Multimed Tools Appl (2010) 46:331–370
DOI 10.1007/s11042-009-0387-4

Enquiring MPEG-7 based multimedia ontologies

Stamatia Dasiopoulou · Vassilis Tzouvaras · Ioannis Kompatsiaris · Michael G. Strintzis

Published online: 16 October 2009

Abstract Machine understandable metadata forms the main prerequisite for the intelligent services envisaged in a Web, which going beyond mere data exchange and provides for effective content access, sharing and reuse. MPEG-7, despite providing a comprehensive set of tools for the standardised description of audiovisual content, is largely compromised by the use of XML that leaves the largest part of the intended semantics implicit. Aspiring to formalise MPEG-7 descriptions and enhance multimedia metadata interoperability, a number of multimedia ontologies have been proposed. Though sharing a common vision, the developed ontologies are characterised by substantial conceptual differences, reflected both in the modelling of MPEG-7 description tools as well as in the linking with domain ontologies. Delving into the principles underlying their engineering, we present a systematic survey of the state of the art MPEG-7 based multimedia ontologies, and highlight issues that hinder interoperability as well as possible directions towards their harmonisation.

Keywords Multimedia ontologies · MPEG-7 · Semantic Web · Multimedia metadata · Interoperability

S. Dasiopoulou (✉) · I. Kompatsiaris · M. G. Strintzis
Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
e-mail: dasiop@iti.gr

I. Kompatsiaris
e-mail: ikom@iti.gr

M. G. Strintzis
e-mail: strintzis@iti.gr

V. Tzouvaras
Department of Electrical and Computer Engineering, National Technical University of Athens, Zographou 15780, Athens, Greece
e-mail: tzouvaras@image.ntua.gr

1 Introduction

Multimedia content is ubiquitous on the Web; popular search engines such as Google and Yahoo images index billions of images, while community portals such as Flickr, Picassa and YouTube, to name but a few, proclaim the tremendous rates at which multimedia documents are produced and consumed. The richness and multiplicity of information communicated through such resources, in combination with the sheer volume involved, make the availability of interoperable content descriptions key for the realisation of practical applications involving content access, retrieval and reuse.

A number of diverse vocabularies have been proposed addressing the varying aspects such descriptions may consider, ranging from format and subject matter descriptions to authoring and privacy concerns [5]. However, providing intelligent content management presupposes more than mutual conformance to a common metadata vocabulary and exchange format: it requires the encoding of the respective semantics independently of the way it is processed, so that content management at a more semantic level can seamlessly take place. Under such context, machine understandable, rather than merely processable, metadata is both a prerequisite and a major challenge.

Towards this goal, the Semantic Web [3] brought forth a number of technologies for capturing, representing, and managing semantics by making it formal and explicit. Languages such as RDFS [6] and OWL [2] have been developed to formalise meaning and promote its sharing among heterogenous systems. Motivated by a kindred vision of information communication and reuse, yet targeting specifically audiovisual documents, ISO developed the *Multimedia Content Description Interface*, commonly referred to as MPEG-7 [24, 25]. MPEG-7 provides a comprehensive set of standardised tools for the description of audiovisual content at multiple granularities, addressing a variety of dimensions that range from structural and low-level descriptions to aspects related to navigation, content organisation, as well as user preferences and usage.

However, as thoroughly elaborated by Ossenbruggen et al. [29, 40], a number of practical obstacles have hindered not only the widespread use of either approach as the means for providing sharable multimedia metadata on the Web, but their synergistic utilisation too. Critical factors preventing the latter are the interoperability problems encountered at syntactic and semantic level. The use of XML Schema for the definition of the MPEG-7 description tools contrasts with the RDF based SW languages, while different standpoints are taken with respect to semantics definition.

Aspiring to reconcile and exploit the complementary assets provided by the two proposals, namely the formal semantics and reasoning capabilities of the SW approach and the multimedia specific description tools of MPEG-7, a number of initiatives have investigated the engineering of MPEG-7 based ontologies [1, 4, 8, 15, 19, 31, 38]. An immediate effect of such efforts is the alleviation of the syntactic barriers. However, confronting the interoperability issues at the semantic level constitutes a task of significantly greater challenge. An elementary cause relates to the lack of a standardised (semantic) correspondence between XML and RDF. Challenges of greater intricacy arise with respect to designating the *intended semantics*. On one hand the use of XML Schema, leaves the largest part of the semantics in the accompanying documentation rather than the description schemes themselves. On the other hand, the flexibility that MPEG-7 allows in the use of certain descriptions

tools has the twofold effect of multiple interpretations per description and variant descriptions of equivalent meaning.

Legitimately different ontology modelling decisions incur, while additional diversity is entailed by the intended context of usage and the envisaged multimedia metadata Web architecture. For example, ontologies aiming to support interoperability between existing MPEG-7 repositories and SW applications need to provide full coverage of the MPEG-7 features. Ontologies on the other hand that focus more on reasoning over media related knowledge, semantics negotiation and alignment, inevitably adhere to more rigorous ontological commitments in order to enforce precise meaning and often restrain the flexibility initially afforded by MPEG-7.

The aforementioned incur a rather obscure setting regarding the interoperability and correlations between the existing MPEG-7 based ontologies, establishing the need for a common framework of reference. Shedding further insight, such framework may allow not only for their effective utilisation but also for the identification of open challenges and future directions. Aiming to contribute towards this direction, in this article we present a systematic survey of the state of the art in MPEG-7 based multimedia ontologies. Key dimensions of this enquiry constitute the two main issues addressed by the existing ontologies, namely the representation of multimedia structural aspects (including decomposition and localisation schemes) and the linking with domain specific ontologies for the purpose of expressing subject matter descriptions. In addition to the thorough examination of the modelling choices taken by the different approaches, we highlight the influential role that the envisaged metadata interoperability architecture has on the latter choices. Finally, through the elucidation of the goals served by the individual ontologies and the differences between ontologies serving the same purpose, possible ways to their harmonisation are outlined.

The rest of the paper is organised as follows. Section 2 presents the state of the art MPEG-7 based multimedia ontologies. Section 3 discusses alternative architectures for a- chieving interoperable semantically-enabled multimedia metadata. Sections 4 and 5 detail the modelling choices with respect to the representation of structural descriptions and linking with domain-specific ontologies respectively. Section 6 summarises the observations, highlighting associations to the different interoperability architectures and possible solutions towards achieving their harmonisation. Related initiatives are presented in Section 7, and Section 8 concludes the paper.

2 MPEG-7 based multimedia ontologies

Although multimedia descriptions may refer to numerous aspects, we focus on the threefold view currently considered by the relevant literature, namely: i) subject matter descriptions expressing the semantics conveyed, ii) structural descriptions pertaining to the decomposition and localisation of content parts, and iii) low-level descriptors covering visual and audio features. Correspondingly, the relevant MPEG-7 parts are:

- part 3, Visual [26] that addresses visual features such as texture, colour, etc.
- part 4, Audio [27] that addresses audio features such harmonicity, specturm, etc., and

– part 5, MDS (Multimedia Description Schemes) [28], and in particular the parts that specify tools for structural (*clause 11*), semantic (*clause 12*), localisation (*clause 6*), and media descriptions (*clause 8*).

In the following, we present the state of the art MPEG-7 based multimedia ontologies, outlining main characteristics including MPEG-7 coverage, representation language, modularity and manual vs automatic engineering. Furthermore, a brief account of the applications where they have been utilised is given.

2.1 The Harmony MPEG-7 based ontology

The MPEG-7 ontology proposed by Hunter in 2001 within the Harmony[1] project constitutes chronologically the first initiative to attach formal semantics to MPEG-7 [19]. The RDF Schema (RDFS) language was proposed to formalise the structural and localisation tools of the MPEG-7 MDS, as well as the descriptors included in the Visual part. Furthermore, a set of descriptors representing information about production, creation, usage and media features were included. The developed ontology was ported later to DAML and eventually to OWL [20]. Further extensions, addressing specific image analysis terms used in the MATLAB Image Processing Toolbox, have been defined to facilitate integration with MATLAB based image analysis implementations [18]; however, being application tuned their interoperability is quite restricted. To address subject matter descriptions, external domain specific ontologies are assumed and linking is achieved through an upper ontology. In the deployed applications, the ABC [22] ontology has been used for this purpose.

The translation of the MPEG-7 definitions into an ontological representation adheres to the original MPEG-7 Schemas. The different MPEG-7 content and segment types are modelled as classes, and so are the visual descriptors, while properties have been used for the modelling the decomposition schemes. Preserving the flexibility afforded by MPEG-7, segment types are treated as multimedia content types too, as illustrated in Fig. 1. Furthermore, adopting a loose axiomatisation, the defined semantic entities are allowed to have more than one semantic interpretation (e.g. the StillRegion class may refer to a still region, or the entire still image, or to a video frame), while different entities may share the same meaning. A a result, the ambiguities present in MPEG-7 are propagated, incurring serious implications on the conceptual clarity and subsequent management of the produced descriptions, as detailed in Section 4.

Hunter's MPEG-7 ontology has been utilised for ontology-based semantic analysis and annotation of fuel cell [20] and pancreatic cell images [18, 23]. The underlying rationale has been to provide a uniform formal representation so that low-level features of image regions can be linked to the domain specific ontologies that describe fuel and pancreatic cells. Exploiting these associations, rule-based reasoning is subsequently applied so as to determine which domain entity is depicted given the low-level features extracted from the examined image regions. The ontology is available at http://metadata.net/mpeg7/mpeg7.owl.

[1] http://metadata.net/harmony/index.html

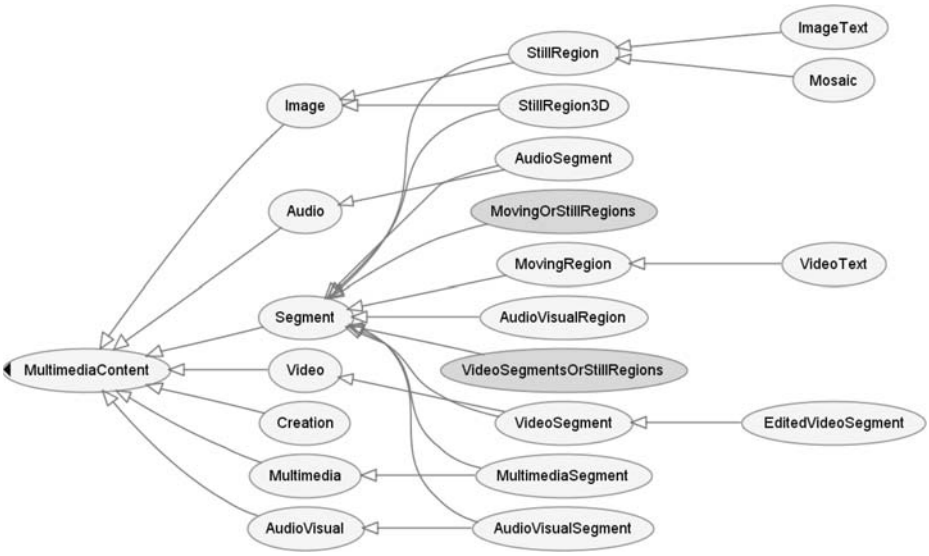


Fig. 1 Class hierarchy of the multimedia content and segment classes in the Harmony MPEG-7 based ontology

2.2 The aceMedia MPEG-7 based ontology

Two RDFS ontologies [4, 34], namely the Multimedia Structure Ontology (MSO) and the Visual Descriptor Ontology (VDO) have been developed within the aceMedia[2] project. MSO covers the complete set of structural description tools from the MDS, while VDO addresses the Visual part.

The modelling choices taken in the development of MSO and VDO follow the same engineering principles with Hunter's ontology resulting in an analogous class hierarchies. Figure 2 illustrates the MSO content and segment classes hierarchy. Thus, similar effects issue, namely preservation of flexibility in the descriptions at the cost of semantic ambiguity. However, from a conceptual perspective MSO enhances Hunter's modelling by introducing new classes (and properties) to capture explicitly some of the different notions implied by the multiple interpretations MPEG-7 attributes to a single description. For example, MSO introduces the mso:Frame and mso:KeyFrame classes (see Fig. 2) in order to model explicitly the frame interpretation that an MPEG-7 VideoSegment or StillRegion description may have. Another example is the definition of the mso:MovingRegion class, which in MSO appears as a direct subclass of mso:VideoSegment, capturing to an extend the temporal correlation between the two segment types, as opposed to Hunter's ontology, where the MovingRegion class it is defined as a direct class of the generic Segment class.

Furthermore, instead of following a monolithic engineering, as Hunter's, the aceMedia approach is modularised, addressing the structural and low-level features definitions in two separate ontologies. An upper ontology is assumed to provide

[2] http://www.acemedia.org

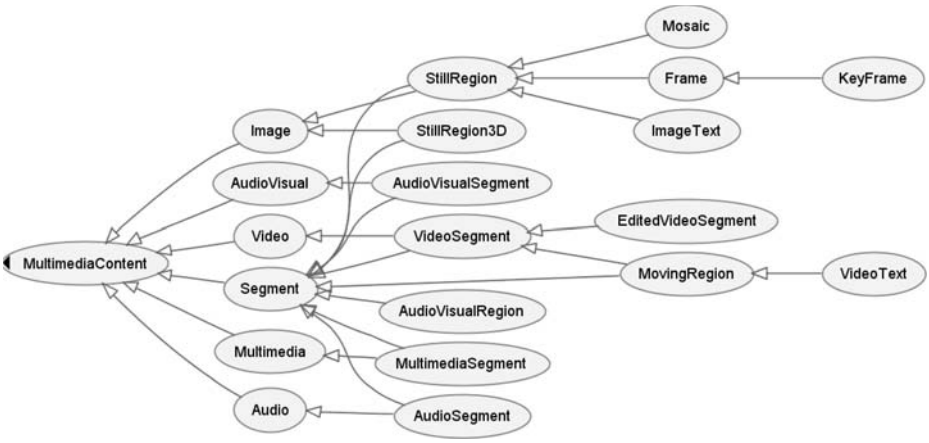


Fig. 2 Class hierarchy of the multimedia content and segment classes in the aceMedia MPEG-7 based ontology

the means for linking the MSO and VDO classes with domain specific ontologies. DOLCE and an ontology specifically developed for this purpose, namely the Annotation Ontology discussed in Section 5, have been used.

