblas, Philip Hofmair, and Rudolf Schlatte. Enabling Multimedia Metadata Interoperability by Defining Formal Semantics of MPEG-7 Profiles. In *1st International Conference on Semantics And digital Media Technology (SAMT'06)*, pages 41–55, Athens, Greece, 2006.
22. Raphaël Troncy and Jean Carrive. A Reduced Yet Extensible Audio-Visual Description Language: How to Escape From the MPEG-7 Bottleneck. In *4th ACM Symposium on Document Engineering (DocEng'04)*, Milwaukee, Wisconsin, USA, 2004.
23. Chrisa Tsinaraki and Stavros Christodoulakis. Interoperability of XML Schema Applications with OWL Domain Knowledge and Semantic Web Tools. In *6rd International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*, 2007.
24. Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In *3rd International Conference on Image and Video Retrieval (CIVR)*, pages 582–591, 2004.
25. Chrisa Tsinaraki, Panagiotis Polydoros, and Stavros Christodoulakis. Interoperability support between MPEG-7/21 and OWL in DS-MIRF. *In Transactions on Knowledge and Data Engineering (TKDE), Special Issue on the Semantic Web Era*, 19(2):219–232, 2007.
26. Shankar Vembu, Malte Kiesel, Michael Sintek, and Stephan Bauman. Towards Bridging the Semantic Gap in Multimedia Annotation and Retrieval. In *1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, 2006.