Multimed Tools Appl
DOI 10.1007/s11042-009-0397-2

VAMP: a service for validating MPEG-7 descriptions w.r.t. to formal profile definitions

Raphaël Troncy · Werner Bailer ·
Martin Höffernig · Michael Hausenblas

© Springer Science + Business Media, LLC 2009

Abstract MPEG-7 can be used to create complex and comprehensive metadata descriptions of multimedia content. Since MPEG-7 is defined in terms of an XML schema, the semantics of its elements has no formal grounding. In addition, certain features can be described in multiple ways. MPEG-7 profiles are subsets of the standard that apply to specific application areas and that aim to reduce this syntactic variability, but they still lack formal semantics. We propose an approach for expressing the semantics explicitly by formalizing the constraints of various profiles using ontologies, logical rules and ad-hoc programming, thus enabling interoperability and automatic use for MPEG-7 based applications. We have implemented *VAMP*, a full semantic validation service that detects any inconsistencies of the semantic constraints formalized. Another contribution of this paper is an analysis of how MPEG-7 is practically used. We report on experiments about the semantic validity

R. Troncy (✉)
EURECOM, 2229 route des crêtes, 06560 Sophia-Antipolis, France
e-mail: raphael.troncy@eurecom.fr, raphael.troncy@cwi.nl

R. Troncy
CWI, 413 Kruislaan, 1098 SJ Amsterdam, The Netherlands

W. Bailer · M. Höffernig
Joanneum Research Forschungsgesellschaft mbH, Institute of Information Systems,
Steyrergasse 17, 8010 Graz, Austria

W. Bailer
e-mail: werner.bailer@joanneum.at

M. Höffernig
e-mail: martin.hoeffernig@joanneum.at

M. Hausenblas
Digital Enterprise Research Institute, National University of Ireland,
IDA Business Park, Lower Dangan, Galway, Ireland
e-mail: michael.hausenblas@deri.org

Published online: 17 October 2009

of MPEG-7 descriptions produced by numerous tools and projects and we categorize the most common errors found.

Keywords VAMP · MPEG-7 semantic validation · Semantic web application · MPEG-7 profile ontology

1 Introduction

The amount of multimedia data being produced, processed and consumed is growing, as is the number of applications dealing with multimedia content. In many of these applications, metadata descriptions of the content are important. MPEG-7 [14], formally named Multimedia Content Description Interface, is designed as a standard for representing these descriptions in a broad range of applications. In order to cover diverse requirement scenarios [19], many *descriptors* and *description schemes*, as well as the relationships between them, have been defined. The descriptors and description schemes are together referred to as *description tools*, and a *description* is a particular instantiation of these. There are description tools for diverse types of annotations on different semantic levels, ranging from very low-level features, such as visual (e.g. texture, camera motion) or audio (e.g. spectrum, harmonicity), to more abstract descriptions (e.g. agent, location, event).

The flexibility of MPEG-7 is based on allowing descriptions to be associated with arbitrary multimedia segments or regions, at any level of granularity, using different levels of abstraction. The downside of the breadth targeted by MPEG-7 is its complexity and its fuzziness [3, 23, 25]. For example, very different syntactic variations may be used in multimedia descriptions with the same intended semantics, while remaining valid MPEG-7 descriptions. Given that the standard does not provide a formal semantics for these descriptions, this syntax variability causes serious interoperability issues for multimedia processing and exchange, for example on the web.

To reduce this syntax variability, MPEG-7 has introduced the notion of *profiles*, that also exist in earlier MPEG standards, to constrain the way multimedia descriptions should be represented for particular applications. Profiles are therefore a way of reducing the complexity of MPEG-7 (i.e. only a subset of the whole standard can be used) and of solving some interoperability issues (i.e. English guidelines are provided on how the descriptors should be used and combined). However, these additional constraints are only represented with XML Schema [26], and, for most of them, cannot be automatically checked for consistency by XML processing tools. In other words, profiles provide only very limited control over the semantics of the MPEG-7 descriptions [9, 17, 21]. Because of this lack of formal semantics, the resulting interoperability problems prevent an effective use of MPEG-7 as a language for describing multimedia.

In this paper, we present VAMP,[1] a semantic VAlidation service for MPEG-7 Profiles. VAMP generalizes the method we proposed for the single DAVP profile [22] by formalizing how MPEG-7 descriptors should be used in commonly-used

[1] VAMP is available as a web application at http://vamp.joanneum.at and as a web service.

profiles. In contrast to other work [1, 6, 9, 24], we do not intend to completely map the MPEG-7 description tools onto an OWL ontology [5, 12], but rather use Semantic Web technologies to represent those MPEG-7 semantic constraints defined in natural language that cannot be expressed using XML Schema. We do not modify or extend the intended semantics of the description tools, but rather capture and formalize it. We have also gathered and analyzed numerous MPEG-7 descriptions generated by various tools. We report in this paper on how semantically valid these descriptions are and we provide a categorization of the most common interoperability problems we found.

The paper is organized as follows. In the next section, we briefly introduce the notion of MPEG-7 profiles and we analyze several MPEG-7 descriptions generated by various tools. In Section 3, we provide a categorization of the most common interoperability problems encountered. In Section 4, we present the VAMP service and we detail how the MPEG-7 profiles can be formalized, building first an OWL ontology and rules capturing the semantic constraints, and developing tools converting the XML-based MPEG-7 descriptions to RDF triples. In Section 5, we compare our approach with other attempts to formalize the MPEG-7 knowledge and we discuss the scope of our methodology before concluding the paper (Section 6).

2 MPEG-7 usage analysis

The MPEG-7 XML Schema defines numerous elements and types, as well as rules for their valid combinations. The standard, however, allows the specification of different descriptions with equivalent semantics. This raises interoperability problems when exchanging MPEG-7 descriptions since applications may use the standard differently. For example, the same decomposition of a video into shots and key frames can be represented by multiple MPEG-7 descriptions [22]. Hence, it is perfectly valid for the same video file to be either described by a `VideoType` under the root `VideoSegment`, or to be described by a `AudiovisualType` and be further decomposed into `VideoSegment` and `AudioSegment`. It has to be noted that the problem comes from a lack of specification, and not from a flaw in one of the applications. Therefore, any implementation would be more or less “lossy”, except if it covers all possible syntactic variations and combinations allowed by the standard, which is not feasible in practice.

This problem has been recognized by both the MPEG working group and the various tools that partially support the standard. Profiles have thus been proposed as a possible solution. In the following, we first introduce the notion of *profiles* (Section 2.1) and we then show how several multimedia annotation tools (Section 2.2) address this interoperability problem by reducing and further constraining the MPEG-7 description tools.

2.1 Profiling MPEG-7

The specification of a profile consists of three parts, namely [15]: i) *description tool selection*, i.e. the definition of the subset of description tools to be included in the profile, ii) *description tool constraints*, i.e. definition of constraints on the description tools such as restrictions on the cardinality of elements or on the use of attributes, and iii) *semantic constraints* that further describe the use of the description tools in the context of the profile.