The MSO and VDO ontologies have been used for supporting semantic image and video analysis and annotation, addressing both personal and commercial content domains, including beach holidays, tennis games, etc. [10, 32]. In a similar vein to the one followed in [18, 20, 23], VDO provides the formalisation of low-level feature descriptions, while MSO models respective image and video parts. Through the use of a so called Annotation Ontology [32], instances of a domain ontology can be linked to the specific parts depicting them and to the low-level descriptors extracted for those parts. By that means, prototype instances are created that enable the subsequent identification of the domain entities depicted by previously unprocessed content. The MSO and VDO ontologies can be found at http://www.acemedia.org/aceMedia/results/ontologies.html.

2.3 The SmartWeb MPEG-7 based ontology

Based on SmartSUMO, a foundational ontology developed on the basis of DOLCE [14] and SUMO [30], a set of ontologies relevant for query-answering and information services on the Web, have been developed within the SmartWeb[3] project. Among them an MPEG-7 based ontology to support the annotation of multimedia content [31, 41].

The developed approach, realising a metamodelling ontological framework, allows to model formally the MPEG-7 descriptions and export them into OWL and RDFS, with corresponding expressivity comprises. The covered MPEG-7 descriptions include the structural, localisation, media and low-level description tools. Further description aspects, such creation and production, are covered indirectly,

[3] http://www.smartweb-projekt.de/start_en.html

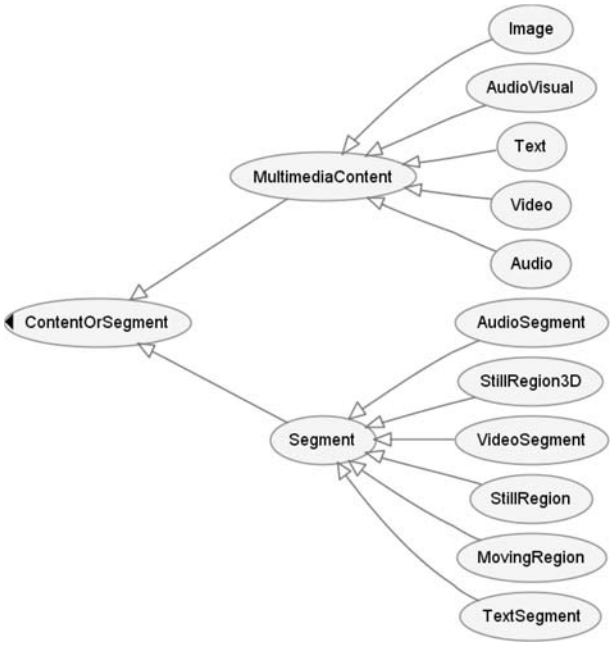


Fig. 3 Class hierarchy of the multimedia content and segment entities classes in the SmartWeb MPEG-7 based ontology

through media-independent classes and properties building on the two foundational ontologies. Linking with domain specific ontologies is accomplished by the SmartWeb Integrated Ontology (SWIntO) infrastructure that aligns the developed set of ontologies.

Contrary to the previously described ontologies, the SmartWeb MPEG-7 ontology does not treat Segment classes as specialisations of the MultimediaContent class, as illustrated in Fig. 3. Furthermore, adopting a different modelling perspective, the decomposition schemes are modelled also as classes, each denoting a valid decomposition pattern per content/segment type and a spatial/temporal dimension. As described in Section 4, where the engineering choices are detailed with respect to the representation of structural and localisation descriptions, this approach may be closer to the original MPEG-7 Schemas but introduces peculiarities and semantic ambiguities, especially in the case of recursive content decomposition.

The developed MPEG-7 based ontology has been utilised in the annotation of soccer videos providing support for addressing structural and low-level features descriptions. The ontology is available at http://smartweb.dfki.de/ontology_en.html.

2.4 The Boemie MPEG-7 based ontology

In an attempt to capture unequivocally the semantics of the MPEG-7 MDS structural descriptions, as well as the Visual and Audio Parts in a more declarative way, two OWL DL ontologies have been developed within the context of the BOEMIE[4]

[4] http://www.boemie.org/

project, namely the the Multimedia Content Ontology (MCO) and the Multimedia Descriptors Ontology (MDO) [8, 9, 11].

Key consideration for both ontologies has been the advocation of a clean conceptualisation to avail of reasoning capabilities. Instead of following a strict translation, MCO re-engineers the MPEG-7 structural and localisation descriptions in order to axiomatise the intended meaning. Not only distinct classes have been introduced to model the different content and segment entities, but disjoint axioms explicitly model that they form non intersecting sets. Figure 4 illustrates the corresponding multimedia content types and segment types hierarchy, where the ¬ symbols represent disjointness axioms defined between the MultimediaContent class (and its subclasses) and the various segment classes. Additionally, the valid decomposition and localisation patterns have been modelled as restrictions in the respective class definitions.

Furthermore, MCO introduces some additional aspects to the MPEG-7 structural description tools. Among them is the discrimination of decomposition into logical units and decomposition into multimedia segments that not necessarily share a logical coherence. For example, a video segment depicting a pole vault attempt may be decomposed into constituent temporal segments based on the different types of displayed motion activity, without necessarily corresponding to distinct semantic subevents. Another additional feature relates to the ability to specifically represent cases where content of one modality is rendered through another, as for example happens when the textual information displayed in an athlete's shirt is rendered as a still region of an image. Linking with domain specific ontologies is implemented through a pair of generic properties that capture the relation between a content/segment instance and the depicted semantics, and the relation between a content/segment instance and its extracted low-level features.

The context under which the two ontologies have been utilised is that of multimedia documents semantic analysis, annotation and retrieval [21]. MDO supports the

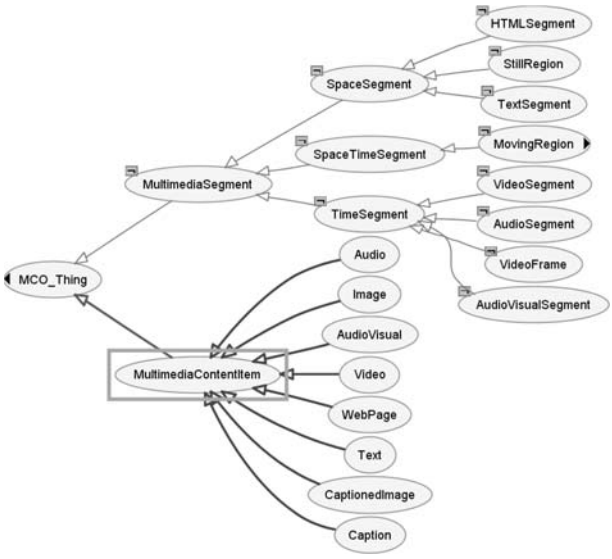


Fig. 4 Class hierarchy of the multimedia content and segment classes in Boemie's MCO ontology

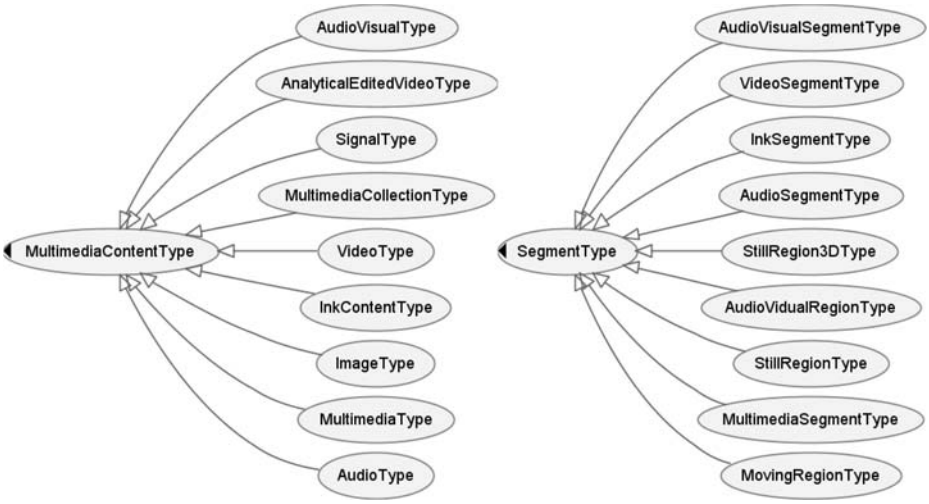


Fig. 5 Class hierarchies of the multimedia content and segment classes in the Rhizomik ontology

representation of automatically extracted low-level features, used for classification training purposes. MCO assists in the representation of the classification results in order to acquire higher level semantic interpretations through reasoning, and also allows for more effective retrieval by allowing access to the exact content parts that are of interest. The ontologies can be accessed at http://www.boemie.org/ontologies.

2.5 The Rhizomik MPEG-7 based ontology

Contrary to the aforementioned efforts that target partial translations of MPEG-7 in a manual fashion, the Rhizomik ontology, developed within the ReDeFer[5] project, proposes a fully automatic translation of the complete MPEG-7 Schema to OWL [15]. It is based on a generic XML Schema to Web Ontology Language mapping, called XSD2OWL, that is combined with a transparent mapping from XML to RDF, the XML2RDF. Applied to the MPEG-7 definitions results in an OWL DL ontology covering all elements of the entire MPEG-7 standard. Human intervention is required only to resolve name conflicts stemming from the independent name domains for complex types and elements in XML.

Figure 5 illustrates the class hierarchies resulting from the respective MPEG-7 content and segment type definition Schemas. Although no sublcass relations connect the two hierarchies as was the case for the Harmony and aceMedia ontologies, Rhizomik preserves the flexibility of the MPEG-7 specifications by explicitly capturing the multiple interpretations in the ontological definitions. Thus, it supports the interpretation of segment type classes as multimedia content classes, as well as all other cases of multiple interpretations per description and of multiple descriptions with same meaning. As a result, all ambiguities present in MPEG-7 are retained (Sections 4 and 5), much as the complexity and length characterising MPEG-7

[5] http://rhizomik.net/redefer

metadata. The latter becomes particularly evident when comparing with the rest MPEG-7 based ontologies, which instead of following a one to one translation, circumvent many elements of the MPEG-7 XML Schemas and target directly the intended meaning of the description tool at hand.

Unlike the MPEG-7 ontologies presented previously, linking with domain specific subject matter descriptions is intrinsic to the Rhizomik ontology as it covers the Semantic DS description tools as well. Specifically, all semantic descriptions are modelled as instances of the different types subclassing the semantic SemanticBaseType, namely the SemanticType, AgentObjectType, ObjectType, SemanticTimeType, SemanticPlaceType, SemanticStateType, ConceptType, and EventType classes. This abstraction model though defines a rather coarse conceptualisation when imposed to existing domain ontologies in the Web, as it requires the re-engineering of existing definitions so that they align under the restricted MPEG-7 models. Consequently, and as described in more detail in Section 5, Rhizomik may provide a very useful mechanism when it comes to making existing MPEG-7 metadata repositories visible to the rest of the Web, yet it results in major challenges when it comes to linking these metadata with existing domain ontologies as it requires for tedious mappings that cannot be easily automated.

Application examples of the resulting MPEG-7 ontology include the semantic integration and retrieval of music metadata, while the XML Schema to OWL translation has been additionally validated in the Digital Right Management and E-Business domains [16, 17].The Rhizomik ontology is available at http://rhizomik.net/ontologies/mpeg7ontos.

2.6 The DS-MIRF MPEG-7 based ontology

Within the DS-MIRF framework [37, 39], the complete MPEG-7 MDS plus the tools from the Visual and Audio parts that are required in MDS descriptions have been manually translated into an OWL DL ontology.

As in the case of the Rhizomik ontology, a one to one translation has been followed taking into account all elements appearing in the respective MPEG-7 description tools. Thus the two share similar class hierarchies and definitions, and the same semantic ambiguities. The DS-MIRF though enhances further the clarity of the translations semantics by making explicit the implicit notions of the initial XML Schemas, a trait that Rhizomik's automated transformation cannot take into account. Specifically, transformation from XML to OWL, and conversely, is supported through a separate OWL DL ontology that holds the mappings between the original XML Schema names and the corresponding OWL entities. The mapping ontology allows additionally the maintenance of XML Schema constructs that cannot be captured with OWL, such as default values and the sequence element.

Regarding the Semantic DS, a systematic methodology has been presented for the integration of domain specific semantics with the general-purpose semantic entities of MPEG-7 [38]. Unlike the Rhizomik ontology, the DS-MIRF framework provides a straightforward manner for translating MPEG-7 semantic descriptions to OWL/RDF metadata ones and also for enhancing them by linking them with domain specific ontologies. As elaborated in Section 5, it alleviates the need from cumbersome, and unavoidably imprecise, mappings from existing ontologies to the coarse semantic conceptualisation of MPEG-7.

