Chapter 2
Multimedia Content Description Using Semantic Web Languages

Roberto García, Chrisa Tsinaraki, Òscar Celma, and Stavros Christodoulakis

2.1 Introduction

During the last decades, digital media have revolutionised media reproduction. The availability of cheap consumer electronic devices that allow the consumption and management of digital multimedia content (e.g. MP3 players, digital cameras, DV camcorders, smart phones) has caused a media availability explosion. The amount of digital media that has been generated and stored, and which continues to be at an exponential rate, has already become unmanageable without fine-grained computerised support.

This, in combination with the media distribution break-up carried out by the World Wide Web and the emergence of advanced network infrastructures that allow for the fast, efficient, and reliable transmission of multimedia content, has formed an open multimedia consumption environment. Digital multimedia content services are provided in this environment, which offer high content quality, advanced interaction capabilities, media personalisation and adaptation according to the user preferences, and access conditions. Such an open environment will be successful only if it is based on standards that allow the services provided by different vendors to interoperate. The specification of different multimedia content description standards poses interoperability requirements and necessitates guidelines for semantic interoperability. These issues are discussed in detail in Tzouvaras and Pan (2007).

The dominant standard in multimedia content description is MPEG-7 (ISO MPEG Group), which provides rich general purpose multimedia content description capabilities, including both low-level features and high-level semantic description constructs. However, the lack of formal semantics in MPEG-7 makes the gap between low- and high-level descriptions difficult to cope with for the existing tools. Consequently, low-level features are common, as they can be easily extracted from the content, but there is a lack of high-level descriptions.

R. García
Human-Computer Interaction Research Group (GRIHO), Universitat de Lleida, Lleida, Spain
e-mail: rgarcia@diei.udl.cat

Y. Kompatsiaris, P. Hobson (eds.), *Semantic Multimedia and Ontologies,*

Low-level approaches, based on signal analysis, are proving to be extremely limiting in making multimedia database systems accessible and useful to the end-users. These content-based descriptors lie far away from what the users recognise as media description means (Celma, Gómez, Janer, Gouyon, Herrera and García 2004). Consequently, recent research has begun to focus on bridging the semantic and conceptual gap that exists between the user and the computer – from content-based to high-level descriptions. One approach to overcome this gap is the use of knowledge-based techniques based on web ontologies. As formal and web-wide shared conceptualisations, ontologies facilitate automated integration and meaningful retrieval of multimedia – both content and metadata – from different sources.

Searching in digital libraries has been widely studied for several years, mostly focusing on retrieving textual information using text-based methods. These queries can be complemented and improved with advanced retrieval methods using content-based descriptors extracted from the audiovisual information by applying signal processing, even though some knowledge management and representation of the content is necessary. Moreover, from the service and content providers' point of view, multimedia metadata represents an added value to audiovisual assets, but then again manual annotation is a labour-intensive and error-prone task. Thus, managing audiovisual essence implies structuring its associated metadata using description schemes, taxonomies, and ontologies in order to organise a meaningful data knowledge representation.

In addition to the syntactic interoperation, which is achieved through the standards, semantic interoperation, which is achieved through the integration of domain knowledge expressed in the form of domain ontologies, is also needed for providing efficient retrieval and filtering services. The domain knowledge is subsequently utilised for supporting semantic personalisation, retrieval, and filtering and has been shown to enhance the retrieval precision (Tsinaraki, Polydoros and Christodoulakis 2007).

This chapter describes the representation of multimedia content descriptions that are structured according to the *MPEG-7* metadata description model and expressed using the Semantic Web languages. The rest of the chapter is structured as follows: Section 2.2 provides an overview of MPEG-7. The general purpose approaches for multimedia content description that are supported by the MPEG-7 standard are presented as well as the limitations of the current MPEG-7 version (mainly a lack of explicit semantics). Section 2.3 presents the existing web ontology languages, while Section 2.4 outlines the efforts made to move the MPEG-7 standard into the Semantic Web. In our case, this is accomplished by interpreting and expressing the informal MPEG-7 semantics using Semantic Web languages. An approach for mapping XML schema (Fallside 2001) constructs to OWL constructs (McGuinness and van Harmelen 2004) is presented in Section 2.5, while Section 2.6 presents two use cases that show the benefits of this approach, including semantic integration and retrieval in the music domain. An integrated ontological infrastructure for the semantic description of multimedia content is presented in Section 2.7. This infrastructure allows for combining the general purpose MPEG-7 constructs with domain and application-specific knowledge through the systematic representation of this knowledge in the form of *web ontology language* (*OWL*) domain and application ontologies integrated with the MPEG-7 semantics. The chapter conclusions are presented in Section 2.8.

2.2 Multimedia Content Description Using MPEG-7

MPEG-7, formally named *multimedia content description interface*, is an ISO/IEC standard developed by the Moving Picture Experts Group (MPEG), the committee that also developed the audiovisual standards: MPEG-1, MPEG-2, MPEG-4, and MPEG-21. MPEG-7 aims to create a standard for the description of the multimedia content. The main goal of the MPEG-7 standard is to provide structural and semantic description mechanisms for multimedia content (Salembier, Manjunath and Sikora 2002; Martínez 2004).

The MPEG-7 standard allows content description for audiovisual content, defining normative elements such as *descriptors*, *DescriptionSchemes*, and a *description definition language* (DDL). The DDL is the basic building blocks for the MPEG-7 metadata language. Descriptors are designed for describing different types of information; low-level audiovisual features, high-level semantic objects, content management, and information about storage media. Description schemes are used to group several descriptors (and description schemes) into structured semantic units using the DDL. Ideally, most descriptors corresponding to low-level features would be extracted automatically, whereas human intervention would be required for producing high-level descriptors.

The standard is divided into four main components: the DDL, the audio part, the visual part, and the information about how these elements are combined in a multimedia scenario – a set of multimedia description schemes that includes all the descriptors for capturing the semantic aspects of multimedia contents, e.g. places, actors, objects, events. Thus, the creation of MPEG-7 documents allows a user to query and retrieve (parts of) multimedia and audiovisual information.

In the rest of this section, we discuss media object information description in Section 2.2.1, text-based media description in Section 2.2.2, low-level feature-based media description in Section 2.2.3, semantic-based media description in Section 2.2.4, and MPEG-7 description retrieval in Section 2.2.5.