The first two parts of a profile specification are used to address the *complexity* problem, that is, the complexity of a description that can be measured by its size or the number of descriptors used. Limiting the number of descriptors and description schemes (either by excluding elements or constraining their cardinality) reduces this complexity. Both the selection and the usage constraints of the description tools are specified using the MPEG-7 DDL. They result in a specific and more constrained XML Schema. The third part of a profile specification tackles the *interoperability* problem. Semantic constraints are expressed in natural language to clarify the ambiguities associated with the use of the remaining description tools selected in the first two parts. This informal specification of the constraints, however, prevents an automated process from checking the correct use of MPEG-7 profiles for describing multimedia content.

Six MPEG-7 profiles are commonly used: the first three have been defined in Part 9 of the standard[2] [15], and we consider three other "de-facto" profiles, not (yet) standardized, but used by the multimedia community and partly based on standard profiles:

Simple Metadata Profile (SMP) describes single instances or collections of multimedia content as complete entities or clips with textual metadata only and no spatial decomposition. The motivation of this profile is to support simple metadata tagging similar to ID3[3] for music and EXIF[4] for images, and to support mobile applications such as 3GPP.[5] A partial mapping from these vocabularies to SMP has been specified.

User Description Profile (UDP) consists of tools for describing the personal preferences and usage patterns of users of multimedia content in order to enable automatic discovery, selection, personalization and recommendation of multimedia content. This profile contains all MPEG-7 description tools that were adopted by the TV-Anytime Forum, and are referenced by the TV-Anytime Metadata specification [20].

Core Description Profile (CDP) consists of tools for describing general multimedia content such as images, videos, audio and collections using the top-level types defined in Part 5 of the standard. A typical use of this profile is the description of the structural and semantic aspects of video content of a TV program and its corresponding materials. This includes managing the media materials, distributing them and archiving them. Just as the two previous profiles, it does not include the visual and audio descriptors defined in Parts 3 and 4 of MPEG-7.

Detailed Audio-Visual Profile (DAVP) describes single multimedia content entities, based on a comprehensive structural description of the content and including a subset of Part 5 (MDS) as well as all audio and visual low-level feature descriptors (Parts 3 and 4).

[2] Five other profiles are actually discussed in [15] but have been later merged or withdrawn.

[3] http://www.id3.org/

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

[5] http://www.3gpp.org/

Table 1 The number of MPEG-7 descriptors and semantic constraints specified in each profile

Profile	Descriptors	Semantic constraints
Simple Metadata Profile (SMP)	45	6 + 0
User Description Profile (UDP)	102	8 + 0
Core Description Profile (CDP)	153	27 + 2
Detailed Audio-Visual Profile (DAVP)	274	35 + 50
TRECVID Profile	20	4 + 9
NHK Metadata Production Framework	193	29 + 32

TRECVID Profile is used to represent master shot boundary reference data of the TREC Video Retrieval Evaluation.[6] It uses a subset of MPEG-7 to describe the shot structure of a video and the key frames representing each shot. As no official XML Schema formalization of the profile is available, we have defined one based on the available TRECVID MPEG-7 documents.[7]

NHK Metadata Production Framework (MPF) data model is an industrial application of the Core Description Profile (CDP) [16]. The authors address the complexity and ambiguity problems of MPEG-7 proposing a metadata model that further restricts CDP by excluding some elements and reducing the cardinality of others. The new version also allows the use of the visual and audio descriptors defined in Parts 3 and 4. The definition of the data model defines a number of semantic constraints for the structure of the description as well as several syntactic and semantic constraints on different elements of the description (called "operational rules").

The six profiles discussed above put different emphasis on the *complexity* and *interoperability* problems previously mentioned. For each profile, we have counted the number of descriptors and we have evaluated the number of semantic constraints it contains (Table 1). More precisely, for each descriptor included in a profile, we looked at its informal semantics written in English in the standard, and we examine the constraints that cannot be represented with XML Schema. Therefore, our evaluation considers both the original MPEG-7 constraints and those specified additionally in the profiles. We observe that the standardized profiles aim at complexity reduction and hence significantly reduce the included set of allowed descriptors (with respect to the 1200 MPEG-7 elements) while defining few semantic constraints. In contrast, DAVP excludes some descriptors such as the user preferences or the collection description schemes, but keeps most of the others [3]. The focus is on the definition of the semantic constraints for the remaining descriptors included in the profile. Similarly, the TRECVID profile has reduced the set of descriptors to those applicable to its specific application area and agreed upon the use of these descriptors. The NHK MPF specification builds on CDP and thus inherits its constraints, but it adds the description tools from Parts 3 and 4 and defines a number of additional constraints on the descriptors included in CDP.

[6] http://www-nlpir.nist.gov/projects/trecvid/

[7] http://vamp.joanneum.at/data/xsd/trecvid_xsd/

2.2 Gathering MPEG-7 descriptions

The W3C Multimedia Semantics Incubator Group maintains a comprehensive list[8] of tools that can generate MPEG-7 descriptions. These tools do not necessarily comply with a profile, but they also try to address the interoperability problem by further constraining the subset of descriptors they support. This complexity reduction, however, comes often with the price of having hard-coded constraints instead of explicit semantics. We present a selection of these tools, categorized according to their predominant media type (image, audio and video), although some of them can handle multiple media.

2.2.1 Image related tools

Caliph & Emir[9] is a semi-automatic annotation tool for images that supports free text and graph-based semantic annotations as well as a number of visual feature extractors. Furthermore, pre-existing metadata, such as EXIF or IPTC tags inside images, is converted into MPEG-7 following the mapping rules given in the SMP profile.

The M-OntoMat-Annotizer[10] supports the manual annotation of still image regions, linking RDF(S) domain specific ontologies to low-level MPEG-7 visual descriptors. The semantics of these visual descriptors is formalized in a Visual Descriptor Ontology (VDO) represented in RDFS [2].

2.2.2 Audio related tools

The MPEG-7 Audio Analyzer[11] implements all 17 low-level audio descriptors defined in Part 4, while the MPEG-7 Spoken Content Demonstrator[12] generates the output of an Automatic Speech Recognition (ASR) system using the `SpokenContent DS`, which is composed of around 20 descriptors.

The MPEG-7 Audio Encoder[13] allows also to extract all the audio descriptors, but it further constrains their use in two new XML Schemas.

2.2.3 Video related tools

IBM VideoAnnex[14] is a semi-automatic annotation tool for videos that generates temporal shot segmentation and allows the spatial decomposition of key frames. The annotations make use of controlled vocabularies defined using the `ClassificationScheme DS` (see Part 5 of [14]).

[8] http://www.w3.org/2005/Incubator/mmsem/wiki/Tools_and_Resources

[9] http://www.semanticmetadata.net/features/

[10] http://www.acemedia.org/aceMedia/results/software/m-ontomat-annotizer.html

[11] http://mpeg7lld.nue.tu-berlin.de/

[12] http://mpeg7spkc.nue.tu-berlin.de/

[13] http://mpeg7audioenc.sourceforge.net/

[14] http://www.research.ibm.com/VideoAnnEx

Frameline 47[15] uses an advanced content schema based on MPEG-7 so as to be able to annotate either entire video files or segments and groups of segments from within video files.