The DS-MIRF ontological framework has been practically tested in the domain of soccer and Formula 1 [39], demonstrating support for MPEG-7 based semantic multimedia for both OWL and MPEG-7 repositories. The ontologies comprising the DS-MIRF framework are available at http://www.music.tuc.gr/ontologies/MPEG703.zip.

2.7 The COMM ontology

The Core Ontology for MultiMedia (COMM) [1] constitutes the most recent approach to the formalisation of MPEG-7 descriptions semantics. COMM is in OWL DL and covers selected tools from the structural, localisation and media description schemes of MDS, as well as low-level descriptors of the visual part. Additionally, it provides the means to include analysis aspects in the annotations, such as information about the algorithm and corresponding parameters used in the extraction of a given description.

In order to provide a common foundational framework for the description of multimedia documents, COMM extends the *Descriptions & Situations (D&S)* and *Ontology of Information Objects (OIO)* design patterns [13] of DOLCE, by re-engineering the MPEG-7 description tools. To support multimedia content descriptions along the threefold perspective of subject matter, structural and low-level feature descriptions, COMM defines four main patterns: i) the decomposition pattern that addresses structural and localisation descriptions, ii) the content annotation pattern that formalises the attachment of metadata to content/segment instances, iii) the media annotation pattern that addresses the representation of the physical instances of multimedia content items, and iv) the semantic annotation pattern that allows the connection of multimedia specific entities to domain specific descriptions. Figure 6 illustrates the decomposition pattern, where the input and output segment roles assume the corresponding multimedia content and segment type classes. In accordance with these patterns the developed ontological definitions are organised into separate modules, advocating a modular architecture. A number of auxiliary basic patterns, such as the digital data and localisation patterns, are introduced to enable the definition of the four previously described description patterns.

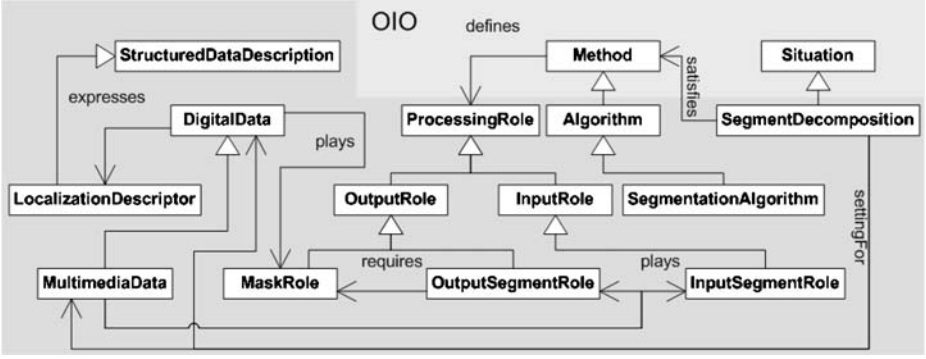


Fig. 6 COMM decomposition pattern (Figure from http://comm.semanticweb.org/Ontology)

Table 1 Generic characteristics of the state of the art MPEG-7 based multimedia ontologies

Multimedia ontology	Representation language	MPEG-7 coverage	Ontology design	Application context
Harmony	OWL full	Structure, visual	Monolithic	Analysis & annotation
aceMedia	RDFS	Structure, visual	Modular	Analysis & annotation
SmartWeb	OWL	Structure, visual	Modular	Analysis & annotation
BOEMIE	OWL DL	Structure, visual & audio	Modular	Analysis & annotation
DS-MIRF	OWL DL	Entire MDS,	Modular	Mpeg7 xml to rdf
Rhizomik	OWL DL	Entire MPEG-7	Monolithic	Mpeg7 xml to rdf
COMM	OWL DL	Structure, visual	Modular	Analysis & annotation

COMM has been utilised for supporting knowledge management of multimedia documents in industrial domains, including competitor car analysis and issue resolution in jet engines, in a fashion similar to previously described applications that relate to semantic content analysis and retrieval. A Java API[6] has been developed to facilitate the creation of COMM based multimedia descriptions, and the implementation of retrieval services, hiding the complex ontological constructs from the user/service developer. COMM is available at http://comm.semanticweb.org/Ontology.

2.8 Summary

Table 1 summarises the previously described MPEG-7 based ontologies. The name of the research project within which ontology development took place is used for reference, unless an established name exists for the ontology per se.

In their majority, the proposed ontologies follow a modular architecture, facilitating separation of concerns, extensibility and effective management of the produced metadata. All ontologies, except for the Rhizomik, have been manually constructed, and OWL, the official recommendation of W3C for the Semantic Web, appears to be the commonly selected representation language. As will be described in following sections though, there is not always a correspondence between the expressive power provided by the adopted representation language, and the constructed ontology definitions that model the intended descriptions. That is, many of the expressive language constructs remain unexploited.

Observing the intended application context, the developed ontologies appear to fall into two categories. The common goal underlying all MPEG-7 ontologies of course is to formalise the meaning of multimedia content descriptions. However, this aspiration is slightly differentiated with respect to the application contexts envisaged by the different ontologies. The DS-MIRD and Rhizomik ontologies, apart from

[6] http://multimedia.semanticweb.org/COMM/api

providing references for multimedia annotations generation, focus in particular on making visible and exploitable to the Semantic Web, existing MPEG-7 metadata repositories created by the XML communities.

In principle, the rest of ontologies could also be used to translate existing MPEG-7 metadata into corresponding RDFS or OWL descriptions, as long as the modelling rationale followed could be captured in a systematic methodology. Given though that their modelling targets more the intended semantics rather than the immediate consideration of all of the original MPEG-7 elements, such task could be considerably challenging. These ontologies though, can be very used to effectively create new multimedia content annotations, with formal semantics and in compliance with the MPEG-7 intended coverage. This trait associates also to the extensive usage of such ontologies in semantic content analysis applications.

As discussed later in the paper, where the complementary roles served by the individual ontologies are examined towards their harmonisation, ontologies such the Rhizomik and DS-MIRF may provide a first MPEG-7 to RDF translation, which could undergo further transformation into a more scalable and semantically effective representation using one of the other ontologies. This relates to the final observation regarding the supported MPEG-7 coverage. All ontologies besides the Rhizomik and DS-MIRF ones address media specific descriptions (i.e. structure and low-level features), leaving semantic aspects to external domain ontologies, to which linking is achieved by utilising an upper ontology that provides generic classes/properties that serve as attachment points. As aforedescribed, COMM is the only ontology that encompasses and formalises these interconnections in itself. Following a different perspective, the Rhizomik and DS-MIRF ontologies model the MPEG-7 Semantic DS description tools as well. Finally, all ontology, but for Rhizomik, have been constructed manually.

We note that not necessarily all elements included in the respective MPEG-7 parts are defined in the corresponding ontology, as in the most cases the goal is not to have a "1-to-1" mapping between MPEG-7 tools and ontology definitions but rather a functional mapping that provides equal description capabilities, and always with respect to the needs of the application at hand.

In the following sections, where different architectures for multimedia repositories are discussed and the modelling choices regarding structural descriptions and linking with domain ontologies are closely examined, the presented characteristics assume deeper insights.

3 Multimedia metadata architectures for semantic interoperability

The key issue in semantic interoperability is the ability to automatically process the information in a machine-understandable manner. The first step for achieving common understanding is a representation language that exchanges the formal semantics of the multimedia information. Systems that understand these semantics (agents, querying engines, etc.) can process the information captured in this language and provide semantically enabled services in the web like search and retrieval. Given the different context of usages, several definitions of semantic interoperability have been proposed in the literature. In the following, we present four alternative architectures for the meaningful communication of multimedia metadata, which

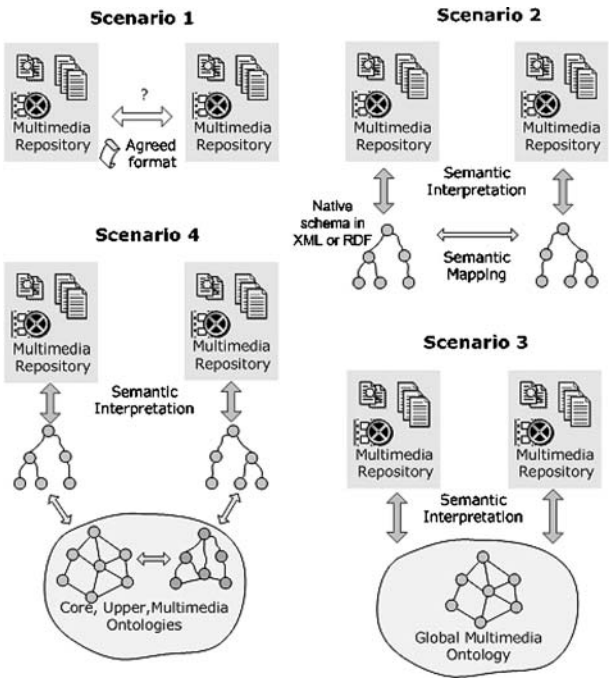


Fig. 7 Scenarios for semantic interoperable multimedia metadata repositories

due to the particular features pertaining to multimedia content, relate not only to the content that is conveyed but also to aspects characterising the document bearing this content (e.g. its structure, low-level features). Figure 7 illustrates possible architectural scenarios.

Under Scenario 1, the multimedia repositories exchange information in a predefined format, mutually agreed upon the involved parties. As the format does not necessarily provide for formal semantics, preservation of the intended meaning is achieved through strict conformance not only to the exchanging syntax, but to the way the syntactic structures should be semantically interpreted so that meaning can be attached to them. MPEG-7 and the various XML-based multimedia vocabularies, such as SMIL [35] and TV-Anytime,[7] fall in this category. Although information coming from multimedia repositories can be exchanged and processed, the lack of declarative semantics, restricts semantic-enabled management only to applications that implement the agreed interpretation. Consequently, there emerge serious concerns about information reuse, extensibility and mapping of semantics between heterogeneous applications. Another significant observation, is the fact that metadata vocabularies adopting this approach appear monolithic in terms of including self-adequate descriptions without foreseeing the need for integration with other (possibly complementary or further enhanced) metadata vocabularies.

Scenario 2 involves the adoption of formal knowledge representation languages that provide explicit, machine-understandable semantic interpretations of

[7] http://www.tv-anytime.org/

information using Unified Recourse Identifiers[8] (URIs). Each multimedia repository provides a formal description of its own metadata in an ontological form, by adopting the most appropriate Multimedia Standard and a formal representation language in accordance with the level of semantics that need to be expressed. Thereafter, using a semantic-preserving mapping framework (automatic or manual), the meaning (or at least an adequate fraction of it) can be exploited by other multimedia repositories. The idea of this scenario is simple and clear, however, there are several challenges, which, in most cases, arise from the diversity of the different multimedia repositories, the difficulties faced in isolated ontology construction, and the inability of automated mapping and reasoning tools to resolve these complexities.

Scenario 3 attempts a first solution to the above problems by proposing a global multimedia metadata ontology that serves as reference for providing a common understanding. The individual multimedia repositories do no construct their own ontologies, but use the global multimedia ontology in combination with other global ontologies (e.g. the Dublin Core [12] element set) in order to describe their metadata. The main drawback of this approach is that all multimedia repositories should use the same global ontologies, which introduces issues related to mutual agreement practices (Scenario 1). Furthermore, the construction of global ontologies that fulfil satisfactorily the needs of diverse applications is an extremely challenging task, even under very coarse assumptions that would render pragmatic the development of a single ontology. Thus in a way, Scenario 3 is a "semantic version" of Scenario 1, thereby inheriting a lot of its disadvantages.

Scenario 4 tries to fulfill the vision of the Semantic Web in its full technological potential. Different multimedia repositories use their own multimedia ontologies to declare the specific organisational view of their content. They also define mappings (manually or automatically) to core and upper multimedia ontologies, among which appropriate mappings / alignments already exist. Thus contrary to the vision of Scenario 3 which considers a global multimedia ontology, the core and upper multimedia ontologies proposed under Scenario 4 aim to serve a role analogous to that of DOLCE and SUMO in the case of domain-specific ontologies. Evidently, in the case of multimedia metadata, the corresponding core ontologies need to address additional concerns pertaining specifically to multimedia metadata and their management, either by extending parts of existing core conceptualisations such as DOLCE, or by introducing new semantic dimensions. Compared to the previous scenarios, this approach possesses increased advantages, affording a more integral solution of high modularity. The main challenges involved relate to the immature yet semantic technologies that would support a satisfactory level of automation in ontology mapping and alignment, especially for highly expressive semantics, with respect to the specific multimedia and subject matter ontologies that will be attached to it.

Naturally, the aforementioned Scenarios and induced considerations apply both to the media and subject-matter specific parts characterising multimedia metadata. Due to the differences in MPEG-7 coverage and intended application context, interoperability for certain MPEG-7 ontologies refers solely to structural descriptions. In this category fall the Harmony, aceMedia, SmartWeb and Boemie ontologies,

[8] http://www.w3.org/TR/uri-clarification/

since all of them assume the existence of separate domain-specific ontologies for subject matter metadata, and of a generic ontology to allow for their linking. Thus, sharing metadata produced with respect to these ontologies amounts to the definition of appropriate mappings among them (Scenario 2) or between them and the core ontology (Scenario 4). Apparently, Scenario 1 is not relevant since formal representation languages are used, while Scenario 3 is not applicable, since instead of using a common global ontology, individual ontologies have been developed to meet the viewpoints taken in each application context. In a sense though, it could be argued that each ontology, aiming to better meet the involved requirements, aspires in a way to serve the role of a global multimedia ontology addressing structural semantics.