2.2.1 Media Object Information Description

Of special interest is part 5 of the MPEG-7 standard, named *MultimediaDescriptionSchemes* (MPEG-7 MDS, ISO/IEC 2003). This part includes a set of description tools dealing with generic features and multimedia descriptors. Figure 2.1 depicts all the components of the MDS. The basic elements component includes basic data types, such as media localisation, time format, and free text annotations. It includes, also, the classification schemes (CS) descriptors. CS descriptors define schemes for

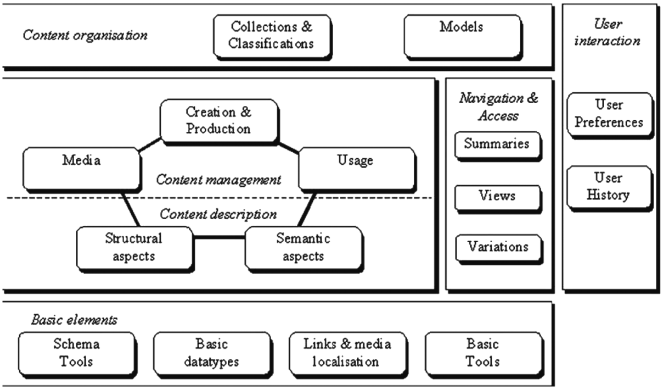

Fig. 2.1 Main elements of the MPEG-7 multimedia description schemes

classifying a subject area with a set of terms, organised into a taxonomy. Similar to the WordNet linguistic ontology, basic relationships among the taxonomy terms are available (e.g. narrow and broader terms, and synonyms).

Among the main components of the MDS are the *ContentManagement and description schemes*. The content management descriptors allow description of the life cycle of multimedia content, from its creation to its usage. They include media information to describe storage format, media quality, media location, etc. Moreover, the content management schemes allow gathering editorial data about the creation and production process of the content. The content description schemes describe the structural aspects (spatial, temporal, and media source structure of the multimedia content) and the semantic aspects.

In detail, the media object information consists of the following:

- The *media information*, which is captured in one of the *MediaInformation*, *MediaInformationRef*, and *MediaLocator* elements. The media information consists of the *media identification*, which allows the unique identification of the media object and its locator, and the *media profile*, which provides media-related information (including media format, quality).
- The *creation information*, which is captured in one of the *CreationInformation* and *CreationInformationRef* elements. The creation information consists of information about the media object *creation* (including title, creators, abstract), *classification* (including genre, subject, language) as well as information about *related material*.
- The *structural information*, which is captured in the *StructuralUnit* element and describes the role of the current multimedia object (segment) within the information context. Thus, the *StructuralUnit* may take values such as "scene", "shot", and "story".
- The *usage information*, which is captured in one of the *UsageInformation* and *UsageInformationRef* elements. The usage information consists of information

about the *rights* associated with the multimedia object, its *financial results*, its *availability*, and its *usage record*.

- Information regarding the importance of the multimedia content from specific *points of view*. This information is captured in the *PointOfView* element.
- The *relationships* of the multimedia content with other media or metadata items as well as the relationships of the semantic entities describing the multimedia content. This information is captured in the *Relation* element, which associates the media object descriptions with instances of the *RelationType* that represent relationships. A relationship may be directed or undirected and features a relationship *type*, the *target* and the *source* of the relationship, and the *strength* of the relationship. The standardised MPEG-7 relationship types are more than 100 and are classified into (a) basic relationship types (equals, inside, refines, etc.), which are specified in the *BaseRelation CS*; (b) graph node relationship types (identity, equivalent, etc.), which are specified in the *GraphRelation CS*; (c) spatial relationship types (over, below, north, etc.), which are specified in the *SpatialRelation CS*; (d) temporal relationship types (precedes, overlaps, contains, etc.), which are specified in the *TemporalRelation CS*; and (e) semantic relationship types (shows, agent, causer, etc.), which are specified in the *SemanticRelation CS*.
- The *matching hints* that allow expression of the criteria for matching the multimedia content with low-level audio and visual descriptors. This information is captured in the *MatchingHint* element.

As an example, consider the MPEG-7 image description of Fig. 2.2, where Chrisa is shown to write an article. The image description consists of the *MediaLocator* element, where the image location (http://www.music.tuc.gr/img01.jpg) is specified,

```
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001
C:\dbxml\mpeg7.xsd">
 <Description xsi:type="ContentEntityType">
  <MultimediaContent xsi:type="ImageType">
   <Image id="img1">
    <MediaLocator>
     <MediaUri>http://www.music.tuc.gr/img01.jpg</MediaUri>
    </MediaLocator>
    <CreationInformation>
     <Creation>
      <Title>Image showing Chrisa writing an article</Title>
     </Creation>
    </CreationInformation>
   </Image>
  </MultimediaContent>
 </Description>
</Mpeg7>
```

Fig. 2.2 MPEG-7 image description example and the *CreationInformation* element, where the image title (image showing Chrisa writing an article) is specified in its *Title* element.

2.2.2 Text-Based Media Description

In this section, we describe the text-based multimedia content description capabilities that are provided by the MPEG-7 MDS (ISO/IEC 2003). The textual annotations are represented by the *TextAnnotation* element of the MPEG-7 segment descriptions. An MPEG-7 textual annotation consists of the following elements, each of which may occur an arbitrary number of times:

- The *FreeTextAnnotation* element, which represents free text annotations;
- The *StructuredAnnotation* element, which represents structured textual annotations in terms of *who* (people and animals), *what object*, *what action*, *where* (places), *when* (time), *why* (purpose), and *how*;
- The *KeywordAnnotation* element, which represents keyword annotations;
- The *DependencyStructure* element, which represents textual annotations with a syntactic parse based on dependency structures.

The confidence in the correctness of a textual annotation and its relevance to the multimedia object being described are represented, in the [0, 1] range, by the *confidence* and *relevance* attributes of the textual annotation.

As an example, consider the textual part of the MPEG-7 image description of Fig. 2.2, which is shown in Fig. 2.3. The textual annotation consists of a free text annotation (captured in the *FreeTextAnnotation* element) and a structured annotation (captured in the *StructuredAnnotation* element).