Muvino[16] is a very simple tool for manually annotating videos (free text annotation and keyword based). It supports some general metadata about the video, the temporal decomposition into segments and some semantic descriptors such as place and time.

The Metadata Editor[17] developed by NHK is an application for producing and storing metadata that conforms to the MPF specifications. The application directly implements the semantics constraints of this profile.

2.3 Summary

We have collected a large set of sample descriptions in order to analyze how MPEG-7 is used in practice, and offered them to the multimedia community in the MPEG-7 Specification Repository[18] available at http://media.cwi.nl/mpeg7/wiki. These examples cover a broad range of applications and use different subsets of MPEG-7 descriptors. Profiles are sometimes used (and even further constrained) or could have been specified from the scope of the application. The interoperability problems, however, cannot be solved by just extending the XML schema and the semantics is often directly hard-coded in the tools. We argue that true interoperability can be obtained if the semantics is made explicit and can be formally checked for consistency.

Some tools generate errors. For example, the IBM VideoAnnex tool automatically produces shot lists of videos. For some video clips the tool produces shot segments with a negative duration, or overlapping segments, even though the `overlap` attribute of the `TemporalDecomposition` has the value `false`. The resulting description will validate according to the XML Schema (of MPEG-7 or one of the profiles) but will not be semantically valid. We have analyzed from these MPEG-7 descriptions the possible errors and identified the semantic constraints that need to be formalized. We detail these errors in the next section and present how the interoperability problem is solved in VAMP.

3 Interoperability problems

In this section, we summarize the errors that we found. We classify them in four categories: the inconsistencies related to the structural information (Section 3.1), the temporal information (Section 3.2), the media information (Section 3.3), and the semantic information (Section 3.4). All the violations discussed here yield perfectly valid documents with respect to the MPEG-7 XML schema but raise inconsistencies with the semantic constraints that express the intended semantics of the standard.

[15] http://frameline.tv/

[16] http://vitooki.sourceforge.net/components/muvino/code/index.html

[17] http://www.nhk.or.jp/strl/mpf/english/editor.htm

[18] The MPEG-7 Specification Repository is a semantic wiki for sharing information relevant for practical work with MPEG-7, e.g. specifications, examples, tools, events, projects, etc.

3.1 Structural-related violations

Many semantic constraints in profile definitions are related to the resulting structure of the descriptions and to the semantics implied by this structure. Such constraints can be typically found in the DAVP and TRECVID profiles and in the additional semantic constraints defined by NHK MPF on top of CDP, but they cannot be expressed in XML Schema:

Decomposition hierarchies such as a video being decomposed into shots and then shots into key frames.

Restrictions on decompositions like allowing several temporal decompositions while only one corresponds to a shot list specified by a criteria attribute.

Misuse of description tools for some segments since some description tools are only permitted on segments corresponding to the entire content or representing a certain type of element in the decomposition hierarchy.

Typical violations of these constraints are misplaced segments in decompositions, repeated and missing segments or decompositions, and missing description tools while they are required or occurring while they are prohibited according to the profile specification.

3.2 Temporal-related violations

The representation of time is an essential component for media having a temporal dimension. MPEG-7, however, defines only a simple syntactic pattern for representing the time points and the time duration. We present common inconsistencies underlying this representation as well as the possible misuse of the temporal decomposition descriptors. We advocate then an alternative time representation.

3.2.1 Common violations

The ISO 8601 standard is generally considered as the reference "specification of the representation of dates in the proleptic Gregorian calendar[19] and times and representations of periods of time" [11]. The corresponding datatypes in XML Schema use lexical formats inspired by the ISO standard and include some deviations such as an optional minus sign in the lexical representation, the possibility of having more than 9999 years or the inclusion of a time zone [26]. Unfortunately, these datatypes are not used in MPEG-7, which instead, redefines a simple pattern format for the media time point:

```
<simpleType name="mediaTimePointType">
  <restriction base="mpeg7:basicTimePointType">
    <pattern value="(\-?\d+(\-\d{2}(\-\d{2})?)?)?(T\d{2}(:\d{2}(:\d{2}
                      (:\d+)?)?)?)?(F\d+)?"/>
  </restriction>
</simpleType>
```

[19] The proleptic Gregorian calendar includes dates prior to 1582 (the year it came into use as an ecclesiastical calendar).

and for the media duration:

```
<simpleType name="mediaDurationType">
  <restriction base="mpeg7:basicDurationType">
    <pattern value="\-?P(\d+D)?(T(\d+H)?(\d+M)?(\d+S)?(\d+N)?)?(\d+F)?"/>
  </restriction>
</simpleType>
```

Based on this decision, the following inconsistencies can be observed:

Invalid time specification. MPEG-7 introduces different new lexical patterns to represent media times and real-world dates and times. The pattern definition allows the specification of invalid dates and times. For example, 31st of February would be a valid date according to the time point pattern shown above. Another shortcoming deals with the frame precision in the media time pattern: for example `T00:01:23:27F25` would be a valid time point whereas it points to the fraction 27 of 25 that is impossible to compute. Similarly, a fraction rate of 0 cannot be computed but could still be represented with this pattern.

Negative segment duration. MPEG-7 segments are described by a start time point and a duration. The optional minus sign of the patterns allows negative duration for segments in a temporal decomposition while this would make no sense.

Inconsistent temporal decomposition. A temporal decomposition of a segment into subsegments is only meaningful if the time range filled by each of the subsegments is at most the time range of the segment being decomposed, i.e. a part of a temporal segment cannot start before or end after its parent segment.

Gap and overlap. A temporal decomposition can be qualified whether the subsegments in the decomposition overlap or have gaps between them. These properties are specified with the `gap` and `overlap` attributes of the decomposition that have a `true/false` value. There is, however, no mechanism to check whether the actual time description of the segments conforms to the value of this boolean attribute or not.

Formalizing the representation of dates and times, for example using OWL-Time [7] solves some of these problems. The 8-ary predicate `duration` is converted into eight binary relations, which are more convenient for description logic-based markup languages such as OWL, so that the consistency of the time specification can be checked. However, OWL-Time would not helped to check if the value of the `gap` and `overlap` attributes match with the actual timecodes of a temporal decomposition.

3.2.2 Analogy with space representation

Similar to the temporal decomposition, the spatial and the spatio-temporal decompositions suffer from the same limitations in MPEG-7. For example, if a region of an image is decomposed into subregions, the subregions must lie inside the parent region. The violations related to the values of the `gap` and `overlap` attributes can thus also be raised. Consistency checking is, however, much more difficult to implement than for the time ranges due to the two-dimensional nature of the regions.

3.3 Media information-related violations

The description of information about properties of the media can be specified at multiple places in MPEG-7. While the presence and cardinality of the elements can be controlled using XML Schema, the semantics between the global media information and the actual description can mismatch. The following inconsistencies can thus be observed:

Inconsistent media content types. The `Content` element in `MediaFormat` is used to describe the content type of the medium being described (e.g. image, video), using a reference to a classification scheme. The same information is contained in the type of the `MultimediaContent` element of the description but these two values can mismatch. For example, the `xsi:type="ImageType"` specifies the multimedia content being described, but the `MediaFormat` could be stated as audio.