The situation is different though, when a multimedia ontology addresses also domain-specific descriptions and defines the in-between the linking. Going though the existing ontologies, such is the case with DS-MIRF, Rhizomik and COMM. The first two serve as structural-specific ontologies, providing both the model and the vocabulary, while with respect to subject matter semantics, they only provide a very abstract model, where more specific ontologies are expected to enrich. COMM abstracts also the structural semantics and the linking with the subject matter descriptions through the extended DOLCE patters, as previously described. Consequently, DS-MIRF and Rhizomik appear to serve the purpose of a global multimedia ontology, while COMM, as also denoted by its name, aspires to serve as a core ontology, implementing a foundational abstract conceptualisation.

Bearing in mind the semantic interoperability view fulfilled by each Scenario, in the next two section we study and compare the modelling approaches taken by the existing MPEG-7 based multimedia ontologies with respect to the representation of media specific information and the representation of subject matter descriptions.

4 Modelling content structure semantics

As indicated in the aforementioned, two main reasons lie behind the modelling differences taken with respect to media specific descriptions. The use of XML Schema leaves the largest part of the intended semantics normative, thus encouraging different interpretations. Furthermore, the intended application context and the correspondingly induced interoperability scenario, introduce different engineering specifications. We note though that as the low-level descriptors correspond to rigid, numerical structures, there is no much room left for alternative interpretations and subsequently differing ontological models. Therefore, in the following we focus on the modelling choices taken with respect to structural semantics, delve into the rationale underlying the alternative proposals and discuss the discrepancies that hinder interoperability.

In order to illustrate the analysis in more intuitive manner, an annotation scenario is used in addition to the theoretical discussions, to serve as a concrete manifestation of the issues examined. As depicted in Fig. 8, the annotation example considers a specific frame from a football game video that depicts the football player Zidane. Consequently, the annotation needs to contain descriptions about the video, its temporal decomposition into a frame, the localisation of the frame, and its conveyed semantics by allowing its linking to the URI selected to identify Zidane.

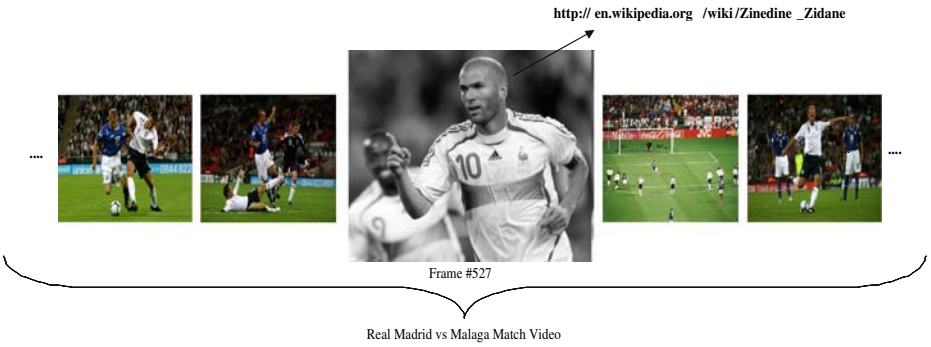


Fig. 8 Annotation example featuring the appearance of the football player Zidane in frame 527 of a football game video

4.1 The Harmony approach

In the approach taken by Hunter, modelling follows rigorously the standard's specifications, preserving the intended flexibility of usage. MPEG-7 specialisation relations are represented using subclass axioms, while consistency with respect to the specifications of the decomposition schemes is modelled by domain and range restrictions that are defined in accordance with the multiple interpretations an MPEG-7 description may have. Inevitably, the ambiguities resulting from non unique semantics are propagated.

The different segment classes (e.g. VideoSegment, StillRegion) subclass both the generic Segment class and the generic MultimediaContent class. As a result segment instances are interpreted as content instances too, with a number of implications. Amongst the convenient effects is that part-whole relations between the multimedia content types and the respective segments are handled indirectly through the subclass semantics. Such behaviour can be advantageous for certain applications. For example, a semantic query for an image depicting Zidane would return images containing a still region depicting Zidane, without requiring any refinement of the query. On the other hand though, such modelling prohibits the discrimination between content items and their constituent segments, a feature hindering the inference services, much as applications such as transcoding and retrieval where the interest lies on specific segments.

In the Harmony ontology, as in MPEG-7, multiple semantics can be attached to the different multimedia segment classes, allowing their interchangeable use with other segment or multimedia content classes. Apart from ensuring the initial flexibility of content descriptions, these multiple interpretations are used for modelling the recursive nature of the MPEG-7 decomposition patterns. To give a more intuitive insight, let us consider the case of recursive spatial decomposition. To represent spatial decomposition schemes, the Harmony ontology provides the *mpeg7:spatial_de-composition* property. The respective domain and range classes are defined to be the *mpeg7:MultimediaContent* class and the *mpeg7:Segment* class. Consequently, representing the examined image as an instance of the *mpeg7:Image* class, and the still region as an instance of the *mpeg7:StillRegion* class, we can link them through the *mpeg7:spatial_decomposition* property. In order to further

decompose this still region though, the original still region instance needs to be interpreted as an instance of the *mpeg7:Image* class, in order to form a valid domain for the *mpeg7:spatial_decomposition* property.

To better illustrate the involved issues, let us consider the annotation example and go through the produced description metadata shown in Table 2. As shown, we represented the video frame as an instance of the *mpeg7:StillRegion* class, resulting from the temporal decomposition of the video. However, according to the ontology definitions the *mpeg7:Video Segment* class could have been used as well, requiring no other modification but the replacement of the triple *(locator1,rdf:type, mpeg7:StillRegion)* with *(locator1,rdf:type,mpeg7:Vi-deoSegment)*. Had we chosen to represent the video frame as the result of spatial decomposition of the video, we could have represented the examined frame as an instance of the *mpeg7:MovingRegion* class. Alternatively, it could be modelled as the result of spatiotemporal decomposition, in which case the video frame could be modelled as an instance of any of the three aforementioned classes. Table 3 summarises all possible combinations for representing a video and its decomposition into a frame, with respect to the Harmony ontology. Apart from the remarkably large number of alternative representations that could be used without conflicting the ontology definitions, one notices that the loose axiomatisation of the Harmony ontology allows even for usages not included in the MPEG-7 specifications.

The discussed different alternatives that ensue from the multiple semantics characterising the ontology definitions, give an indicative example of the ambiguities suffered when attempting to model or interpret a description. Querying for example for *mpeg7:StillRegion* instances does not guarantee that all video frames fulfilling the query conditions are going to be retrieved, since some may be declared as instances of the *mpeg7:VideoSegment* or the *mpeg7:MovingRegion* classes. Similarly, a query centered on the *mpeg7:videoSe- gment_temporal_decomposition* property would fail in the general case, as the respective spatial or temporal decomposition properties may have been used.

Table 2 Annotation metadata for Fig. 1 using Hunter's approach

@prefix mpeg7:http://www.metadata.net/mpeg7/mpeg7.owl.
@prefix abc:http://metadata.net/harmony/ABC/ABC.owl.

:video1 rdf:type mpeg7:Video.
:video1 mpeg7:MediaLocator "http://multimedia.repository/soccer.mpeg".
:frame1: rdf:type mpeg7:StillRegion.
:video1 mpeg7:videoSegment_temporal_decomposition :frame1.

:frame1 hasLocator :locator1.
:locator1 rdf:type mpeg7:ParameterTrajectory.
:locator1 mpeg7:elipseFlag "true".
:locator2 rdf:type mpeg7:Box.
:locator1 mpeg7:InitialRegion :locator2.
:locator2 mpeg7:Coords "0 0 0 1024 1280 1024".
:time1 mpeg7:MediaDuration "0".
:time1 mpeg7:StartTime "527".
:locator1 mpeg7:mediaTime :time1.

:frame1 abc:realizes "http://en.wikipedia.org/wiki/Zinedine_Zidane".

Table 3 Alternative ways for the representation of a video decomposed into a frame in the Harmony ontology

Video representation	Decomposition dimension	Frame representation
Video	mediaSource_decomposition	Still region
		Moving region
		Video segment
Video	spatial_decomposition	Still region
		Moving region
		Video segment
Video	spatio-temporal_decomposition	Still region
		Moving region
		Video segment
Video segment	videoSegment_spatial_decomposition	Moving region
Video segment	videoSegment_temporal_decomposition	Still region
		Video segment
Video segment	videoSegment_spatio-temporal_decomposition	Still region
		Moving region

Additional observations related to the need for introducing the *hasLocator* property so as to enable the association of segments of content with their respective locators (spatial, spatiotemporal, etc.), and the lack of purely temporal locators. Due to the latter the direct localisation of the considered frame is not feasible; instead it requires its modelling as a moving region (covering the entire frame and with zero time spanning) and its subsequent identification through the provided *mpeg7:ParameterTrajectory* descriptor. Both observations reflect the immediate effect of the fact that each of the proposed ontologies has been developed within a particular context of usage, intended to fulfil a specific set of functionalities, and not necessarily address all elements included in a description tool.

A final comment concerns ontology engineering choices. The use of metamodelling, brings the developed ontology in OWL Full. OWL Full might be the more expressive of the three OWL species, yet it is undecidable, incurring implications in terms of efficient reasoning. Furthermore, the use of OWL Full initiates questions with respect to the poor utilisation of the expressive constructs supported. Class definitions include only subclass relations, when property restrictions could have been exploited to capture more precisely the modelled semantics.

4.2 The aceMedia approach

As aforementioned, the modelling rationale followed in the construction of MSO, is very similar to that of Hunter's, i.e. adherence to the MPEG-7 flexibility in handling multimedia items and segments, at any level of granularity. As an enhancement to Hunter's approach, a special class has been introduced for representing explicitly the notion of a video frame, namely the *mso:Frame*. In addition, the localisation of temporal decomposition is made straightforward through properties such as *mso:hasStartFrame*, *mso:hasEndFrame*, and *hmso:hasIndexFrame*. The introduction of classes and properties to explicit model semantically distinct notions contributes to a cleaner modelling. Yet, the use of the RDFS language does not allow one to benefit in terms of capturing complex underlying associations, for example, to

infer that an instance belongs to the Video class in case its temporal decomposition includes only video frames or video segments. Diverging slightly from the MPEG-7, the MSO treats the *MovingRegion* class as a specialisation of the *VideoSegment* one, demonstrating an alternative, complementary interpretation.

Although the *mso:VideoFrame* concept prohibits one from representing a frame as either a video segment, a moving region or a still region, as was the case in Hunter's approach, a large part of the flexibility in the original MPEG-7 Segment DS is preserved entailing corresponding ambiguities. Compliant with the MPEG-7 specifications, a video frame is treated as a specialised type of *mso:StillRegion* (note that *mso:VideoSegment* or *mso:MovingRegion* are also valid choices according to the MPEG-7 specifications). However, the lack of differentiation between content items and their decomposition segments, renders a video frame a specialisation of the *mso:Image* class also. Thus, the only way for one to determine whether an instance of *mso:Image* refers to a still image or a video frame is to check whether this instance has been used as filler in assertions representing spatio-temporal or temporal decompositions. As MSO does not explicitly provide the means to specify the physical locator of a media item, the *hasMediaLocator* property was introduced.

In Table 4, the annotation metadata generated in compliance with the Multimedia Structure Ontology are shown. Excluding the representation of video frames, all issues confronted due to multiple alternative representations and the entailed semantic ambiguities that apply to the Harmony ontology hold as well for the MSO ontology; as such we do not go into further detail here.

4.3 The SmartWeb approach

Unlike the approaches taken in the Harmony and the aceMedia ontologies, in the SmartWeb approach the generic Segment class is no longer modelled as a subclass of the generic MultimediaContent class. To allow the recursive application of decomposition schemes, explicit classes and corresponding properties have been introduced to capture each valid decomposition pattern. As a result, in close resemblance to the respective MPEG-7 decomposition elements, a set of concepts and properties of the form of "content-resultingSegment" and "segment-decompositionDimesions" are introduced (Fig. 9). Let us consider the case of a recursive temporal decomposition for video content. A video instance may be decomposed using the

Table 4 Annotation metadata for Fig. 1 using the aceMedia approach

```
@prefix mso:http://www.acemedia.org/ontologies/SCHEMA#.
@prefix ann:http://www.acemedia.org/ontologies/ANNOTATION#.

:video1 hasMediaLocator "http://multimedia.repository/soccer.mpeg".
:video1 rdf:type mso:Video.

:frame1: rdf:type mso:Frame.
:video1 mso:temporal_decomposition_of_a_video_segment :frame1.
:frame1 hasFrameIndex "527".

:frame1 ann:depicts :zidane1 rdf:type football:Player.
"http://en.wikipedia.org/wiki/Zinedine_Zidane" ann:hasDegreeOfConfidence "0.78".
```

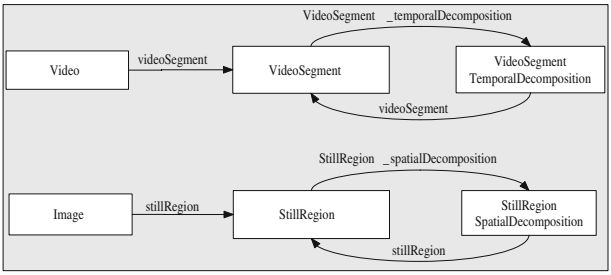


Fig. 9 The SmartWeb ontology model of recursive decomposition for video and images across the temporal and spatial dimension respectively

mpeg7:videoSegment property into instances of the *mpeg7:VideoSegment* class. To temporally decompose the latter further, the *mpeg7:VideoSegment_temporalDecomposition* (or its super-property the *mpeg7:temporalDecomposition*) is used to create instances of the *mpeg7:VideoSegmentTemporalDecomposition* class. These can be in turn decomposed into *VideoSegment* instances using the *mpeg7:video Segment* property, and the whole pattern is iterated until the desired level of partitioning is reached.