2.2.3 Low-Level Feature-Based Media Description

MPEG-7 (ISO MPEG Group) allows associating, in the MPEG-7 multimedia object descriptions, low-level visual and audio features with the media objects being described. According to the MPEG-7 MDS (ISO/IEC 2003), the MPEG-7

```
                                ...
<TextAnnotation confidence="0.9" relevance="1">
 <FreeTextAnnotation>Chrisa writes an article</FreeTextAnnotation>
 <StructuredAnnotation>
  <Who><Name>Chrisa</Name></Who>
  <WhatObject><Name>Article</Name></WhatObject>
  <WhatAction><Name>Writes</Name></WhatAction>
 </StructuredAnnotation>
</TextAnnotation>
                                ...
```

Fig. 2.3 Textual part of the MPEG-7 image description of Fig. 2.2

descriptions that describe (segments of) multimedia objects having a visual component (e.g. images, videos, audiovisual segments) may represent the visual features of the described (segments of) multimedia objects through the *VisualDescriptor* and the *VisualDescriptionScheme* elements using, respectively, visual descriptors and visual description schemes. The MPEG-7 descriptions that describe (segments of) multimedia objects having an audio component (e.g. audio segments, audiovisual segments) may represent the audio features of the described (segments of) multimedia objects through the *AudioDescriptor* and the *AudioDescriptionScheme* elements using, respectively, audio descriptors and audio description schemes (ISO/IEC 2001b). In the rest of this section, we will present the low-level feature-based multimedia content description capabilities that are provided by MPEG-7, focusing on the visual features.

A set of basic low-level descriptors are defined in the MPEG-7 visual, including the basic *color descriptors*, the basic *texture descriptors*, the basic *shape descriptors*, and the basic *motion descriptors*.

MPEG-7 also provides supplementary textual structures for colour spaces, colour quantisation, and multiple 2D views of 3D objects. It also allows for using static (image) descriptors on video content and for the spatial as well as the temporal localisation of media object descriptors.

2.2.4 Semantic-Based Media Description

In this section, we describe the semantic-based multimedia content description capabilities provided by the MPEG-7 MDS (ISO/IEC 2003). The semantic multimedia content descriptions are represented by the *Semantic* element of the MPEG-7 segments, where a set of semantic entities describing the segment content may be defined or referenced. It has been shown in Tsinaraki, Polydoros, Kazasis and Christodoulakis (2005) that the MPEG-7 semantic description capabilities allow, in addition to the representation of semantic multimedia content descriptions, the representation of domain ontologies using pure MPEG-7 constructs (details of this methodology are provided in Section 2.7).

The semantic entities participating in MPEG-7 descriptions are instances of the subtypes of the abstract type *SemanticBaseType*, which represent semantic entities of specific types in a narrative world. The *AbstractionLevel* element of the *SemanticBaseType* specifies whether a semantic entity is abstract or concrete. *AbstractionLevel* has one attribute, *Dimension*, of non-negative integer type. When *AbstractionLevel* is not present in a semantic description, the description refers to specific audiovisual material. When *AbstractionLevel.Dimension*=0, it is a description of a reusable semantic entity (e.g. the person Chrisa) that is referenced from every segment where the entity appears. When *AbstractionLevel* has a non-zero *Dimension*, it specifies classes for the description of abstract semantic entities (e.g. the *Article* semantic entity, with *AbstractionLevel.Dimension*=1, represents the class of the articles). The subtypes of *SemanticBaseType* that represent different types of semantic entities are the following:

- The *SemanticType*, which is a concrete type used for the description of collections of semantic entities;
- The *AgentObjectType*, which is a concrete type used for the description of the actors that appear in a segment. The actors are specified in the *Agent* element of *AgentObjectType*. Actors in general are represented using the subtypes of the abstract type *AgentType*. *PersonType*, *OrganizationType*, and *PersonGroupType* are the subtypes of *AgentType* and are used for the representation of persons (e.g. a student), organisations (e.g. a university), and groups of persons;
- The *ObjectType*, which is a concrete type used for the description of objects and object abstractions in the material world (e.g. a desk);
- The *EventType*, which is a concrete type used for the description of events that take place in a semantic world (e.g. writing);
- The *ConceptType*, which is a concrete type used for the description of concepts present in an audiovisual segment (e.g. cooperation);
- The *SemanticStateType*, which is a concrete type used for the description of a state of the world described in an audiovisual segment and the parametric description of its features (e.g. the average of a student's grades before and after an examination period);
- The *SemanticPlaceType*, which is a concrete type used for the description of a place in a semantic world (e.g. Crete);
- The *SemanticTimeType*, which is a concrete type used for the description of semantic time (e.g. New Year's Eve).

As an example, consider the semantic part of the MPEG-7 image description of Fig. 2.2, which is shown in Fig. 2.4.

Notice that the *ChrisaArticle* object is the result of the *Writes* event. The agent of the event is the person represented by the *Chrisa* semantic entity.

Semantic entity (abstract or concrete) definitions may occur in the context of either segment descriptions or independent semantic descriptions. The semantic entity definitions occurring in independent semantic descriptions may then be referenced from the segment descriptions they appear in. This is very useful both for the ontology classes and for the reusable semantic entities.

2.2.5 Retrieving Information from MPEG-7 Descriptions

The eXtensible Markup Language (XML) has been adopted as the format to represent MPEG-7 descriptors. Also the MPEG-7 DDL is an extension of the W3C XML schema. XML schema provides the means for defining the structure of XML documents, that is, simple and complex data types, type derivation and inheritance, element occurrence constraints, and, finally, namespace-awareness for element and attribute declarations. The MPEG-7 DDL extends the XML schema and covers the ability to define array and matrix data types, and provides specific temporal descriptions (by means of the *basicTimePoint* and *basicDuration* types).

```
...
<Semantic>
 <Label><Name>Chrisa writes an article</Name></Label>
 <SemanticBase xsi:type="ObjectType" id="ChrisaArticle">
  <AbstractionLevel dimension="0"/>
  <Label><Name>Chrisa's Article</Name></Label>
 </SemanticBase>
 <SemanticBase xsi:type="AgentObjectType" id="Chrisa">
  <AbstractionLevel dimension="0"/>
  <Label><Name>Chrisa</Name></Label>
  <Agent xsi:type="PersonType">
   <Name>
    <GivenName>Chrisa</GivenName>
    <FamilyName>Tsinaraki</FamilyName>
   </Name>
  </Agent>
 </SemanticBase>
 <SemanticBase xsi:type="EventType" id="Writes">
  <AbstractionLevel dimension="0"/>
  <Label><Name>Writes</Name></Label>
  <Relation source="#Writes" target="#Chrisa"
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agentOf"/>
  <Relation source="#Chrisa" target="#Writes"
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent"/>
  <Relation source="#ChrisaArticle" target="#Writes"
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:resultOf"/>
  <Relation source="#Writes" target="#ChrisaArticle"
type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:result"/>
 </SemanticBase>
</Semantic>
...
```