Inconsistent modality information. The `MediaProfile` describes the visual and audio encoding (e.g. a master quality and a low resolution preview), or each stream if several streams in different encoding are available. This information must also match the content type, but again, there is no way to check that the values are consistent. For example, different modalities can be present in the structural description (e.g. one video and two audio channels) even though the media information contains contradicting information about the modalities (e.g. states that the content is mono-audio).

3.4 Classification scheme-related violations

An MPEG-7 `ClassificationScheme` is a generic mechanism for defining multilingual and controlled vocabularies. The set of terms and definitions belonging to a scheme is organized in a taxonomy, and is identified by a URI to be further referenced as values for descriptors. Part 5 of the standard already defines some basic classification schemes, e.g. for enumerating the media types, the different encoding, or some TV genres.

The appropriateness of a classification scheme in a certain context is a source of possible violations of the semantic constraints. More precisely, the `ClassificationSchemeBaseType` has two attributes: `uri` which identifies the classification scheme and `domain` which gives a list of XPath expressions containing the MPEG-7 description schemes that can reference the terms of the scheme. A description, however, can contain unforeseen descriptors using terms from this scheme, i.e. the classification scheme does not contain appropriate terms for the context in which it is used. Once a classification scheme is dereferenced, the terms identified might not be retrieved, i.e. there are broken links. A classification scheme can also import other classification schemes which makes the task of resolving the referenced terms more difficult.

The errors detailed in this section cannot be checked with XML Schema validators. Semantic constraints are defined informally in the standard and cannot be processed by automated tools. We therefore propose a method for formalizing these constraints, implemented in the VAMP service.

4 VAMP: a semantic validation service for MPEG-7 descriptions

The violations of the semantic constraints trigger interoperability problems even though the result is perfectly valid MPEG-7 descriptions. In previous work, we have analyzed the semantic constraints of the Detailed Audiovisual Profile (DAVP) and formalized a subset of them [22]. We have addressed the problem of temporal semantic constrains in [8]. Here, we generalize further these approaches to other profiles and we present VAMP, a Semantic Web application for validating the conformance of MPEG-7 documents to the semantics of a given profile (Section 4.1). We show that the formalization of the semantic constraints amounts to explicitly capturing the semantics of a given profile as well as some additional logical rules and ad-hoc programming (Section 4.2). We describe the implementation of the VAMP service, available as a web interface for humans, and as a REST-style web service for agents (Section 4.3). Finally, we provide some statistics of the usage of VAMP which is running for 1 year (Section 4.4).

4.1 General methodology

We propose the following layered approach to validate *semantically* the conformance of MPEG-7 descriptions to a given profile:

XML/syntactic well-formedness: The well-formedness[20] of the input description is verified;

XML/syntactic validity: The XML validity of the input description against the MPEG-7 schema and the selected profile schema is checked, including syntactic validation of patterns defined in MPEG-7 DDL (e.g. for time points and durations);

RDF/semantics constraints: The consistency of the input description with the ontology and logical rules formalizing the semantic constraints of a profile is computed.

Figure 1 depicts these various steps in the VAMP service. We propose to use Semantic Web languages to formalize the semantic constraints when possible, and later inference tools to check the semantic consistency of the descriptions. This is carried out with an appropriate combination of the following languages [10]:

- XML Schema [26] to define the structural constraints, that is, which types are allowed and how they can be combined.
- OWL-DL [5] to formally capture the intended semantics of the descriptors contained in a profile which have semantic constraints and to model the formal representation of temporal segments.
- Horn clauses [4] to express relationships between syntactically different but semantically equivalent descriptors. Horn clauses are also used to perform closed world checks of the descriptors and are created with respect to a profile ontology.
- XSLT to convert MPEG-7 descriptions into RDF depending on a profile ontology. The RDF data asserts the class-membership of particular descriptors