Note that two different properties, the *mpeg7:videoSegment* and the *mpeg7: temporal-VideoSegment_te- mporalDecomposition*, are used for exactly the same purpose, i.e. to express the temporal decomposition of video segments. Furthermore, the property *mpeg7:vi- deoSegment* is interchangeably used to link either video to video segments, or video segments to video segments. Adding to the aforementioned the use of both *mpeg7:VideoSegment* and *mpeg7: VideoSegmentTemporalDecomposition* classes for the representation of temporal sequences of frames, instead of advocating precise meaning, the ontology ends up with complex equivocal definitions. Similar observations apply for the rest decomposition dimensions. Note that in case of more than two levels of temporal decomposition, it is not possible to avoid using both classes since there is no decomposition property that can be recursively applied to any of them, nor are the two classes related through a subclass relation (or a common subsumer) to the classes used in the properties' domain and range restrictions.

Let us examine the aforementioned effects with respect to our exemplar annotation scenario, shown in Table 5 results. As illustrated in order to reach the granularity of a frame, first an artificial instance of *mpeg7:VideoSegment* needs to be defined, so that through, an artificial again, decomposition, we can eventually reach an instance of *VideoSegmentTemporalDecomposition*. It is the latter to which the *mpeg7:stillRegion* decomposition property can be applied and allow linking with the *mpeg7:StillRegion* instance representing the video frame under consideration. In the SmartWeb ontology, frames are treated as special types of still regions solely, and not as moving regions or video segments also, as in MPEG-7 and in the approaches described in Sections 2.1 and 2.2. This restriction however is delivered in a normative way only, through the use of *rdfs:comment*.

Another observation relates to the representation of media localisation information. Although the class *mpeg7:MediaInstance* is defined for representing physical content entities, it is not possible to link it with instances of the *mpeg7:MultimediaContent* class that represent the respective multimedia data (in Table 5*:video1* instantiates both classes to enable the description of the media physical location). An approach could be the use of the set of classes and properties implementing the *Ontology of Information Objects (OIO)* design pattern of DOLCE

Table 5 Annotation metadata for Fig. 1 using the SmarWeb approach

@prefix mpeg7:http://smartweb.semanticweb.org/ontology/mpeg7#.
@prefix smartmedia:http://smartweb.semanticweb.org/ontology/smartmedia#.

:annotation1 smartmedia:aboutMediaInstance :video1
:video1 rdf:type mpeg7:MediaInstance.
:video1 rdf:type mpeg7:Video.
:locator1 rdf:type mpeg7:MediaLocator.
:video1 mpeg7:mediaLocator :locator1.
:locator1 mpeg7:mediaURI "http://multimedia.repository/soccer.mpeg".

:frame1: rdf:type mpeg7:StillRegion.
:videoSeg1 rdf:type mpeg7:VideoSegment.
:videoSeg1 mpeg7:mediaTime :time1.
:vide01 mpeg7:videoSegment :videoSeg1.
:videoSeg1 mpeg7:temporalDecomposition temporalDecomposition1.
:temporalDecomposition1 rdf:type mpeg7:VideoSegmentTemporalDecomposition.
:temporalDecomposition1 mpeg7:stillRegion :frame1.

:ann1 rdf:type smartmedia:ContentAnnotation.
:ann1 smartmedia:relevance "0.78".
:ann1 smartmedia:aboutMediaInstance :frame1.
:ann1 smartmedia:aboutDomainInstance "http://en.wikipedia.org/wiki/Zinedine_Zidane".

that have been included in the SmartWeb Integrated Ontology (SWIntO). More specifically, *:video1* could be regarded an information object *realised* by the soccer.mpeg file. However, this would presuppose that the *mpeg7:MultimediaContent* can act as an instance of the *smartdns:information-object*, requirement that currently is not satisfiable as *mpeg7:MultimediaContent* neither subclasses the *smartdns:information-object* nor can be linked to the latter through a property.

In a recent publication [33], the authors present an integrated model, the so called *SmartMediaLing*, that formally describes the alignment between the different ontology components, through the application of the DOLCE D&S (Description&Situation) and OIO (Ontology of Information Objects) patterns (see Fig. 10). Citing from [33], Section 4.2: "...we identify the picture itself as an *information object* that is *about* and *entity*..The picture can be decomposed to different segments..A *SegmentDecomposition* is an *information object* carrying the result of a segmentation process, a *perdurant* applied by some *agent*...". However, the corresponding ontology files are not yet publicly available.

4.4 The Boemie approach

As described in Section 2.4, the principle underlying the engineering has been the explicit axiomatisation of the addressed MPEG-7 description tools semantics. Two disjoint classes, the *mco:MultimediaContent* and *mco:MultimediaSegment*, make explicit the discrimination between content items and respective segments that results from their decomposition. The semantics of the various multimedia content items, multimedia segments, locators and decompositions have been formally described through restrictions. For example, the definition of the *mco:VideoFrame* class does not include only the subclass relations with respect to its more generic

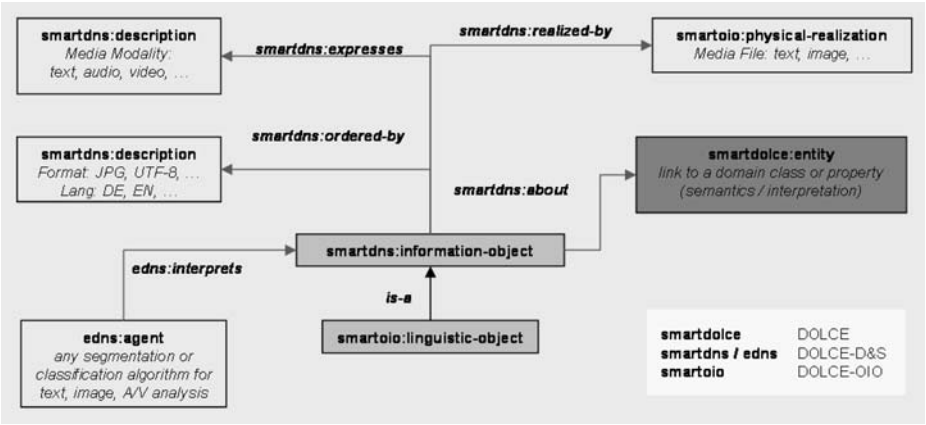


Fig. 10 The Smartmedia ontology integrated in the OIO module (Figure from [16])

mco:VideoSegment class. In addition, property restrictions are defined in order to capture that *mco:VideoFrame* instances can result only through a time point temporal decomposition applied to either *mco:Video* or *mco:VideoSegment* (temporal video segment whose duration spans over a non-zero time interval). Thus, ambiguities and complexities in retrieval, such as those aforementioned of frame instances represented as still regions, images, video segments or moving regions, and modelled as the outcome of decomposition across different dimensions and multimedia entities, are overcome.

The description generated when using the MCO ontology is shown in Table 6. Unlike the previous approaches, where multiple combinations of different classes and properties could be used and still deliver the same semantics, in the Boemie approach there is a unique correspondence between the ontological definitions the semantics captured. This may appear restrictive compared to the multiple ways afforded by the original MPEG-7 Schemas and the ontologies who preserve them, we note though that it these multiplicity in descriptions of semantically equivalent notions that entails the ambiguities and complexity that characterises MPEG-7 metadata and hinders their semantic management.

Table 6 Annotation metadata for Fig. 1 using the Boemie approach

@prefix mco:http://www.boemie.org/BOEMIE_ontologies/mco_v2.8.owl.

:video1 rdf:type mco:Video.
:video1 mco:hasURL “http://multimedia.repository/soccer.mpeg”.
:frame1 rdf:type mco:VideoFrame.
:video1 mco:hasMediaDecomposition :frame1.

:index1 rdf:type mco:TimeLocator.
:frame1: mco:hasSegmentLocator :index1.
:index1 mco:hasStartOffset “527”.
:index1 mco:hasEndOffset “527”.

:frame1 mco:depicts “http://en.wikipedia.org/wiki/Zinedine_Zidane”.

As in the case of COMM, described in the following (Section 4.7), the richer axiomatisation entails increased complexity when it comes to tasks of consistency checking and entailment, i.e. when validating that the produced RDF descriptions are in compliance with the MCO conceptualisaton and when making explicit through inferences information that is implicit. The richer semantics affect also engineering, challenging best practises in ontology design. The use of existential restrictions ensures in general that the entailed inferences will be the intended ones; when used in concept definitions however, they result in unnecessary strict conditions. For example, defining an image as $Image \equiv \exists\ hasMediaDecomposition.StillRegion$, prohibits the existence of an image that is not decomposed. Using a universal restriction instead, captures better the intended semantics, but may result in unexpected inferences due to the open world assumption semantics. The approach followed in MCO, was the adoption of universal restrictions to precisely reflect the intended semantics, and the combined use of existential and universal restrictions only when truly required, e.g. to ensure the existence and appropriateness of locators for the segments resulting from decomposition.

OWL DL does not allow to fully capture all intended semantics. In some cases qualified cardinality restrictions would be required to enhance the precision of the definitions, while more complex constructs, such as role chains or even rules, are required to make explicit some of the underlying associations. Propagating the subject matter descriptions from a segment to the entire content item is an example case that would benefit from the usage of rules. It is important to stress that such concerns, although of great importance from a knowledge engineering point of view, can be practically reconciliated in many cases, especially when considering the trade off between expressivity and reasoning efficiency. For example, when querying for videos depicting a particular athlete, expanding the query to address all types of segments that form valid decompositions of video content, will retrieve all items expected. Note that this is different from the query customisations mentioned in Section 4.1 for example, where the lack of precise semantics is the reason that renders complex queries necessary.

4.5 The Rhizomik approach

As described previously, the Rhizomik approach consists in the translation of the complete MPEG-7 Schema through automatic XSD2OWL and XML2RDF mappings. This means that description tools and elements that have been omitted or partially modelled in the aforementioned approaches, are now included. As a result, the number of triples required for semantically equivalent descriptions is significantly increased. Let us consider for example the representation of a video. Using one of the aforementioned ontologies, it would involve a single triple where the instance representing the video item is linked through rdf:type to the class representing Video entities. Under the Rhizomik approach though, the complete MPEG-7 ContentEntityType description schema is instantiated.

In Sections 4.1, 4.2 and 4.3, we detailed how adherence to the MPEG-7 flexibility resulted in classes and properties with equivocal semantics. The Rhizomik approach, constituting a rigorous translation of the XML Schema based definitions of MPEG-7 descriptions tools suffers similar ambiguities. It is important to note, that these ambiguities result from the syntactic variability allowed in MPEG-7, and not from

the lack of explicitly axiomatising the intended semantics in the Rhizomik ontology. Contrary to the Hunter approach, which utilises poorly the expressivity provided by OWL, and the aceMedia and SmartWeb approaches that stay in RDFS, the Rhizomik approach, makes the MPEG-7 intended semantics formal through the use of OWL DL constructors. Consequently, querying for video frames may still require customisation in order to include still regions, moving regions and video segments instances, yet, property restrictions used in the content and segment class definitions ensure that only valid decomposition schemes can be applied.

However, staying strictly compliant to the MPEG-7 specifications, common interpretations shared among different classes and properties are allowed. As an example, let us consider the properties *mpeg7:StillRegion* and *mpeg7:SpatialDecomposition*. Both of them represent decomposition along the spatial dimension. Their difference is that *mpeg7:StillRegion* is intended to be used when the outcome of the spatial decomposition is a still region. In addition, there is the *mpeg7:StillRegionSpatialDecompositionType* class defined as a restriction on the *mpeg7:StillRegion* property, allowing only instances of *mpeg7:StillRegionType* as fillers. Although such redundancy is justifiable in the MPEG-7 XML Schema structures, so that the intended usage and application of description tools is represented, in the case of logic based languages such as OWL DL it leads to a rather perplexed and messy conceptualisation.