Fig. 2.4 Semantic part of the MPEG-7 image description of Fig. 2.2

The MPEG-7 XML schemas define 1182 elements, 417 attributes, and 377 complex types. The size of this standard makes it quite difficult to manage. Moreover, the use of XML technologies implies that a great part of the semantics remains implicit. Therefore, each time an MPEG-7 application is developed, semantics must be extracted from the standard and re-implemented.

The next two examples depict how to retrieve information from MPEG-7 MDS documents using the *XQuery* (Siméon, Chamberlin, Fernández, Boag, Florescu and Robie 2007) language and an XML database. The first example in Listing 1 shows an expression to retrieve MPEG-7 audiovisual segments containing any media information. The output is presented as simple HTML code, containing a link to the media file – with the title and type of file as the text link.

```
for $segment in//AudioVisualSegment
let $title:=$segment/CreationInformation/Creation/Title/text()
order by $title
return
 for $media in $segment/MediaInformation/MediaProfile
 let $file:=$media/MediaInstance/MediaLocator/MediaUri/text()
 let $type:=$media/MediaFormat/Content/Name/text ()
 return
  <a href="{$file}">{ $title," [",$type,"]" }</a>
```

Listing 1 *XQuery* expression to retrieving a list of multimedia items (title and format type)

```
for $creator in
 /Mpeg7/Description/MultimediaContent/*/CreationInformation/
 Creation/Creator
 where
  $creator/Role[@href="urn:opendrama:cs:SingerCS:%"]
  and
  $creator/Agent[@xsi:type="PersonType"]
order by $creator/Agent/Name/FamilyName
return
 <agent>
 {
  let $completeName:= $creator/Agent/Name
  let $name:= $completeName/GivenName/text()
  let $surname:= $completeName/FamilyName/text()
  return
   <singer> { $name," ",$surname }</singer>
}
{
 let $completeName:= $creator/Character
 let $name:= $completeName/GivenName/text()
 let $surname:= $completeName/FamilyName/text()
 return
  <character> { $name, " ",$surname }</character>
}
</agent>
```

Listing 2 *XQuery* example to retrieving the singers and the characters they play

The second example (Listing 2) shows an *XQuery* expression to retrieve all MPEG-7 person agents, whose role is *Singer*, and the characters they play. This query uses a taxonomy that defines different types of singers' roles (soprano, contralto, tenor, and bass).

The previous examples only illustrate one kind of difficulty derived from the use of just syntax-aware tools. In order to retrieve any kind of MPEG-7 *SegmentType* descriptions from an XML database, one must be aware of the hierarchy of segment types and implement an *XQuery* that covers any kind of multimedia segment (i.e. *Audio-VisualType*, *VideoSegmentType*, *AudioSegmentType*). On the other hand, once the hierarchy of segments is explicitly defined in an ontology (e.g. in OWL form), semantic queries benefit from the, now, explicit semantics. Therefore, a semantic query for *SegmentType* will retrieve all the subclasses without requiring additional efforts. This is necessary because although XML schemas capture some semantics of the domain they model, XML tools are based on syntax. The captured semantics remain implicit from the XML processing tools point of view. Therefore, when an *XQuery* searches for a *SegmentType*, the *XQuery* processor has no way to know that there are many other kinds of segment types that can appear in its place, i.e. they are more concrete kinds of segments. At this stage, a possible solution to avoid this is to use wildcards' syntax (see the second and fifth lines of Listing 2). However, this corresponds to a unconstrained generalisation, i.e. any element satisfies it and it is not possible to constrain it to just a kind of element, e.g. all the AudioVisualType subtypes.

Therefore, MPEG-7 constitutes a valuable starting point for more specific developments as it can be seen as an "upper-ontology" for multimedia. However, the lack of explicit semantics makes MPEG-7 very difficult for third-party entities to extend in an independent way. This lack of facilities for easy extension has been one of the main motivations to build solutions that make MPEG-7 semantics formal and thus easily machine-processable. Some solutions to this problem are detailed in Section 2.4.

2.3 Web Ontology Languages

The World Wide Web has changed the way people communicate with each other. Most of today's Web content is suitable for human consumption. Keyword-based engines have helped users to find the information they are seeking on the net. Yet, search engines present some limitations: the results are single web pages, results are highly sensitive to the vocabulary (semantically similar queries should return similar results), and usually there is a high recall and low precision of the result set (i.e. there is too much noise on the web page results) (Antoniou and van Harmelen 2004).

The main problem of the current Web, at this stage, is that the meaning of the content is not accessible by machines. Information retrieval and text processing tools are widely used, but there are still difficulties when interpreting sentences, or extracting useful information for users. The development of the Semantic Web, with machine-readable content, has the potential to revolutionise the current World Wide Web and its use.

2.3.1 Overview of the Semantic Web

The definition and vision that had Tim Berners-Lee (1999) is that the Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes but for automation, integration, and reuse of data across various applications (Berners-Lee, Hendler, and Lassila 2001; Shadbolt, Berners-Lee and Hall 2006).

The previous ideas and principles to enhance the Web are being put into practice under the guidance of the World Wide Web Consortium (W3C). The next statement presents their view:

> The semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The mix of content on the web has been shifting from exclusively human-oriented content to more and more data content. The semantic web brings to the web the idea of having data defined and

> linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications. For the web to reach its full potential, it must evolve into a semantic web, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people.
>
> — W3C Semantic Web Activity Statement

The Semantic Web technologies have been arranged into a layered architecture. The key technologies include explicit metadata, ontologies, logic and inferencing, and intelligent agents. Each layer, from the bottom to the top, has an increasing level of complexity, yet it offers more expressivity.