[20] http://www.w3.org/TR/REC-xml/#sec-well-formed

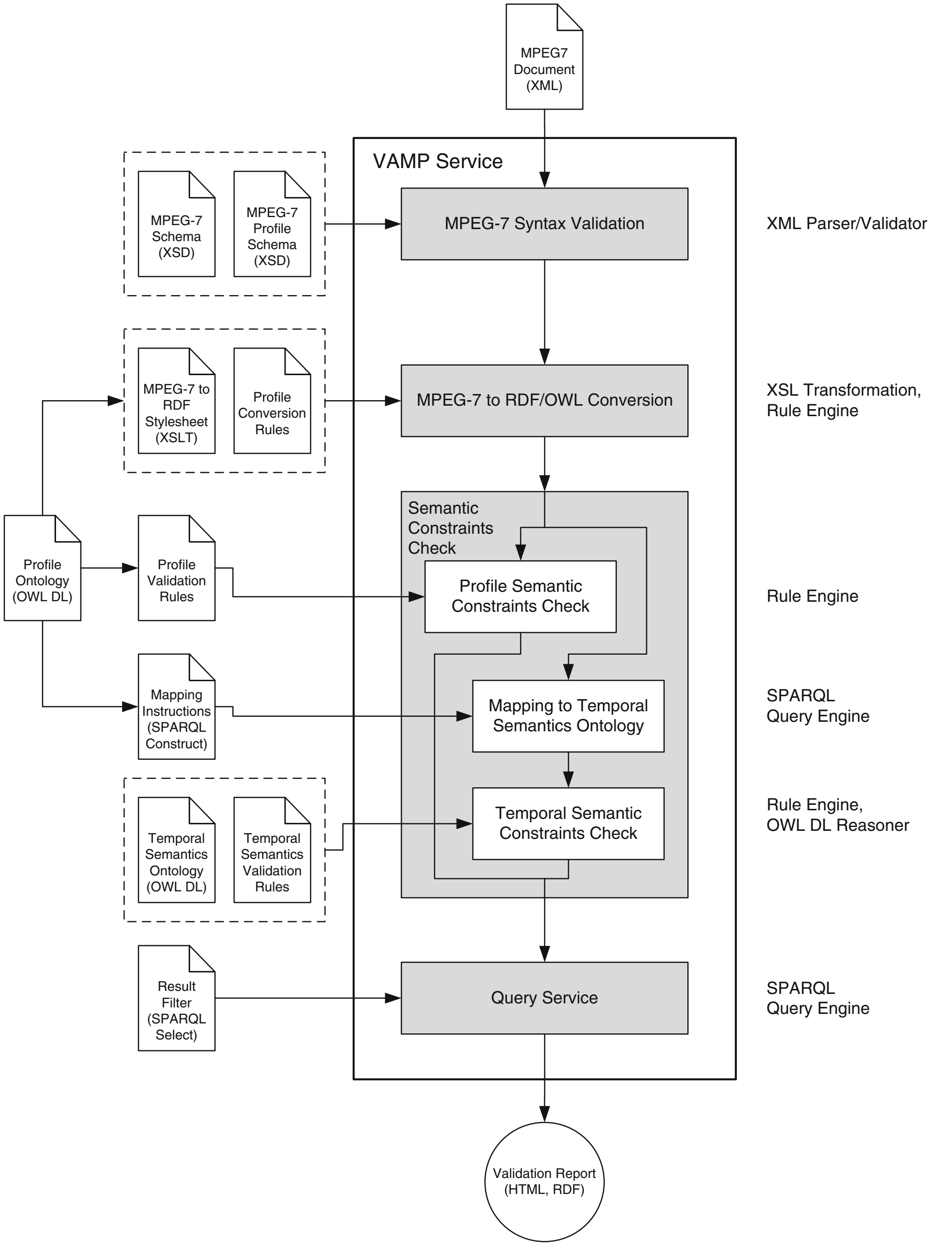

MPEG7 Document (XML)
VAMP Service
MPEG-7 Schema (XSD)
MPEG-7 Profile Schema (XSD)
MPEG-7 Syntax Validation
XML Parser/Validator
MPEG-7 to RDF Stylesheet (XSLT)
Profile Conversion Rules
MPEG-7 to RDF/OWL Conversion
XSL Transformation, Rule Engine
Semantic Constraints Check
Profile Ontology (OWL DL)
Profile Validation Rules
Profile Semantic Constraints Check
Rule Engine
Mapping Instructions (SPARQL Construct)
Mapping to Temporal Semantics Ontology
SPARQL Query Engine
Temporal Semantics Ontology (OWL DL)
Temporal Semantics Validation Rules
Temporal Semantic Constraints Check
Rule Engine, OWL DL Reasoner
Result Filter (SPARQL Select)
Query Service
SPARQL Query Engine
Validation Report (HTML, RDF)

Fig. 1 General architecture of the VAMP service

given their properties. Further classification rules are needed after the XSL transformation to complete the MPEG-7 conversion into RDF.

- SPARQL to map instances from the selected profile ontology to the temporal semantics ontology (construct query) and to retrieve semantic violations (select query).

First, the MPEG-7 input document is checked for syntactic validity against both the MPEG-7 and the selected profile XML schemas. A syntactically valid MPEG-7 input document is a necessary precondition to start the semantic validation. Second, the MPEG-7 description is converted into RDF with respect to an ontology capturing the semantics of the selected profile. In this step, an XSL transformation and additional conversion rules are applied. This results in a set of RDF triples that is the input data for the semantic constraints check, i.e. the validation of profile-specific and temporal semantic constraints. For the validation of the former, only validation rules derived from the profile ontology are applied (cf. Fig. 3 for an example of a profile-specific validation rule). To validate the temporal constraints a mapping from the profile ontology to the (profile-independent) temporal semantics ontology is first needed. For this purpose a SPARQL construct query is used to map instances from the selected profile ontology to the temporal semantics ontology. Then, temporal validation rules are applied and the temporal validation results are classified using a OWL-DL reasoner as described in [8], for example, to determine gap and overlap violations between segments. All possible profile-specific validation violations are flagged by profile validation rules (cf. Fig. 3) while temporal violations are marked by the use of an OWL-DL reasoner after performing temporal validation rules [8]. Finally, these marked violations are reported using a SPARQL select query. In the following, we discuss these steps using the example of a structural profile semantic constraint.

4.2 Formalizing the MPEG-7 semantic constraints

One structural semantic constraint of DAVP is that a decomposition of a shot can only include key frames. This constraint is quite simple but it cannot be checked with XML processing tools (shot and key frames being both of `VideoSegmentType`) and needs to be formalized semantically.

4.2.1 Modeling semantic constraints within an ontology

Figure 2 gives a partial formalization of the class `KeyframesTemporalDecomposition` and the object property `hasSegment` in the OWL

```
Namespace(rdf = <http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Namespace(owl = <http://www.w3.org/2002/07/owl#>)
Namespace(xsd = <http://www.w3.org/2001/XMLSchema#>)
Namespace(rdfs = <http://www.w3.org/2000/01/rdf-schema#>)
Namespace(davp = <http://iis.joanneum.at/mpeg-7/davp/semantics/MPEG7#>)

Class(davp:KeyframesTemporalDecomposition partial
  restriction(davp:hasSegment allValuesFrom(davp:Keyframe))
  ... )

ObjectProperty(davp:hasSegment InverseFunctional
  inverseOf(davp:isSegmentOf)
  domain(davp:Decomposition)
  range(davp:Segment))
```

Fig. 2 Formalization of the class `KeyframesTemporalDecomposition` in OWL DL

Abstract Syntax (OWL-AS) [18]. It starts with the namespace declarations, followed by the definition of the concepts used. The universal restriction in class `KeyframesTemporalDecomposition` defines that any instance of this class can only have `hasSegment` relations to instances of class `Keyframe`.

The XML representation of the description can then be converted into RDF using the ontology capturing the semantics of the profile. The OWL-DL expressivity is, however, insufficient for capturing all the semantic constraints. For example, the boolean values of the `gap` and `overlap` attributes can mismatch their actual truth values based on the actual time points delimiting the segments. Horn clauses [4] and ad-hoc programming are also necessary to check the consistency of such information.

4.2.2 Deriving validation rules from the ontology

Logic programs (LP) [4] is a knowledge representation formalism. The commonly used expressiveness of full LP includes features such as negation-as-failure, priorities and procedural attachments, that are not expressible in First-Order-Logic (FOL). An ordinary logic program is a set of rules each having the form:

$$H \leftarrow B_1 \wedge \ldots \wedge B_m \wedge \sim B_{m+1} \wedge \ldots \wedge \sim B_n$$

where

- H and B_i are *atomic formulae*,
- $\sim$ is a logical connector called *negation as failure*,
- $\leftarrow$ is to be read as *if*, so that the overall rule should be read as “[head] if [body]”,
- and $n \geq m \geq 0$.

The left-hand side of the rule is called the rule’s head (or conclusion/consequent); the right-hand side is called the rules body (or premise/antecedent). Note that no restriction is placed on the arity of the predicates appearing in these atoms. Logical variables, and logical functions (with any arity), may appear unrestrictedly in these atoms.

The logical rule depicted in Fig. 3 is used to detect segments which are not key frames, but part of a temporal decomposition into key frames. If the rule finds

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix davp: <http://iis.joanneum.at/mpeg-7/davp/semantics/MPEG7#> .