Table 7 illustrates the RDF description when using the Rhizomik ontology. Even in this short annotation example, one can observe the classes and properties referring

Table 7 Annotation metadata for Fig. 1 using the Rhizomik approach

```
@prefix mpeg7:http://rhizomik.net/ontologies/2005/03/Mpeg7-2001.owl#.

:mm1 mpeg7:Description :contentType1.
:contentType1 rdf:type mpeg7:ContentEntityType.
:contentType1 mpeg7:MultimediaContent :videoType1.
:videoType1 rdf:type mpeg7:VideoType.
:videoType1 mpeg7:Video :video1.
:video1 rdf:type mpeg7:VideoSegmentType.
:video1 mpeg7:MediaLocator :locator1.
:locator1 mpeg7:MediaUri “http://multimedia.repository/soccer.mpeg”.

:video1 mpeg7:SpatialDecomposition :spatialDecomposition1.
:frame1 rdf:type mpeg7:MovingRegionType.
:spatialDecomposition1 rdf:type mpeg7:VideoSegmentSpatialDecompositionType.
:spatialDecomposition1 mpeg7:MovingRegion :frame1.

:frame1 mpeg7:TemporalMask :mask1.
:mask1 rdf:type mpeg7:TemporalMaskType.
:time1 rdf:type mpeg7:MediaTimeType.
:mask1 mpeg7:SubInterval :time1.
:loc1 rdf:type mpeg7:PositionType.
:loc1 mpeg7:TimePoint “157”.
:loc1 mpeg7:Duration “0”.

:frame1 mpeg7:Semantic “http://en.wikipedia.org/wiki/Zinedine_Zidane”.
```

to MPEG-7 elements that are omitted by the other approaches. As already implied in Section 2, this is an immediate effect of the intended application context. The main concern of the Rhizomik ontology is to translate existing MPEG-7 metadata to RDF descriptions, and thereby make them visible to the Semantic Web and the available services; consequently, covering all MPEG-7 elements is an indispensable requirement. The one to one translation followed, means that as in the case of the Harmony, aceMedia and SmartWeb ontologies, multiple varying descriptions may be used to convey the intended meaning. As the examples previously given cover this issue in detail, we do not go through specific examples here.

4.6 The DS-MIRF approach

The transformation principles from MPEG-7 XML to RDF underlying the DS-MIRF approach are in the same vein with the Rhizomik approach. This is a direct consequence of the aspiration, the two approaches share, for using the proposed MPEG-7 ontologies as an upper model for the description of multimedia content. Thus, all traits discussed previously for the modelling of structural description in the Rhizomik ontology, apply with respect to the DS-MIRD structural descriptions representation.

Specifically, examining the DS-MIRF modelling in detail, one observes that the representation of the decomposition semantics is practically identical, besides slight variations incurring from the more explicit modelling of specific XML constructs. As a result, the aforementioned issues with respect to semantic ambiguities of descriptions with more than one meanings hold for the DS-MIRF as well. For example, one could use one of the *mds:MovingRegionType*, *mds:VideoSegmentType* or *mds:VideoType* to represent the frame under consideration, employing different types of decomposition respectively. Thus, semantic interoperability issues persist. However, DS-MIRF and Rhizomik differ substantially in the way the linking with external domain specific ontologies is modelled. As detailed in Section 5, the DS-MIRF framework allows the systematic integration of domain ontologies in MPEG-7 descriptions.

Table 8, shows a DS-MIRF compliant RDF description. As for the Rhizomik ontology, different segment classes could have been used to represent the frame under consideration, and the same holds for the selected decomposition representation.

4.7 The COMM approach

Based on the Ontology of Information (OIO) and Descriptions&Situations (D&S) design patterns, COMM axiomatises the description of content decomposition and annotation, at semantic and media level. The decomposition pattern models content partitioning as a *dns:si- tuation* satisfying a particular *decomposition (segmentation)* algorithm, which in turn defines the input and output multimedia entities, as well as the parameters related to the segmentation algorithm and the localisation of the output segment.

It is this different perspective that constitutes the re-engineering of MPEG-7 specifications when compared to the aforementioned approaches. This axiomatisation, though seeming a bit cumbersome, makes straightforward the construction of

Table 8 Annotation metadata for Fig. 1 using the DS-MIRF approach

```
@prefix mds:http://127.0.0.1:8080/ontologies/MPEG703/MDS#.
@prefix visual:http:////127.0.0.1:8080/ontologies/MPEG703/Visual#.

:content1 rdf:type mds:ContentEntityType.
:content1 mds:MultimediaContent :videoType1.
:videoType1 rdf:type mds:VideoType.
:videoType1 mds:Video :video1.
:video1 rdf:type mds:VideoSegmentType.
:video1 mds:MediaLocator :mediaLoc1.
:mediaLoc1 rdf:type mds:MediaLocator.
:mediaLoc1 mds:MediaUri "http://multimedia.repository/soccer.mpeg".

:video1 mds:TemporalDecomposition :videoSegType1.
:videoSegTemp1 rdf:type mds:VideoSegmentTemporalDecompositionType.
:videoSegTemp1 mds:Mediatime "527".
:videoSegTemp1 mds:StillRegion :frame1.
:frame1 rdf:type mds:StillRegionType.

:frame1 mpeg7:Semantic "http://en.wikipedia.org/wiki/Zinedine_Zidane".
```

such descriptions, defining precisely the individuals required and how they should be interrelated. In most of the aforementioned approaches, this is only implicitly indicated, i.e. one starts with one of the main entities to be represented (e.g. the video or frame representation in our example) and following properties linked to them through domain and range axioms, figures out how to express the decomposition description. The use of more expressive constructors in the Boemie and Rhizomik approaches, assists in making the description process more explicit, providing in a way "representation patterns" through the class definitions (e.g. in Boemie, a still region is defined as a specialisation of those entities that have as prerequisite a spatial locator and that may be further spatially decomposed into still regions).

COMM however, stays in a quite high abstraction level when defining the respective multimedia patterns. For example, in the definition of the *visual:still-region-spatial-deco- mposition* class, which corresponds to the MPEG-7 StillRegionSpatialDecompositionType, there is no restriction on the allowed type of segments that may result. That is, only still region segments that are identified by spatial locators and can be further decomposed only along the spatial dimension (as axiomatised in the Boemie MSO ontology). This on one hand, leaves each application responsible for ensuring correct usage, and on the other hand, introduces effects of less formal interpretations that following strictly the MPEG-7 Schemas, end up with classes and properties, whose semantics are not distinct. The properties *visual:still-region-spatial-decomposition* and *visual:spatial-segment-decomposition* are such an example of the latter.

Following the COMM approach, the RDF metadata describing the temporal decomposition and the semantic annotation are shown in Table 9. For brevity, and since they are also modelled as a a *dns:situation*, the localisation descriptions have been omitted.

Table 9 Decomposition metadata for Fig. 1 using the COMM approach

```
@prefix core:http://comm.semanticweb.org/core.owl#.
@prefix dns:http://comm.semanticweb.org/extended-dns-very-lite.owl#.
@prefix loc:http://comm.semanticweb.org/localization.owl#.
@prefix visual:http://comm.semanticweb.org/visual.owl#.

:video1 rdf:type core:video-data.
:video1 dns:realized-by “http://multimedia.repository/soccer.mpeg”.

:inputSegmRole1 rdf:type core:input-segment-role.
:video1 dns:plays :inputSegmRole1.
:frame1 rdf:type core:image-data.
:stillRegionRole1 rdf:type core:image-data.
:frame1 dns:plays :stillRegionRole1.
:temporalDecomposition1 rdf:type visual:temporal-segment-decomposition.
:segmentationAlgorithm1 rdf:type core:segmentation-algorithm.
:temporalDecomposition1 dns:satisfies :segmentationAlgorithm1.

:temporalDecomposition1 dns:setting-for :video1.
:temporalDecomposition1 dns:setting-for :frame1.
:temporalDecomposition1 dns:setting-for :frameLocDD1.
:segmentation1 dns:defines :still-region-role1.
:segmentation1 dns:defines :input-segment-role1.
:segmentation1 dns:defines :temporalMaskrole1.
:temporalMaskRole1 rdf:type loc:temporal-mask-role.
:still-region-role1 dns:requires :temporalMaskRole1.

:manualAnn1 rdf:type dns:method.
:semAnn1 rdf:type core:semantic-annotation.
:semAnn1 dns:satisfies manualAnn1.
:manualAnn1 dns:defines annDataRole1.
:annDataRole1 rdf:type core:annotated-data-role.
:frame1 dbs:plays annDataRole1.
:manualAnn1 dns:defines semLabelRole1.
:semLabelRole1 rdf:type core:semantic-label-role.
“http://en.wikipedia.org/wiki/Zinedine_Zidane” dns:setting semAnn1.
:frame:1 dns:setting semAnn1.
```

4.8 Summary

Table 10 summarises the previously elaborated characteristics of the examined MPEG-7 based ontologies concerning the modelling of content structure descriptions.

Apart from the Rhizomik and DS-MIRF ontologies, that follow a one to one mapping, covering all MPEG-7 Schema elements, the rest of the ontologies follow a simplified modelling, aiming to provide an effective, yet efficient representation of the MPEG-7 descriptions meaning. As a result, the RDF metadata produced when using these two ontologies, confront the same complexity and size issues that apply to MPEG-7 XML metadata. These issues are further aggravated by the fact that due to the formalised now semantics, the produced descriptions are expected to undergo semantic management on the basis of the meaning conveyed. Yet, as detailed above,

Table 10 Summary of structural descriptions modelling

MM ontology	Structure semantics modelling
Harmony:	Multiple meaning class&properties, segment types subclass MM content, types, property-centric decomposition model, domain/range restrictions model valid decompositions, multiple representations/interpretations per description
aceMedia:	Multiple meaning class&properties, similar to Harmony, enhanced by explicit entities, multiple representations/interpretations per description
SmartWeb:	Multiple meaning class&properties, distinguishes segment types from mm content types, class centric decomposition model, multiple representations/interpretations per description
BOEMIE:	Unique meaning class&properties, disjoint segment and mm content classes, definitions wrt to decomposition & localisation restrictions unique reprsentation/interpretation per description
DS-MIRF:	One to one MPEG-7 translation, multiple meaning class&properties, makes explicit all XML elements, multiple representations/interpretations per description
Rhizomik:	One to one MPEG-7 modelling, multiple meaning class&properties, multiple representation/interpretations per description
COMM:	Re-engineers MPEG-7 descriptions, extends DOLCE design patters, abstracts description entities and their interconnection unique reprsentation/interpretation per description

this one to one translation from the MPEG-7 definitions, propagates all semantic ambiguities, rendering the intended meaning ambivalent.

Following a modified modelling though, does not make the rest of the ontologies impervious to such ambiguities. The Harmony, aceMedia and SmartWeb ontologies, preserve the flexibility of usage intended in the original MPEG-7 descriptions. Thus, though many description elements are circumvented, multiple interpretations are allowed per the defined semantic entities, and different ontological constructs are allowed to share equal semantics. The Harmony ontology is the one closer to the MPEG-7 normative specifications with respect to the usage of descriptions. The aceMedia ontology brings some more conceptual clarity by adding some explicit classes and properties to capture notions otherwise attributed to descriptions with other meanings as well. The SmartWeb ontology contributes further to clean semantics by distinguishing the notions of segment types and multimedia content types. As elaborated in Section 6, these multiple possible interpretations issue a challenging setting when it comes to aligning these ontologies in order to make the descriptions conforming to each of them interoperable to each other.

Aiming to provide for well-founded semantics, the Boemie and COMM ontologies take a different standpoint. They both re-engineer the MPEG-7 descriptions in order to axiomatise the semantics identified in the specifications. The Boemie ontology introduces disjoint classes to represent the different multimedia content and segment types semantics, and definitions based on property restrictions that capture decomposition and localisation patterns. Thereby, it advocates clean semantics, subject to reasoning. Furthermore, it enables to enrich the structural descriptions with additional aspects. COMM defines foundational patterns allowing to formalise MPEG-7 descriptions as well as processes, by extending DOLCE's design patters. The resulting multimedia patterns, abstract the entities involved in the descriptions as well as their interconnection.

In terms of complexity, the study of the individual modelling choices, reveals that half of the examined ontologies utilise in practise only a very small subset of the expressive power afforded by the representation language. The Rhizomik, DS-MIRF and Boemie, and COMM ontologies make extensive use of the language constructs provided, falling indeed in the OWL DL subset. The Harmony ontology on the other hand, is a counterexample, being in OWL Full due to heavy metamodelling usage, while the definitions it considers lie in RDFS.

5 Linking with domain-specific ontologies

In the previous section, we examined how the different MPEG-7 based ontologies interpret and translate the MPEG-7 structural descriptions. In this section, we address the other key dimension pertaining to multimedia metadata, namely linking with domain specific ontologies in order to represent subject matter annotations.

The Harmony ontology assumes the existence of an upper ontology in order to be interconnected with other ontologies. In the deployed applications, the ABC upper model has been used for this purpose. The ABC model [22] integrates a number of basic entities and relationships, including among others agents, places, tangible objects, time and object modifications. It has been built with two purposes in mind. First, to provide a core model based on which domain specific ontologies&vocabularies can be constructed. Second, to provide foundational definitions based on which mappings between different ontologies can be implemented. The ABC classes serve as attachments points for both the MPEG-7 entities and the domain specific ones, while the the ABC properties define in which ways they can be linked. More specifically, the domain specific definitions can be aligned with ABC by specialising corresponding ABC abstract classes and properties, such as *abc:Agent*, *abc:Artifact*, *abc:hasAgent*, *abc:inPlace*, etc. The multimedia specific ontologies on the other hand are aligned by subclassing the *abc:Manifestation* class.

In the aceMedia approach, DOLCE Lite provides the abstract classes and properties that both the multimedia and domain specific ontologies may extend in order to interrelate. In addition, a specially dedicated ontology, the so called Annotation Ontology (AO), has been constructed. The latter provides specialises DOLCE entities in order to provide the concrete means to interlink the ontologies. Two properties of particular interest are the *ann:depicts* property, which implements the linking of multimedia instances to the domain semantics depicted, and the *ann:hasDegreeOfConfidence* property, which represents the plausibility of an annotation.