The two base layers (unicode and URI, and the XML family) are inherited from the current Web. Section 2.2 already has presented some technologies relating to XML. The upper layers compose the Semantic Web, over the existing basic technologies. The next sections overview these layers, that is, the Resource Description Framework (RDF), the RDF Schema (RDFS), and the Web Ontology Language (OWL).

2.3.2 Resource Description Framework

The RDF (Brickley and Guha 2004) vocabulary is similar to other knowledge representation formalisms such as conceptual graphs (CG) or semantic nets. CG express meaning in a form that is logically precise, humanly readable, and computationally tractable. CG serve as an intermediate language for translating computer-oriented formalisms to and from natural languages. With a clear graphic representation, they serve as a readable – but formal – design and specification language. The next figure, Fig. 2.5, shows an example of a semantic net, which relates music bands, artists, and basic data.

Graph representation is a powerful tool for human understanding. However, in our context we need machine-processable representations.

The RDF vocabulary allows formally describing the previous example, and even serialising it using the XML language. RDF is, then, a data model for objects (resources) and the relations (properties) between them, and it provides simple

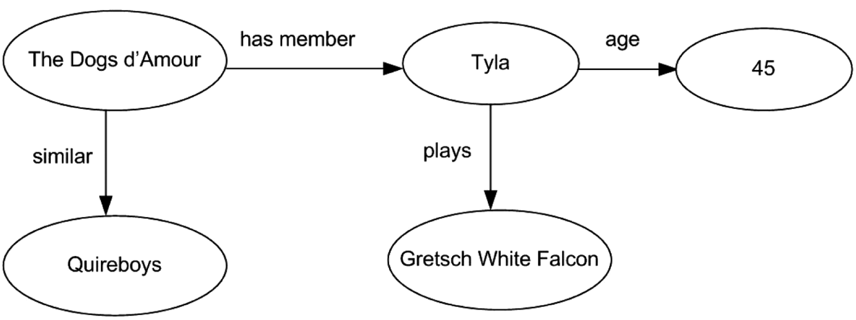

Fig. 2.5 A semantic net

semantics. A resource is an object, a thing we want to talk about. A resource has a URI (Uniform Resource Identifier). Properties are a special kind of resource that describe relations between resources (e.g. related with, age, plays). Properties are identified by URIs.

Statements assert the properties of resources. From a natural language point of view, a statement is composed of a *Subject–Predicate–Object* triple. From a more computer science point of view, this is equivalent to an *Object-Attribute-Value* triple, or in this context a *Resource-Property-Value* triple. A triple [x, P, y] is equal to a logical formula $P\ (x, y)$, where the binary predicate P relates the object x to the object y. Values can be either resources or literals (e.g. strings).

A possible statement could be, "Oscar Celma is the owner of the web page http://foafing-the-music.iua.upf.edu". This triple is equal to the graph statement, Fig. 2.6.

It is a directed graph, where the nodes correspond to the objects and the labelled arc is a property. The same statement can be represented in XML syntax (also known as RDF/XML):

```
<rdf:Description
 rdf:about="http://foafing-the-music.iua.upf.edu">
 <mydomain:owner>Oscar Celma</mydomain:owner>
</rdf:Description>
```

The `rdf:Description` makes a statement about the resource (a web page) `http://foafing-the-music.iua.upf.edu`. The property (*owner*) is used as a tag within the description, and the value is the content of the tag. Moreover, we can describe the person "Oscar Celma" by the resource with URL

http://www.mydomain.org/people/#44521:

```
<rdf:Description
 rdf:about="http://www.mydomain.org/people/#44521">
 <mydomain:name>Oscar Celma</mydomain:name>
 <mydomain:title>Associate Professor</mydomain:title>
</rdf:Description>
```

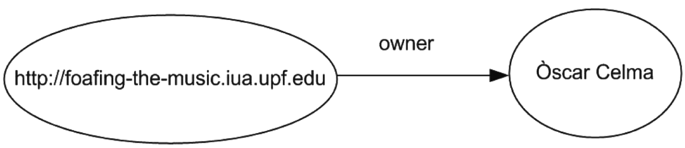

Fig. 2.6 Graph representation of a triple

In this case, the rdf:Description corresponds to two statements about the resource http://www.mydomain.org/people/#44521 (the name, and the title of that person). Now, we can define a *course* that *is taughtby* that resource:

```
<rdf:Description
 rdf:about="http://www.tecn.upf.es/˜ ocelma/edi2">
 <uni:courseName>Introduction to Databases</uni:courseName>
 <uni:creditsNumber>6</uni:creditsNumber>
 <uni:isTaughtBy rdf:resource="http://www.mydomain.org/people
    /#44521" />
</rdf:Description>
```

The resulting graph of the three previous examples is depicted in Fig. 2.7.

By now, we have defined a set of statements, but there are still no restrictions about them. For instance, we should state that the property *isTaughtBy* is only applied to courses (the subject) and professors (the object), or that an associate professor is a particular type of professor, with some restrictions (maximum number of hours, needs to hold a PhD, etc.). The RDF Schema vocabulary is intended to describe this information.

2.3.3 RDF Schema

RDF Schema (RDFS) (Manola and Milles 2004) is a vocabulary for describing properties and classes of RDF resources, and provides hierarchies of such properties and classes. The RDFS vocabulary allows definition of the semantics of the RDF statements.