[Check_KeyframesTemporalDecomposition_hasSegment_only_Keyframe:

  (?keyframesTemporalDecomposition rdf:type
     davp:KeyframesTemporalDecomposition),
  (?keyframesTemporalDecomposition davp:hasSegment ?segment),
  noValue(?segment rdf:type davp:Keyframe),
  ->
  (?segment davp:hasError
     davp:MisplacedSegmentInKeyframeTemporalDecomposition)
]
```

Fig. 3 Formalization of `KeyframesTemporalDecomposition` with additional Horn clauses

a segment (`?segment`), which is not a key frame (`Keyframe`), but part of temporal decomposition of key frames (`KeyframesTemporalDecomposition`), an error is flagged (`hasError`) and typed (`MisplacedSegmentInKeyframeTemporalDecomposition`) to be further processed in order to give a meaningful explanation of the violation to the end-user.

4.2.3 Semantic constraints and reasoning

Once the semantic constraints have been formalized, they need to be checked for consistency. In contrast to the Semantic Web, VAMP is a closed system. Actually, we assume that all information needed to validate an MPEG-7 description is available: in the MPEG-7 input document itself, in the profile-dependent transformation, in the semantic constraints profile ontology and in the semantic constraints profile rule base. The semantic constraints profile ontology is used as an indirect input: the ontology is only the basis for the transformation instructions (XSLT stylesheet) and the rule base. We are aware of the possibility of using DL-safe rules [13], however, our approach is to work with OWL-DL and rules in an independent manner. This is the direction the evolution of the VAMP service has taken.

4.3 Implementation

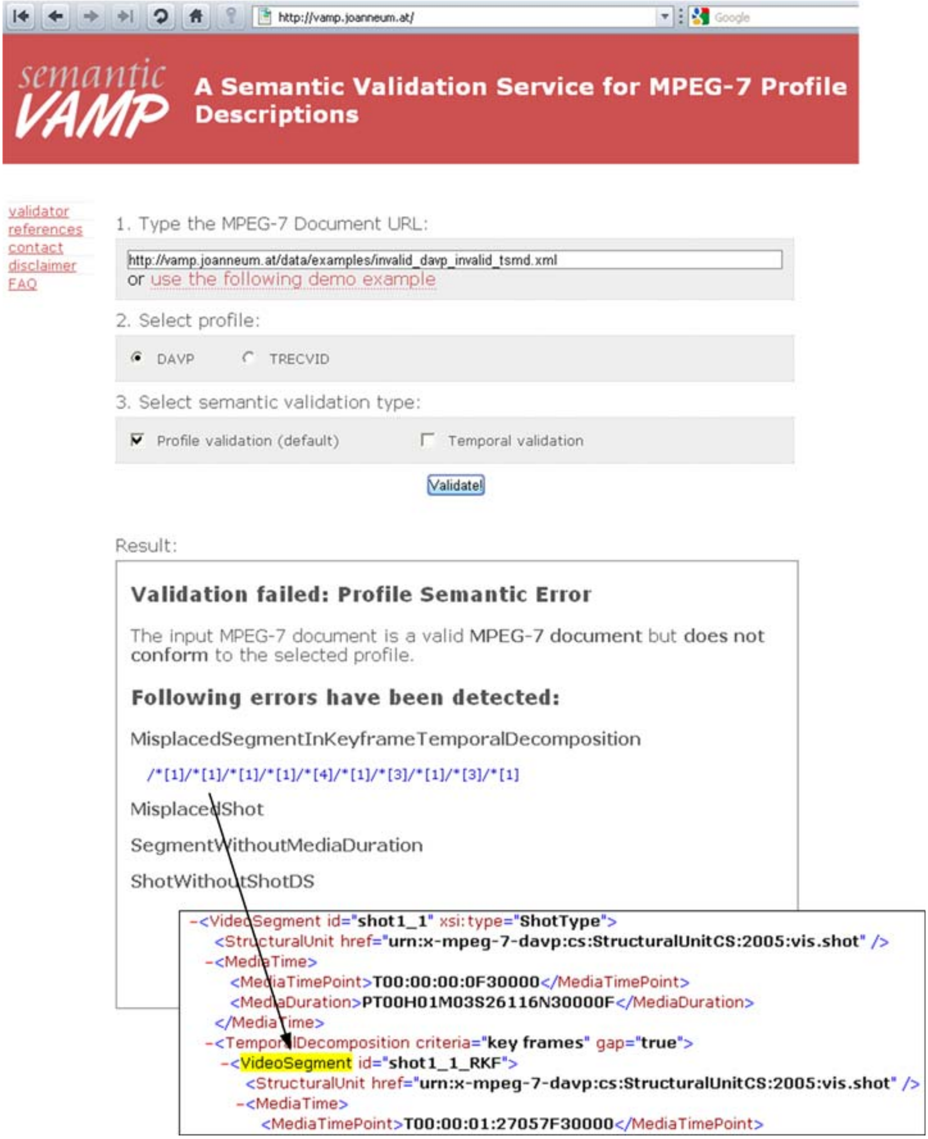
This methodology has been implemented in the VAMP service, available as a web interface for humans and as a REST-style web service for agents. For the RDF processing, Jena 2.4[21] is used. The validation of the semantic constraints is done by the Jena general purpose rule engine.[22] Jena rules provide for a sound, and integrated reasoning system that allows for both forward and backward reasoning. For the classification of the temporal validation results, the OWL DL reasoner Pellet[23] is used.

The interface for a human user is the VAMP web interface, depicted in Fig. 4. The web application uses Ajax and Java servlet technologies. First the users enters the URI of the MPEG-7 document to be validated. For the demonstration of VAMP, some demo examples are provided and can be selected alternatively. The next step is to select the MPEG-7 profile, to which the input MPEG-7 description should conform to. Note that the validation of media information-related violations (Section 3.3) and classification scheme-related violations (Section 3.4) is currently not implemented. Then the semantic validation type is selected. Therefore two different semantic validation types are available, which can be combined: profile validation and temporal validation. The `Validate` button provides a meaningful validation report of all detected semantic errors. For each semantic error, the XML elements which caused this error are listed. These XML elements are identified by XPath expressions which enables the direct observation of the error locations in the input MPEG-7 document.

[21] http://jena.sourceforge.net/

[22] http://jena.sourceforge.net/inference/index.html#rules

[23] http://clarkparsia.com/pellet



A Semantic Validation Service for MPEG-7 Profile Descriptions

1. Type the MPEG-7 Document URL:

http://vamp.joanneum.at/data/examples/invalid_davp_invalid_tsmd.xml
or use the following demo example

2. Select profile:

● DAVP ○ TRECVID

3. Select semantic validation type:

☑ Profile validation (default) □ Temporal validation

Result:

Validation failed: Profile Semantic Error

The input MPEG-7 document is a valid MPEG-7 document but does not conform to the selected profile.

Following errors have been detected:

MisplacedSegmentInKeyframeTemporalDecomposition

/*[1]/*[1]/*[1]/*[1]/*[4]/*[1]/*[3]/*[1]/*[3]/*[1]

MisplacedShot

SegmentWithoutMediaDuration

ShotWithoutShotDS