SmartWeb follows a similar rationale, assuming an upper ontology. In the implemented framework, this role is served by SmartSUMO the foundational ontology designed with the project by coupling DOLCE and SUMO. Going further than the previously described upper ontologies approaches, SmartSUMO not only provides the generic classes, which the multimedia and domain ontologies are supposed to subclass, but supports the formalisation of the annotation process itself. To represent the latter, the *ContentAnnotation* class is defined. A content annotation is linked to domain semantics through the *aboutDomainInstance* property and to multimedia items through the *aboutMediaInstance* property. The range of the *aboutDomainInstance* is the generic *dolce:Entinty* class, that superclasses all domain specific classes.

The range of the *aboutMediaInstance* is the *smartmedia:ContentOrSegment* class, that it the union of the multimedia content and segment classes. In the BOEMIE approach, a very simple approach has been adopted for integration with subject matter ontologies. A property called *mco:depicts* has been defined, having as domain the union of *mco:MultimediaContent* and *mco:MultimediaSegment* classes and as range the top level concept of the domain concepts hierarchy.

The DS-MIRF framework implements a diametrically different approach, setting MPEG-7 in the role of a core ontology for integrating media and domain related knowledge. The abstract semantic entities defined in the MPEG-7 Schema, translated into respective OWL constructs, play in the DS-MIRF context, a role equivalent to that of the abstract classes and properties of core ontologies such as DOLCE or the ABC. Since the semantic breadth covered by MPEG-7 is quite restricted, in this aspect closer to ABC rather than DOLCE, domain ontologies cannot be expected to extend the provided axiomatisation. Hence, DS-MIRF provides a methodology so that the domain assertions can be translated into ontological statements compliant to the MPEG-7 conceptualisation, thus realising the interoperability of semantics. It is important to note that under the DS-MIRF domain assertions translation, no assumption is made about the ontological representation of the media related aspects of the content.

Espousing MPEG-7 as the core ontology for structuring and integrating multimedia content descriptions is the paradigm taken by the Rhizomik approach too. However, when it comes to integration with domain specific ontologies, the Rhizomik approach is applicable only under the presumption that these domain ontologies have been beforehand re-engineered so that they are compliant to the classes resulting from the corresponding Semantic DS structures. In this sense, despite sharing a common goal, the Rhizomik and DS-MIRF approaches are substantially different. DS-MIRF does not require for the MPEG-7 Schema to be extended (e.g. by domain specific definitions) before translating the content descriptions into MPEG-7 compliant, ontological assertions. The Rhizomik approach on the other hand, needs to have available the domain specific XML Schemas that extend MPEG-7 Semantic description tools, before applying the XML2OWL and XML2RDF mappings that generate the ontology and the RDF statements populating it.

From the considered MPEG-7 based ontologies, COMM is the one that addresses the integration with domain ontologies in a most formal fashion. Building on the OIO and D&S design patterns of DOLCE, COMM defines the *Semantic Annotation* pattern to allow the linking of multimedia descriptions with domain descriptions provided by external domain ontologies. Instead of directly attaching a domain instance to the conveyed semantics, it is related to the way this semantic description was obtained. The alternative methods modelled include its manual acquisition or its automatic generation through the application of a corresponding semantic analysis algorithm.

Summing up, the afore described methodologies outline three alternative perspectives towards linking with domain specific ontologies (Table 11). The first one considers the use of an upper ontology, which provides on one hand generic classes that the individual multimedia and domain ontologies may subclass, and on the other hand generic properties that may be used per se, or further specialised, in order to connect them. This is the perspective taken in the applications where the Harmony and aceMedia approaches have been employed. The framework adopted in the case

Table 11 Summary of the approaches taken to support linking with domain ontologies

MM ontology	Linking with domain ontologies
Harmony:	Generic classes and properties provided by an upper ontology, (ABC has been used in Harmony applications)
aceMedia:	Generic classes and properties provided by an upper ontology (DOLCE and AO have been used in aceMedia applications)
SmartWeb:	Combination of DOLCE&SUMO generic classes and properties enhanced with DOLCE OIO pattern
BOEMIE:	Generic properties whose domain&range restrictions span across the multimedia and domain ontologies
DS-MIRF:	MPEG-7 Semantic DS abstract classes and properties
Rhizomik:	MPEG-7 Semantic DS abstract classes and properties
COMM:	Foundational multimedia patterns that extend DOLCE's OIO and D&S patters

of the Boemie ontology, is a simplification, where the interlinking properties have been included in the multimedia ontology, with their domain and range restrictions spanning both the multimedia and domain ontologies.

In the second methodology, the MPEG-7 Semantic DS abstract entities are used as the foundational model to which the domain specific ontologies are expected to align. This is the approach realised in the Rhizomik and DS-MIRF application frameworks, although following quite different rationales. In the DS-MIRF approach, a systematic methodology allows to integrate domain knowledge by specialising the MPEG-7 Semantic DS classes. In the Rhizomik approach, linking can be achieved only as long as the domain ontologies have been re-engineered into MPEG-7 compliant descriptions. The third methodology consists in the formalisation not only of the linking between multimedia and domain ontologies, but axiomatises the process of linking as well. Under this category fall the linking approaches taken by SmartWeb and COMM. SmartWeb models the linking as a content annotation instance, through the *ContentAnnotation* class, to which the corresponding multimedia and domain instances interconnect. COMM advances the foundational grounding by extending the OIO and D&S design patterns of DOLCE. Thereby, content annotation is represented as the result of an annotation method, including an assertions serving as the semantic annotation, and a description of the multimedia data that is being annotated.

We conclude the discussion on the proposed linking approaches, with some remarks regarding the interchangeability of the methodologies undertaken by the individual MPEG-7 based ontologies. In all cases, there is no modelling correlation between the way the linking and the representation of structural descriptions. Consequently, any of the linking schemes could have used in combination with any of the multimedia ontologies. Leaving out the DS-MIRF and Rhizomik ontologies that adhere more to the MPEG-7 XML to RDF translation as detailed in previous sections, all combinations are viable. For example, linking with respect to the Harmony ontology could be achieved using DOLCE and the Annotation Ontology

as in the aceMedia approach, using the *depicts property* of the Boemie approach and defining its range as the MultimediaContent class of the Harmony ontology (i.e. the class superclassing all content and segment classes), using COMM's semantic annotation pattern, by having the MultimediaClass subclass COMM's DigitalData class, etc. Analogously, any of the rest ontologies (parts sharing the same coverage) could have been used in the place of the Harmony ontology in the integration framework under the ABC ontology.

6 Towards interoperable multimedia metadata

In the previous sections, we examined closely on one hand the modelling choices regarding the representation of structural aspects, and on the other hand the linking with domain descriptions provided by external domain ontologies. Furthermore, comparing the different approaches taken, we outlined issues regarding their interoperability, the complementary roles served, intended usage, etc. In this section, we assess holistically the ontologies with respect to the metadata interoperability architectures presented in Section 3 and discuss possible harmonisation approaches.

The Harmony, aceMedia and Boemie ontologies seek to formally describe content structure, relying on domain specific ontologies to provide subject matter descriptions. Linking is aspired in the form of interconnected generic classes provided by an upper ontology or by class / property definitions that span over different ontologies. As such, with respect to structural content description, the three ontologies fall under Scenario 2. Each of them reflects a different viewpoint, pertaining to differing requirements amongst semantic multimedia description applications. To make interoperable the three ontologies, their conceptualisations should be aligned. As outlined previously, the cleaner the semantics of an ontology, the easier in principle the determination of mappings. As result, metadata conforming to the Harmony and aceMedia are harder to reconcile and share among heterogenous systems, since noth ontologies are characterised by unequivocal meaning and semantic ambiguities.

SmartWeb and COMM adhere to a different rationale, attempting to formalise linking with domain ontologies. In the SmartWeb approach, the process of linking is modelled as an annotation instance to which the corresponding multimedia and domain instances are attached. Taking a modularised view, the part of the SmartWeb ontology that addresses content structure may be used as a reference point to which other ontologies that target structural semantics can be aligned. Although in a lesser degree than other ontologies (see the aceMedia or the corresponding Rhizomik parts), the SmartWeb ontology suffers still significant semantic ambiguities though. COMM formalises further the representation of multimedia and domain ontologies connection, and through the extension of DOLCE's design patters provides a cleaner and more rigorous foundational grounding. COMM is by definition intended to serve as a core multimedia ontology that formalises the description of multimedia content across the different aspects involved, while allowing further specialisations through more specific, multimedia or domain, ontologies. Again, the more conceptually clear these specific ontologies are, the easier the definition of specialisation and relevance mappings.

The DS-MIRF and Rhizomik ontologies place MPEG-7 in the role of an upper ontology, as an effect of their end goal, i.e. to make existing MPEG-7 XML multi-

media metadata repositories visible to the Semantic Web, and vice versa. Treating MPEG-7 as the standardised, most comprehensive set of tools for the description of multimedia content, multimedia descriptions are expected to conform to its, formalised now, specifications. DS-MIRF allows further interaction with domain specific ontologies, through the systematic integration methodology provided. Under these considerations, both ontologies fall within Scenario 3. Serious considerations emerge though with respect to using the MPEG-7 abstract Semantic DS model as the upper foundational definitions to which domain specific ontologies should align. A simple comparison with foundational ontologies such as DOLCE and SUMO indicates concisely the inappropriateness of MPEG-7 to serve such a role.

Considering solely the parts that address the representation of content structure, two lines of argumentation incur with respect to the considered architectural scenario. Under Scenario 3, the structural model of the DS-MIRF / Rhizomik ontologies would correspond to an upper model, expected to be used by all applications wishing to formally describe content structure. However, as shown in previous sections, the one to one translation of MPEG-7 adopted by both ontologies, entails serious semantic ambiguities that challenge the ontology engineering principles and pose severe obstacles to the utilisation of typical inference services. Clearly, generating multimedia metadata with inherently vague meaning is problematic. Considering the two ontological models as the effect of different application requirements, thus under Scenario 2, achieving interoperability with other multimedia ontologies pertains to similar issues to the ones discussed above for the cases of the Harmony and aceMedia ontologies.

Table 12 Summary of MPEG-7 based ontologies with respect to interoperability architecture scenarios

MM ontology	Interoperability architecture scenarios
Harmony:	Scenario 2, interoperable content structure metadata through mappings to other content structure modelling ontologies
aceMedia:	Scenario 2, interoperable content structure metadata through mappings definition
SmartWeb:	Scenarios 2 and 4, interoperable structure metadata through mappings, linking with domain ontologies through foundational ontologies
BOEMIE:	Scenario 2, interoperable content structure metadata through mappings, cleaner semantics, easier to resolve mappings
DS-MIRF:	Scenarios 2 and 3, MPEG-7 as an upper multimedia ontology, linking with domain ontologies wrt MPEG-7 Semantic DS interoperable structure descriptions through mappings
Rhizomik:	Scenarios 2 and 3, MPEG-7 as an upper multimedia ontology, interoperable structure descriptions through mappings
COMM:	Scenario 4, core ontology, formalises content media specific description patters and linking with domain ontologies

The aforementioned considerations, summarised in Table 12, sketch possible approaches to benefit the most from the experiences drawn by the individual initiatives and to allow for their interoperability. Thinking of the Semantic Web architecture, COMM undoubtedly provides the more appealing and integral approach towards the formal integration of multimedia descriptions on the Web. However, as described in Section 4.7, staying in a quite abstract level, the, at least currently defined multimedia design patterns, need to be further extended to account for the different semantic structures defined in MPEG-7. The Boemie ontology, due its well-founded semantics, forms a fitting candidate, while facilitating alignment with other less rigorous ontologies.

Finally, taking into account the importance of bringing in the SW the already existing MPEG-7 based metadata, the approaches taken in the Rhizomik and DS-MIRF projects are of crucial importance. An immediate advantage of the Rhizomik approach is that it allows for the automatic translation of the complete MPEG-7 intro corresponding OWL DL descriptions. Due to the similar definitions resulting form the one to to one translation, the two ontologies could potentially combined in order to enhance the semantic management and integration of existing MPEG-7 repositories. In order to achieve a higher degree of conceptual clarity and allow for intelligent management at the level of meaning though, the descriptions should be enhanced with respect to ontologies with better defined semantics.

7 Relevant work

Advancing multimedia awareness in the Semantic Web is an undisputable requirement and goal for supporting content management and sharing at a semantic, thus more user oriented, level. Its significance is asserted by the continually increasing number of research projects addressing issues related to multimedia content annotation, management and distribution.[9]

Related efforts addressing the investigation and comparison of different MPEG-7 based ontologies include the activities undertaken within the Multimedia Semantics (MMSEM) Incubator Group,[10] and more specifically the deliverables on *Multimedia Vocabularies on the Semantic Web* [5] and *MPEG-7 and the Semantic Web* [7]. The work presented in this article is in fact the continuation and result of the systematic study of issues initiated within the MMSEM context. The Common Multimedia Ontology Framework[11] consists another relevant initiative, which in addition addresses questions with respect to the types of knowledge that is of interest in multimedia related applications. The newly chartered W3C Media Annotation[12] and Media Fragments[13] WGs, as a continuation of the efforts initiated within MMSEM, further manifest the strong emphasis placed upon achieving cross community multimedia data integration.

[9] http://ontoworld.org/wiki/KWTR:_multimedia

[10] http://www.w3.org/2005/Incubator/mmsem/

[11] http://www.acemedia.org/aceMedia/reference/multimedia_ontology/

[12] http://www.w3.org/2008/01/media-annotations-wg.html

[13] http://www.w3.org/2008/WebVideo/Fragments/

Finally, in [36], four MPEG-7 based ontologies are discussed in terms of coverage, scalability of generated metadata, and the methodology proposed for integration with domain ontologies. Although providing interesting comparison, the it does not examine closely the different modelling choices and their implications on interoperability, nor the complementary roles served and how they could be reconciled.