As is common in other disciplines, to describe a particular domain one can use classes and properties. RDFS provides mechanisms to define a particular domain using classes (and properties), hierarchies, and inheritance. Classes model the entities (and their restrictions) of the domain, whereas properties provide relationships

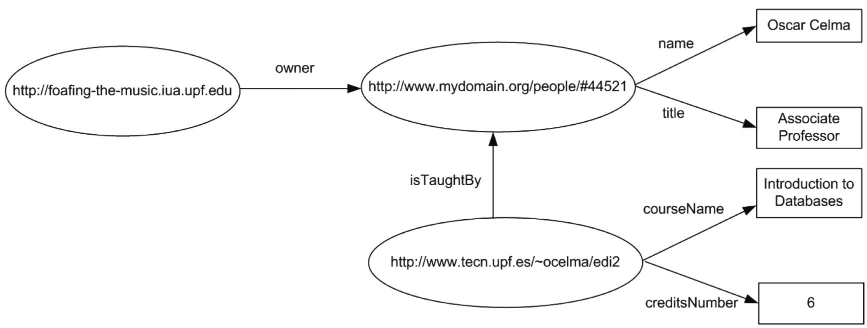

Fig. 2.7 Graph representation of the previous RDF statements

among the classes. Properties have a domain and range (similarly to mathematical functions), to impose restrictions on the values of the property. Yet there are some important missing features of RDFS:

- There are no local scope properties: *rdf:range* defines the range of a property for all classes. We cannot declare range restrictions that apply to some classes only;
- There is no disjointness of classes;
- Missing Boolean combinations of classes: union, intersection, and complement;
- No cardinality restrictions: restrictions on how many distinct values a property may or must take (“a person has two parents”);
- No special characteristics of properties: transitive (greater than), unique (is mother of), and inverse (eats and is eaten by).

These limitations are solved in the OWL language, presented in the next section. To conclude this section, a simile can be established among the existing technologies on the current Web, and the ones proposed by the Semantic Web community: while the XHTML language makes the Web behave like a global book when viewed at the worldwide level, RDF and RDF Schema make it behave like a global database. Regarding the data structures, the basic RDF primitive is a directed graph, whereas the XML representation is based on a tree. Thus, an RDF graph is on its own basically unrestricted and more powerful in terms of expressiveness.

2.3.4 Ontology Vocabulary

An ontology is an explicit and formal specification of a conceptualisation (Gruber 1993). In general, an ontology describes formally a domain of discourse. The requirements for ontology languages are a well-defined syntax, a formal semantics, and a reasoning support that checks the consistency of the ontology, checks for unintended relationships between classes, and automatically classifies instances in classes.

The web ontology language (OWL) has a richer vocabulary description language for describing properties and classes than RDFS. OWL has relations between classes, cardinality, equality, characteristics of properties, and enumerated classes. The OWL is built on RDF and RDFS and uses RDF/XML syntax. OWL documents are, then, RDF documents.

The next example shows the definition of two classes:

```
<owl:Class rdf:ID="Singer">
 <rdfs:subClassOf rdf:resource="#Artist" />
</owl:Class>
<owl:Class rdf:ID="Song" />
```

Object property elements relate objects to other objects. For instance, “a singer *sings* songs”.

```
<owl:ObjectProperty rdf:ID="sings">
 <rdfs:domain rdf:resource="#Singer"/>
 <rdfs:range rdf:resource="#Song"/>
</owl:ObjectProperty>
```

Data-type properties relate objects to data-type values. For example, the data property that denotes the age of an *Artist*:

```
<owl:DataProperty rdf:ID="age">
 <rdfs:domain rdf:resource="#Artist"/>
 <rdfs:range rdf:resource="&xsd;nonNegativeInteger"/>
<owl:DataProperty>
```

Property restrictions on classes are based on the use of rdfs:subclassOf. To say that class C satisfies certain conditions is equivalent to state that C is a subclass of C’, where C’ collects all objects that satisfy the conditions. For instance, a restriction on the kind of values the property can take:

```
<owl:Class rdf:about="#GuitarPlayer">
<rdfs:subClassOf>
 <owl:Restriction>
  <owl:onProperty rdf:resource="#plays"/>
  <owl:allValuesFrom rdf:resource="#Guitar"/>
 </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

Or cardinality restrictions (a *music band* comprises, at least, two members):

```
<owl:Class rdf:about="#Band">
 <rdfs:subClassOf>
   <owl:Restriction>
    <owl:onProperty rdf:resource="#hasMember"/>
    <owl:minCardinality ="&xsd;nonNegativeInteger">
    </owl:minCardinality>
  </owl:Restriction>
 </rdfs:subClassOf>
</owl:Class>
```

OWL offers some special properties, such as: *owl:TransitiveProperty* (e.g. “has better grade than”, “is taller than”, “is ancestor of”), *owl:SymmetricProperty* (e.g. “has same grade as”, “is sibling of”), *owl:FunctionalProperty* (a property that has almost one value for each object, e.g. “age”), and *owl:InverseFunctionalProperty* (a property for which two different objects cannot have the same value, e.g. “socialSecurityNumber”). For example, a *playedwith* property is symmetric:

```
<owl:ObjectProperty rdf:ID="playedWith">
 <rdf:type rdf:resource="&owl;SymmetricProperty"/>
 <rdfs:domain rdf:resource="#Artist"/>
 <rdfs:range rdf:resource="#Artist"/>
</owl:ObjectProperty>
```

There are three different OWL sublanguages. Each sublanguage offers a level of expressivity. OWL Full is the most expressive of the three sublanguages. There are no special constraints about how the OWL primitives can be used. Therefore, the greatest level of expressivity of the language can be achieved. However, on the other hand, the language becomes undecidable, so efficient reasoning is not guaranteed.

OWL-DL is based on description logics. It has vocabulary partitioning, that is, any resource is allowed to be only a class, a data-type, a data-type property, an object property, an individual, a data value, or part of the built-in vocabulary. Also, there is explicit typing in OWL-DL, so the vocabulary partitioning must be stated explicitly.

Property separation implies that the following can never be specified for data type properties: *owl:inverseOf*, *owl:FunctionalProperty*, *owl:InverseFunctionalProperty*, and *owl:SymmetricProperty*. Additionally, there is a restriction for anonymous classes: they are only allowed to occur as the domain and range of either *owl:equivalentClass* or *owl:disjointWith* and as the range of *rdfs:subClassOf*.

These constraints on how OWL primitives are combined guarantee that to reason on OWL-DL expressions is decidable and tractable, i.e. it will terminate in a finite and a not too large amount of time. This is so because OWL-DL is in the family of description logics. Description logics allow the specification of a terminological hierarchy using a restricted set of first-order formulae. Restrictions ensure that description logics have nice computational properties, but the inference services are restricted to subsumption and classification. Subsumption means, given formulae describing classes, the classifier associated with certain description logic will place them inside a hierarchy. On the other hand, classification means that given an instance description, the classifier will determine the most specific classes to which the particular instance belongs.

Finally, OWL Lite has the same restrictions as OWL-DL plus it is not allowed to use *owl:oneOf*, *owl:disjointWith*, *owl:UnionOf*, *owl:complementOf,* or *owl:hasValue*. Regarding cardinality statements: only values 0 and 1 are possible. These additional constraints reduce even more the expressivity of the language but, on the other hand, make reasoning more efficient.

2.4 MPEG-7 Ontologies

As shown in Section 2.2, MPEG-7 allows for the semantic annotation of multimedia content and the systematic representation of domain knowledge using MPEG-7 constructs. The domain knowledge is usually expressed today in the form of domain ontologies, and several ontology description languages have been proposed, based on the OWL language presented in Section 2.3. Thus, it is expected both that many OWL domain ontologies will be developed and that many developers will be familiar with OWL and will use it for ontology definition. It is therefore very important for the multimedia community to have a methodology for the interoperability of OWL with MPEG-7 and for the integration of domain knowledge expressed in OWL within MPEG-7. This way, the MPEG-7 constructs will become Semantic Web objects and the Semantic Web tools (such as reasoners) and methodologies may be used with MPEG-7. This feature is useful for several applications, e.g. knowledge acquisition from multimedia content. In this section, we present the ontologies expressed in the Semantic Web languages that capture (fully or partially) the MPEG-7 semantics. As a consequence, although MPEG-7 is a standard, hence it enhances interoperability at least at the syntactic level, the several different ontological MPEG-7 representations are not standard, and are not compatible or interoperable with each other. Thus, a new interoperability issue appears, which is discussed in detail in the *Harmonization of Multimedia Ontologies activity* of the aceMedia project (2007) and Celma, Dasiopoulou, Hausenblas, Little, Tsinaraki and Troncy (2007).

Chronologically, the first efforts towards a semantic formalisation of MPEG-7 were carried out, during the MPEG-7 standardisation process, by Jane Hunter (1999). The proposal used RDF (Brickley and Guha 2004) and RDF Schema (Manola and Milles 2004) to formalise a small part of MPEG-7 and later incorporated some DAML+OIL constructs (McGuinness, Fikes, Hendler and Stein 2002) to further detail their semantics (Hunter 2001), where the DAML+OIL ontology definition language was used to partially describe the MPEG-7 MDS and visual metadata structures. The ontology has been recently translated into OWL. However, it continues to show one of its major shortcomings, the limited coverage of the MPEG-7 constructs.

Another proposal based on RDF/RDFS that captures the MPEG-7 visual has been presented in Simou, Tzouvaras, Avrithis, Stamou and Kollias (2005). The same shortcomings are observed due to the expressivity limitations of RDF/RDFS. Consequently, this ontology also provides limited support for the representation of MPEG-7 formal semantics.

An OWL full ontology that captures the whole MPEG-7 standard was presented in García and Celma (2005). This ontology has been automatically produced using the mappings from XML schema constructs to the OWL constructs, which are detailed in Section 2.5. This mapping is complemented with an XML to RDF one that makes it possible to map existing MPEG-7 data to RDF data based on the previous ontology.

The disadvantage of modelling the whole standard is that this ontology is OWL full, which means that computational completeness and decidability of reasoning are not guaranteed. However, this limitation is unavoidable due to the structure of the MPEG-7 standard XML schemas. The only way to avoid it, if all the semantics implicit in the schemas are formalised, is to restrict the ontology to just a part of the standard.

This is the approach of the OWL-DL ontology presented in Tsinaraki, Polydoros and Christodoulakis (2004b), which captures the full MPEG-7 MDS (including the classification schemes) and just the parts of the MPEG-7 visual and audio that are necessary for the complete representation of the MPEG-7 MDS. The ontology was manually developed, according to a methodology that allows the transformation of the XML schema constructs of MPEG-7 in OWL-DL.

The methodology consists of the following steps:

- The MPEG-7 simple data-types are imported from the XML schema syntax, as OWL does not directly support simple type definition.
- The MPEG-7 complex types are represented as OWL classes, which have the complex type names as identifiers. The attributes and the simple type elements (of type string, integer etc.) of the complex types are represented as OWL data type properties that have the OWL classes that represent the complex types as domain and the simple types as range. The complex type elements are represented as OWL object properties that have the OWL classes that represent the complex type as domain and the OWL classes that represent the element types as range (if the latter do not already exist, it is defined from scratch).
- For the representation of the subtype/supertype relationships that hold for a complex type, the following actions are performed: (a) If the complex type is a subtype of another complex type, the subclass relationship is represented using the OWL/RDF subclassing construct; and (b) If the complex type is a subtype of a simple type, a data-type property is defined that has as identifier "*type_name*Content", where *type_name* is the type of the supertype (e.g. string, integer). The data type property has the supertype as range and the OWL class that represents the complex type as domain.
- The XML schema restrictions are transformed to the analogous OWL constructs. Thus, a fixed attribute value is transformed to an OWL "hasValue" restriction, and the minOccurs/maxOccurs attributes are transformed to either simple cardinality restrictions (i.e. cardinality, minCardinality, and maxCardinality) or groups of cardinality restrictions, grouped using the OWL unionOf (in case of choices) and intersectionOf (in case of sequence) constructs.