```
-<VideoSegment id="shot1_1" xsi:type="ShotType">
   <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.shot" />
  -<MediaTime>
     <MediaTimePoint>T00:00:00:0F30000</MediaTimePoint>
     <MediaDuration>PT00H01M03S26116N30000F</MediaDuration>
   </MediaTime>
  -<TemporalDecomposition criteria="key frames" gap="true">
    -<VideoSegment id="shot1_1_RKF">
       <StructuralUnit href="urn:x-mpeg-7-davp:cs:StructuralUnitCS:2005:vis.shot" />
      -<MediaTime>
         <MediaTimePoint>T00:00:01:27057F30000</MediaTimePoint>
```

Fig. 4 The VAMP web interface

VAMP is also available as a web service so that the validation functionality can be embedded into any application. We provide a REST-style web service interface for the validation service. Similar to the graphical user interface, the client of the web service provides an input MPEG-7 description to be validated, the profile the description should conform to, and the semantic validation type.

The service can then generate the results of the SPARQL query in two different formats: i) an XML format, which can be easily further processed by XSLT depending on the application's needs; ii) the RDF graph that is built up in the service containing all the instances contained from the document.

4.4 VAMP usage statistics

The VAMP service is online for more than one year. We have analyzed the logs for the last year in order to find out how the service has been used and how valid or erroneous were the documents submitted. In this analysis, we have first excluded the users who come from the organizations that have contributed to the development of the service. We have also excluded the descriptions provided as examples on the VAMP web page.

In total, 476 validations have been performed, originating from 36 different sites. 16% could not be evaluated because the URL provided was invalid or an internal error occurred (probably due to memory constraints). On the remaining documents, 32% were not well-formed XML documents and 46% were not valid with respect to the MPEG-7 XML Schema. 20% were valid MPEG-7 documents but did not conform to the selected profile schema. Interestingly, only 2% passed both the profile XML Schema and the semantic validation.

5 Related work and discussion

Several attempts have been made to map the MPEG-7 description tools onto an OWL ontology,[24] which we present in Section 5.1. We then argue why MPEG-7 and its formal representation should co-exist (Section 5.2). We finally discuss the scope of our approach which goes beyond the validation of MPEG-7 descriptions (Section 5.3).

5.1 Existing MPEG-7 ontologies

Automatic mappings from the MPEG-7 XML Schema to OWL covering the whole standard have been proposed [6, 24]. The resulting ontology, however, is unable to capture the intended semantics not represented in the XML schema without re-engineering work. Other attempts have manually modeled an MPEG-7 ontology. The result is, however, either restricted to the upper level elements and types of MPEG-7 [9], or adapted to a very specific use of the standard in a particular application [21]. These ontologies could be used in the VAMP service as an alternative modeling of the semantic constraints as soon as a transformation into RDF and appropriate rules are provided. The validity would then not be checked against a particular profile.

[24] http://www.w3.org/2005/Incubator/mmsem/wiki/Vocabularies

5.2 Using MPEG-7 and its formalization

Considering the various shortcomings of the MPEG-7 schema-based representation with respect to a formal representation of its semantics, and the existing work for obtaining a formal model, one can wonder if it is worth keeping the MPEG-7 XML-based format. We argue that both representations are useful and are suitable for different purposes.

Describing the structure of audiovisual content, such as the sequence of shots contained in a video, is fundamental for many applications. Representing a structure with the current semantic web languages is often too complex. Due to the directed graph model with unordered edges used by OWL/RDF, it is not possible to determine the order of segments in the ontology-based representation without explicitly representing it [21]. Furthermore, numerous MPEG-7 low-level descriptors are characterized for having numerical values such as vectors and matrices while encapsulating few semantics. Hence, there is little or no advantage in having a formal representation for these concepts since: i) it is inefficient for typical operations such as similarity matching, ii) it will generate too many triples that might go beyond the current scale of RDF stores (consider for example the description of visual descriptors of the key frames of several hours of video).

5.3 Generalizing the VAMP approach

The approach presented in this paper is not limited to validating MPEG-7 documents. The basic idea of formalizing some semantic constraints of specific XML-based languages can be useful in a range of other applications. For example, VAMP could be used to validate semantically SMIL documents. In the advanced options, one would need to specify the URI of a SMIL ontology along with some associated logical rules capturing the intended semantics of this standard, and then provide the XSLT transformation. The SMIL document could then be checked with VAMP, even though the human-readable explanation of the various error types would need to be adapted. Mappings between various XML-based metadata formats as envisioned by the W3C Media Annotations Working Group[25] could thus benefit from the VAMP service.

6 Conclusion and future work

In this paper, we proposed a general approach to overcome the interoperability problems that result from the lack of formal semantics of the MPEG-7 description tools by formalizing their semantic constraints. The approach is based on the definition of profiles, which are not just subsets of the MPEG-7 standard, but that also define a set of semantic constraints that specify the use of the descriptors in a particular

[25] http://www.w3.org/2008/WebVideo/Annotations/

context. Our methodology advocates the specification of an ontology that includes the concepts being described in a profile, plus additional logical rules to fully capture the semantic constraints. We have demonstrated the feasibility of this approach by implementing VAMP, which is available both as a web application and as a web service. We have collected and analyzed numerous MPEG-7 descriptions from various tools from the multimedia community, and we have successfully applied VAMP for checking the constraints related to time ranges in temporal decompositions and to media information, highlighting the errors produced sometimes by these tools. The validation service is also now available for checking the semantics conformance of the MPEG-7 format used for representing shot boundary references, which is really useful for the TRECVID community when exchanging results.

When formalizing semantic constraints, the question of strictness consistency arises. There is, of course, always a tradeoff between flexibility and strictness with respect to description tool semantics. If we require the semantic constraints to be very strict, this might prevent the use of any structures in the description not foreseen in the profile definition, even if they are used as an extension and do not interfere with the structures defined in the profile. Thus it could be an option to introduce different levels of conformance to the profile semantics. We are working on this concept that we name “semantic levels”, by analogy to the levels of profiles in MPEG standards allowing different complexity. The idea is to define several levels of strictness in terms of semantic constraints for each profile which can then be used depending on application requirements. The definition starts with the most “liberal” semantic level: an ontology and a set of rules modeling the most basic semantic constraints of the profile. These constraints should only solve interoperability problems by avoiding ambiguities, but not unnecessarily restrict the use of optional elements or extensions. Based on this simple definition, stricter levels can be derived by adding further constraints to the ontology and defining additional rules.