8 Conclusions

In this article, we presented a systematic survey of the state of the art MPEG-7 based ontologies, comparing them across the two main annotation dimensions prevailing the literature, i.e. content structure descriptions and linking with domain ontologies. Through a close examination of the undertaken modelling choices, we highlighted implications on the interoperability of the resulting metadata, and discussed associations with the intended usage context. The different modelling approaches followed by the individual ontologies illustrate the significance of a formally founded standardised description models, such as the paradigm taken by COMM. Furthermore, the significant challenges ensuing from semantic ambiguities when resolving mappings between different multimedia ontologies and also for practical semantic metadata management services, highlight the significance for conceptual clarity and well-defined semantics, such as the paradigm followed in the Boemie approach. Finally, approaches such as the Rhizomik and the DS-MIRF bring to focus the already existing repositories of MPEG-7 XML metadata that currently remain poorly exploitable and closed to the rest of the Web.

Concluding, two main challenges regarding future directions towards a multimedia-aware Semantic Web besides the aforementioned considerations, relate to the scalability of the representations and to the capturing of contextual information. Both constitute crucial requirements and challenges, considering the interlinked architecture of the Web and the sheer volumes of available content, which urge the discovery, representation and utilisation of the semantic interrelations between pieces of multimedia information found in diverse resources.

Acknowledgement This work was partially supported by the European Commission under contracts FP6-001765 aceMedia, FP6-507482 KnowledgeWeb, and FP6-027538 BOEMIE.

References

1. Arndt R, Troncy R, Staab S, Hardman L, Vacura M (2007) COMM: designing a well-founded multimedia ontology for the web. In: Proc international Semantic Web conference, Busan, Korea, 11–15 November 2007
2. Bechhofer S, van Harmelen F, Hendler J, Horrocks I, McGuinness D, Patel-Schneider P, Stein L (2004) OWL Web ontology language reference. W3C recommendation, 10 February 2004. http://www.w3.org/TR/owl-ref/
3. Berners-Lee T, Hendler J, Lassila O (2003) The Semantic Web. Sci Am 284:34–43
4. Bloehdorn S, Simou N, Tzouvaras V, Petridis K, Handschuh S, Avrithis Y, Kompatsiaris I, Staab S, Strintzis M (2004) Knowledge representation for Semantic multimedia content analysis and reasoning. In: Proc of European workshop on the integration of knowledge, semantics and digital media technology (EWIMT), London, UK, 25–26 November 2007
5. Boll S, Burger T, Celma O, Halaschek-Wiener C, Mannens E, Troncy R (2007) Multimedia vocabularies on the Semantic Web. In: W3C incubator group report, 24 July 2007

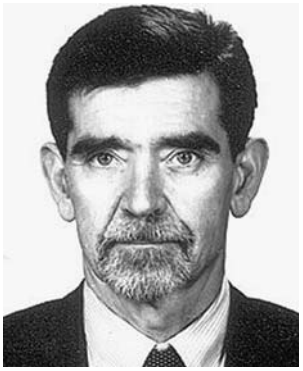
6. Brickley D, Guha R (2004) RDF vocabulary description language 1.0 - RDF schema. W3C recommendation, 10 February 2004. http://www.w3.org/TR/rdf-schema/
7. Celma O, Dasiopoulou S, Hausenblas M, Little S, Tsinaraki C, Troncy R (2007) Mpeg-7 and the Semantic Web. In: W3C incubator group editor's draft, 14 August 2007
8. Dasiopoulou S, Tzouvaras V, Kompatsiaris I, Strintzis M (2007) Capturing MPEG-7 semantics. In: Proc 2nd international conference on metadata and semantics (MTSR), Corfu, Greece
9. Dasiopoulou S, Dalakleidi K, Tzouvaras V, Kompatsiaris I (2007) D3.4 - mulitmedia content and descriptor ontologies—version 2. Boemie Technical Report
10. Dasiopoulou S, Saathoff C, Mylonas P, Avrithis Y, Kompatsiaris Y, Staab S, Strintzis M (2008) Introducing context and reasoning in visual content analysis: an ontology-based framework. In: Hobson PM, Kompatsiaris Y (eds) Semantic multimedia and ontologies: theory and applications. Springer, New York
11. Dasiopoulou S, Dalakleidi K, Eleftherohorinou H, Mailis T, Tzouvaras V, Papastathis V, Kompatsiaris I, Avrithis Y (2009) D3.1 - mulitmedia content and descriptor ontologies—version 1. BOEMIE Technical Report
12. Dublin Core Metadata Initiative (2008) Dublin Core Metadata Element Set, Version 1.1. DCMI recommendation. http://dublincore.org/documents/dces/
13. Gangemi A (2005) Ontology design patterns for Semantic Web content. In: 4th international Semantic Web conference (ISWC), Galway, Ireland, 6–10 November 2005, pp 262–276
14. Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L (2002) Sweetening ontologies with DOLCE. In: 13th international conference on knowledge engineering and knowledge management (EKAW), Siguenza, Spain, 1–4 October, pp 166–181
15. García R, Celma O (2005) Semantic integration and retrieval of multimedia metadata. In: Proc international Semantic Web conference (ISWC), Galway, Ireland
16. García R, Gil R (2007) Facilitating business interoperability from the Semantic Web. In: Proc 10th international conference on business information systems (BIS), Poznan, Poland, pp 220–232
17. García R, Gil R, Delgado J (2007) A web ontologies framework for digital rights management. Artif Intell Law 15(2):137–154
18. Hollink L, Little S, Hunter J (2005) Evaluating the application of semantic inferencing rules to image annotation. In: 3rd international conference on knowledge capture (K-CAP), Banff, Alberta, Canada, pp 91–98
19. Hunter J (2001) Adding multimedia to the Semantic Web: building an MPEG-7 ontology. In: Proc the first Semantic Web working symposium, SWWS'01, Stanford University, California, USA
20. Hunter J, Drennan J, Little S (2004) Realizing the hydrogen economy through Semantic Web technologies. IEEE Intell Syst J—Special issue on eScience 19:40–47
21. Karkaletsis V, Paliouras G, Spyropoulos C (2005) A bootstrapping approach to knowledge acquisition from multimedia content with ontology evolution. In: Proc international conference on adaptive knowledge representation and reasoning (AKRR), Helsinki, Finland, pp 98–105
22. Lagoze C, Hunter J (2001) The ABC ontology and model. J Digit Inf 2(2)
23. Little S, Hunter J (2004) Rules-by-example—a novel approach to semantic indexing and querying of images. In: International Semantic Web conference, pp 534–548
24. Martínez J (2002) MPEG-7: overview of MPEG-7 description tools, part 2. IEEE Multimed 9(3):83–93
25. MPEG-7 (2001) Multimedia content descritpion interface. Standard no. iso/iec 15938
26. MPEG-7 Visual (2001) ISO/IEC 15938-3/FDIS information technology—multimedia content description interface - part 3 visual. ISO/IEC JTC 1/SC 29/WG 11/N4358, Sidney
27. MPEG-7 Audio (2001) ISO/IEC 15938-4:2001(E)/FDIS information technology—multimedia content description interface—part 4 audio. ISO/IEC JTC 1/SC 29/WG 11/N4224, Sidney
28. MPEG-7 MDS (2001) ISO/IEC 15938-5/FDIS information technology—multimedia content description interface—part 5 multimedia description schemes. ISO/IEC JTC 1/SC 29/WG 11/N4242, Singapore
29. Nack F, van Ossenbruggen J, Hardman L (2005) That obscure object of desire: multimedia metadata on the Web, part 2. IEEE MultiMed 12(1):54–63
30. Niles I, Pease A (2001) Towards a standard upper ontology. In: 2nd international conference on formal ontology in information systems (FOIS), Ogunquit, Maine, USA, 17–19 October, pp 2–9
31. Oberle D, Ankolekar A, Hitzler P, Cimiano P, Sintek M, Kiesel M, Mougouie B, Baumann S, Vembu S, Romanelli M (2007) DOLCE ergo SUMO: on foundational and domain models in the SmartWeb integrated ontology (SWIntO). J Web Sem 5(3):156–174

32. Petridis K, Bloehdorn S, Saathoff C, Simou N, Dasiopoulou S, Tzouvaras V, Handschuh S, Avrithis Y, Kompatsiaris I, Staab S (2006) Knowledge representation and semantic annotation of multimedia content. IEE Proc Vis Image Signal Process 153:255–262
33. Romanelli M, Buitelaar P, Sintek M (2007) Modeling linguistic facets of multimedia content for semantic annotation. In: 2nd international conference on semantic and digital media technologies (SAMT), Genova, Italy, pp 240–251
34. Simou N, Saathoff C, Dasiopoulou S, Spyrou E, Voisine N, Tzouvaras V, Kompatsiaris I, Avrithis Y, Staab S (2005) An ontology infrastructure for multimedia reasoning. In: Proc international workshop on very low bitrate video coding (VLBV 2005), Sardinia, Italy
35. SMIL 1.0 (1998) Synchronized multimedia integration language specification. W3C recommendation, 15 June. http://www.w3.org/TR/REC-smil/
36. Troncy R, Celma O, Little S, GarciaGarcía R, Tsinaraki C (2007) Mpeg-7 based multimedia ontologies: interoperability support or interoperability issue? In: 1st workshop on multimedia annotation and retrieval enabled by shared ontologies (MARESO), Genova, Italy, pp 2–16
37. Tsinaraki C, Christodoulakis S (2007) Interoperability of XML schema applications with OWL domain knowledge and Semantic Web tools. In: On the move to meaningful internet systems (OTM), confederated international conferences, Vilamoura, Portugal, pp 850–869
38. Tsinaraki C, Polydoros P, Christodoulakis S (2004) Integration of OWL ontologies in MPEG-7 and TV-anytime compliant semantic indexing. In: 16th international conference on advanced information systems engineering (CAiSE), Riga, Latvia, 7–11 June 2004, pp 398–413
39. Tsinaraki C, Polydoros P, Christodoulakis S (2007) Interoperability support between MPEG-7/21 and OWL in DS-MIRF. IEEE Trans Knowl Data Eng 19(2):219–232
40. van Ossenbruggen J, Nack F, Hardman L (2004) That obscure object of desire: multimedia metadata on the Web, part 1. IEEE MultiMed 11(4):38–48
41. Vembue S, Kiesel M, Sintek M, Bauman S (2006) Towards bridging the semantic gap in multimedia annotation and retrieval. In: Proc workshop on Semantic Web annotations for multimedia (SWAMM), Edinburgh, Scotland

Stamatia Dasiopoulou received her Diploma degree in Electronic and Computer Engineering at Technical University of Crete, Hania, Greece, in 2003, and defened her PhD thesis "Extraction and Representation of Visual Content Semantics" at the Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece, in September 2008. She is currently working as a PostDoctoral Research Fellow with the Multimedia Knowledge Laboratory at Informatics and Telematics Institute, Thessaloniki, Greece. Her research interests include Semantic Multimedia Analysis, Multimedia Ontologies and MPEG-7, Knowledhe Representation and Reasoning in Multimedia Understanding.

Vassilis Tzouvaras received the Diploma Degree in Electronics Systems Engineering, Department of Electrinics Systems, University of Shefield, UK in 1998, Master degree in Automatic Control, Department of Automatic Control and Systems Engineering, University of Shefield, UK, in 1999, and PhD degree from Department of Electrical and Computer Engineering, National Technical University of Athens, Greece in 2005. His research interests include Computer Vision, Neural Networks, Fuzzy Sets, Multimedia Semantics, Ontological Representation, Semantic Web.

Ioannis Kompatsiaris received the Diploma degree in electrical engineering and the Ph.D. degree in 3-D model based image sequence coding from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 1996 and 2001, respectively. He is a Senior Researcher with the Informatics and Telematics Institute, Thessaloniki and currently he is leading the Multimedia Knowledge Laboratory. Prior to his current position, he was a Leading Researcher on 2-D and 3-D Imaging at AUTH. His research interests include multimedia content processing, multimodal techniques, multimedia and the Semantic Web, multimedia analysis and annotation ontologies, knowledge-based, context aware inference for semantic multimedia analysis, semantic metadata representation, semantic adaptation, personalization and retrieval, MPEG-4 and MPEG-7 standards. His involvement with those research areas has led to the co-authoring of 6 book chapters, 20 papers in refereed journals and more than 60 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. Since 1996, he has been involved in more than 15 projects in Greece, funded by the EC, and the Greek Ministry of Research and Technology. Ioannis Kompatsiaris is an IEEE member and a member of the Technical Chamber of Greece.



Michael G. Strintzis received the Diploma in Electrical Engineering from the National Technical University of Athens, Athens, Greece in 1967, and the M.A. and Ph.D. degrees in Electrical Engineering from Princeton University, Princeton, N.J. in 1969 and 1970, respectively. He then joined the Electrical Engineering Department at the University of Pittsburgh, Pittsburgh, Pa., where he served as Assistant (1970–1976) and Associate (1976–1980) Professor. Since 1980 he is Professor of Electrical and Computer Engineering at the University of Thessaloniki, and since 1999 Director of ITI-CERTH. Since 1999 he serves as an Associate Editor of the IEEE Trans. on Circuits and Systems for Video Technology. His current research interests include 2D and 3D Image Coding, Image Processing, Biomedical Signal and Image Processing and DVD and Internet data authentication and copy protection. In 1984, Dr. Strintzis was awarded one of the Centennial Medals of the IEEE.