Representing formally the semantic constraints of the MPEG-7 description tools is not only useful for semantically validating the descriptions, but also for establishing mappings between profiles and heterogeneous MPEG-7 descriptions. Actually, the greatest potential with semantic definitions of MPEG-7 profiles is in the ability to use these descriptions to relate the content to other audiovisual segments described using alternative MPEG-7 profiles or other domain ontologies such as EXIF or the ID3 tags. Current multimedia applications on the web need to index multimedia metadata from heterogeneous sources. Formalizing the semantics of the profiles used for representing this metadata allows to express mappings between heterogeneous descriptions based on their semantics. In the future, we plan to investigate further how the approach presented in this paper can be used in this particular use case.

Acknowledgements The authors would like to thank Alia Amin (CWI) for the design of the VAMP interface, Philip Hofmair and Rudolf Schlatte (JRS) for their help in the implementation of VAMP, and Lynda Hardman (CWI) for her feedback on earlier versions of this paper. The research leading to this paper was partially supported by the European Commission under the contracts FP6-027026, “Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space”, IST-2-511316, “IP-RACINE: Integrated Project - Research Area CINE” and FP6-027122, “SALERO: Semantic AudiovisuaL Entertainment Reusable Objects”.

References

1. Arndt R, Troncy R, Staab S, Hardman L, Vacura M (2007) COMM: designing a well-founded multimedia ontology for the web. In: 6th International semantic web conference (ISWC'07). Busan, South Korea, pp 30–43
2. Athanasiadis T, Tzouvaras V, Petridis K, Precioso F, Avrithis Y, Kompatsiaris Y (2005) Using a multimedia ontology infrastructure for semantic annotation of multimedia content. In: 5th International workshop on knowledge markup and semantic annotation (SemAnnot'05), Galway, Ireland
3. Bailer W, Schallauer P (2006) The detailed audiovisual profile: enabling interoperability between MPEG-7 based systems. In: 12th International multimedia modelling conference (MMM'06). Beijing, China, pp 217–224
4. Baral C, Gelfond M (1994) Logic programming and knowledge representation. J Log Program 19–20:73–148
5. Dean M, Schreiber G (2004) OWL Web ontology language: reference. W3C Recommendation. http://www.w3.org/TR/owl-ref/
6. Garcia R, Celma O (2005) Semantic integration and retrieval of multimedia metadata. In: 5th International workshop on knowledge markup and semantic annotation (SemAnnot'05). Galway, Ireland
7. Hobbs JR, Pan F (2006) Time ontology in OWL. W3C working draft. http://www.w3.org/TR/owl-time/
8. Höffernig M, Hausenblas M, Bailer W (2007) Semantics of temporal media content descriptions. In: Multimedia metadata applications workshop (M3A). Graz, Austria, pp 155–162
9. Hunter J (2001) Adding multimedia to the semantic web—building an MPEG-7 ontology. In: First international semantic web working symposium (SWWS'01), Stanford
10. Hunter J, Lagoze C (2001) Combining RDF and XML schemas to enhance interoperability between metadata application profiles. In: 10th International world wide web conference (WWW'01), Hong Kong, pp 457–466
11. International Organization for Standardization (2000) Representations of dates and times, 2nd edn. ISO 8601, 15 December 2000
12. Manola F, Miller E (2004) RDF (Ressource Description Framework) Primer. W3C Recommendation, 10 February 2004. http://www.w3.org/TR/rdf-primer/
13. Motik B, Sattler U, Studer R (2005) Query answering for OWL-DL with rules. J Web Semantics 3(1):41–60
14. MPEG-7 (2001) Multimedia content description interface. ISO/IEC 15938
15. MPEG-7 (2005) Information technology—multimedia content description interface—Part 9: profiles and levels. ISO/IEC 15938-9:2005
16. MPF (2008) Metadata production framework specifications (v. 2.0.2E). Technical report, NHK science and technical research laboratories. http://www.nhk.or.jp/strl/mpf/english/index.htm
17. Nack F, van Ossenbruggen J, Hardman L (2005) That obscure object of desire: multimedia metadata on the web (Part II). IEEE Multimed 12(1):54–63
18. Patel-Schneider PF, Hayes P, Horrocks I (2004) OWL web ontology language: semantics and abstract syntax. W3C Recommendation, 10 February 2004. http://www.w3.org/TR/owl-semantics/
19. Pereira F (2001) MPEG-7 requirements document v.16. ISO/IEC JTC1/SC29/WG11/N4510. Pattaya, Thailand
20. Pfeiffer S, Srinivasan U (2000) TV anytime as an application scenario for MPEG-7. In: Workshop on standards, interoperability and practice, Los Angeles
21. Troncy R (2003) Integrating structure and semantics into audio-visual documents. In: 2nd International semantic web conference (ISWC'03), Sanibel Island, pp 566–581
22. Troncy R, Bailer W, Hausenblas M, Hofmair P, Schlatte R (2006) Enabling multimedia metadata interoperability by defining formal semantics of MPEG-7 profiles. In: 1st International conference on semantics and digital media technology (SAMT'06), Athens, pp 41–55
23. Troncy R, Carrive J, Lalande S, Poli J-P (2004) A motivating scenario for designing an extensible audio-visual description language. In: The international workshop on multidisciplinary image, video, and audio retrieval and mining (CoRIMedia), Sherbrooke
24. Tsinaraki C, Polydoros P, Christodoulakis S (2004) Interoperability support for Ontology-based video retrieval applications. In: 3rd International conference on image and video retrieval (CIVR'04), Dublin

25. van Ossenbruggen J, Nack F, Hardman L (2004) That obscure object of desire: multimedia metadata on the web (Part I). IEEE Multimed 11(4):38–48
26. XML Schema (2001) W3C Recommendation, 2 May 2001. http://www.w3.org/XML/Schema

Dr. Raphaël Troncy is an Assistant Professor in the Multimedia Communications Department in the EURECOM Institute in Sophia Antipolis (France). He has obtained with honors his Master's thesis in Computer Science at the University of Grenoble (France) after one year spent in the University of Montreal (Canada). He benefited from a PhD fellowship at the National Audio-Visual Institute (INA) of Paris where he received with honors his PhD in 2004 from INRIA. He selected as an ERCIM Post-Doctorate Research Associate (2004–2006) where he visited the National Research Council (CNR) in Pisa (Italy) and the National Research Institute for Mathematics and Computer Science (CWI) in Amsterdam (The Netherlands). He was a senior researcher in the Interactive Information Access group at CWI Amsterdam (2006–2009).

Raphaël Troncy is co-chair of the W3C Incubator Group on Multimedia Semantics and the W3C Media Fragments Working Group, contributes to the W3C Media Annotations Working Group and actively participated in the K-Space Network of Excellence (2006–2009). He is an expert in audio-visual metadata and in combining existing metadata standards (such as MPEG-7) with current Semantic Web technologies. He works closely with the IPTC standardization body on the relationship between the NewsML language family and Semantic Web technologies.

Werner Bailer received a degree in Media Technology and Design from the Hagenberg University of Applied Sciences (Upper Austria) in 2002 for his diploma thesis on "Motion Estimation and Segmentation for Film/Video Standards Conversion and Restoration".

He is working at the Institute of Information Systems & Information Management since 2001 and has been involved in a number of national and European research projects in the area of audiovisual archiving, digital cinema production, digital film restauration and quality analysis and interactive TV. He has experience in development of image and video processing algorithms, metadata modelling and software architectures. Since 2007 he is working on a PhD thesis on the topic of multimedia content abstraction.

Martin Höffernig studies Telematics at the Technical University of Graz, Austria and is about to complete his diploma thesis on "Formalising Semantic Constraints for MPEG-7 Profiles". This work is done at the Institute of Information Systems and Information Management where he is working as a researcher since 2006. His research interests are Semantic Web Technologies and Multimedia Metadata.

Michael Hausenblas is a postdoctoral researcher at DERI where he acts as the coordinator of the Linked Data Research Centre. From 2001 to 2008 he worked at Joanneum Research where he also obtained his PhD from the Graz University of Technology. He has been and is involved in several EC projects around media semantics and has published over 40 papers at conferences, workshops and in journals. In mid 2008 he initiated the interlinking multimedia community effort and is now creating demonstrators for it. Michael is an active contributor to W3C standardisation activities.