- The MPEG-7 classification schemes are represented as individuals of the *ClassificationSchemeType* class, which represents the homonym MPEG-7 type that specifies the structure of the classification schemes.

The main advantage of the above methodology is that, thanks to the manual effort, an OWL-DL ontology has been produced, which accurately captures the semantics of the MPEG-7 constructs (including both the named and the unnamed – nested – ones).

2.5 Mapping Approach

The approach used to map XML schema constructs in the MPEG-7 standard to OWL constructs is based on a generic XML schema to OWL mapping combined with an XML to RDF translation. It has already shown its usefulness with other quite big XML schemas in the Digital Rights Management domain, such as MPEG-21 and ODRL (García, Gil and Delgado 2007), and also in the E-Business domain (García and Gil 2007).

The main contribution of this approach is that it exploits the great amount of metadata that has been already produced by the XML community. There are many attempts to move metadata from the XML domain to the Semantic Web. Some of them just model the XML tree using RDF primitives (Klein 2002). Others concentrate on modelling the knowledge implicit in XML language definitions, i.e. DTDs or the XML schemas, using Web ontology languages (Amann, Beer, Fundulak and Scholl 2002; Cruz, Xiao and Hsu 2004; Halevy, Ives, Mork and Tatarinov 2003). Finally, there are attempts to encode XML semantics integrating RDF into XML documents (Lakshmanan and Sadri 2003; Patel-Schneider and Simeon 2002).

None of the previous approaches facilitate an extensive transfer of XML metadata to the Semantic Web in a general and transparent way. Their main problem is that the implicit interpretation of XML schema in terms of RDF(S) and OWL semantics is not formalised when XML metadata instantiating this schema is mapped. Therefore, they do not benefit from XML semantics and produce RDF metadata almost as semantics-blind as the original XML. Or, on the other hand, they capture these semantics but they use additional ad hoc semantic constructs that produce less transparent metadata. Therefore, we have chosen the XML semantic reuse methodology (García 2006) implemented by the ReDeFer project. It combines an XML achema to Web ontology mapping, called XSD2OWL, with a transparent mapping from XML to RDF, XML2RDF. The ontologies generated by XSD2OWL are used during the XML to RDF step in order to generate semantic metadata that makes XML schema semantics explicit. Both steps are detailed next. To conclude, in order to improve the transfer from MPEG-7 XML metadata to the Semantic Web, there is also a simple MPEG-7 classification scheme to OWL mapping called CS2OWL. It maps MPEG-7 classification hierarchies, e.g. TV-anytime hierarchy of contents or formats, to an OWL hierarchy of classes.

2.5.1 XSD2OWL Mapping

The XML schema to OWL mapping is responsible for capturing the schema informal semantics. These semantics are derived from the combination of XML schema constructs. The mapping is based on translating these constructs to the OWL ones that best capture their meaning. These mappings are detailed in Table 2.1.

The XSD2OWL mapping is quite transparent and captures the semantics implicit in XML schema following the interpretations in Table 2.1. The same names used for

Table 2.1 XSD2OWL translations for the XML schema constructs and their interpretations in terms of the corresponding OWL constructs

<table>
<tr><th>XML schema</th><th>OWL</th><th>Shared semantics</th></tr>
<tr><td>element|attribute</td><td>rdf:Property
owl:DatatypeProperty
owl:ObjectProperty</td><td>Named relation between nodes or nodes and values</td></tr>
<tr><td>element@substitutionGroup</td><td>rdfs:subPropertyOf</td><td>Relation can appear in place of a more general one</td></tr>
<tr><td>element@type</td><td>rdfs:range</td><td>The relation range kind</td></tr>
<tr><td>complexType|group |attributeGroup</td><td>owl:Class</td><td>Relations and contextual restrictions package</td></tr>
<tr><td>complexType//element</td><td>owl:Restriction</td><td>Contextualised restriction of a relation</td></tr>
<tr><td>extension@base|restriction@base</td><td>rdfs:subClassOf</td><td>Package concretises the base package</td></tr>
<tr><td>@maxOccurs
@minOccurs
None specified</td><td>owl:maxCardinality
owl:minCardinality
owl:cardinality(1)</td><td>Restrict the number of occurrences of a relation. 1 implicit if not specified</td></tr>
<tr><td>Sequence Choice</td><td>owl:intersectionOf
owl:unionOf</td><td>Combination of relations in a context</td></tr>
</table>

XML constructs are used for the OWL ones, although in the new namespace defined for the ontology, XSD and OWL constructs names are identical. This usually produces uppercase-named OWL properties because the corresponding element name is uppercase, although this is not the usual convention in OWL.

Therefore, XSD2OWL produces OWL ontologies that make explicit the semantics of the corresponding XML schemas. The only caveats are the implicit order conveyed by *xsd:sequence* and the exclusivity of *xsd:choice*. For the first problem, *owl:intersectionOf* does not retain its operands' order; there is no clear solution that retains the great level of transparency that has been achieved. The use of RDF lists might impose order but introduces ad hoc constructs not present in the original metadata. Moreover, as has been demonstrated in practise, the element ordering does not contribute much from a semantic point of view. For the second problem, *owl:unionOf* is an inclusive union, and the solution is to use the disjointness OWL construct, *owl:disjointWith*, between all union operands in order to make it exclusive.

The XSD2OWL mapping has been applied to the MPEG-7 XML schemas producing the complete MPEG-7 ontology. This ontology has 2372 classes and 975 properties. The only adjustment that has been done to the automatically generated ontology is to resolve a name collision between an OWL class and an RDF property. This is due to the fact that XML has independent name domains for complex types and elements while OWL has a unique name domain for all constructs. Moreover, the resulting OWL ontology is OWL full because the XSD2OWL translator has been forced to employ *rdf:Property* for those *xsd:elements* that have both data type and object type ranges. Table 2.2 shows an example of an XML schema ComplexType mapping to the corresponding OWL class.

Table 2.2 XML schema to OWL mapping example (namespaces omitted for readability)

XML schema	OWL (abstract syntax)

`<complexType name="AudioType">`
`<complexContent>`
`<extension base="MultimediaContentType">`
`<sequence>`
`<element name="Audio" type="AudioSegmentType"/>`
`</sequence>`
`</extension>`
`</complexContent>`
`</complexType>`
	`Class (AudioType complete MultimediaContentType restriction(Audio allValuesFrom(AudioSegmentType) cardinality(1)))`

2.5.2 XML2RDF Mapping

Once all the XML schemas for the metadata under consideration are available as mapped OWL ontologies, it is time to map the XML metadata that instantiates them. The intention is to produce RDF metadata as transparently as possible. Therefore, a structure-mapping approach has been selected (Klein 2002). It is also possible to take a model-mapping approach (Tous, García, Rodríguez and Delgado 2005). XML model mapping is based on representing the XML information set using semantic tools. This approach is better when XML metadata is semantically exploited for concrete purposes. However, when the objective is semantic metadata that can be easily integrated, it is better to take a more transparent approach. Transparency is achieved in structure-mapping models because they only try to represent the XML metadata structure, i.e. a tree, using RDF. The RDF model is based on the graph so it is easy to model a tree using it.

Moreover, we do not need to worry about the loose semantics produced by structure mapping. We have formalised the underlying semantics into the corresponding ontologies and we will attach them to RDF metadata using the instantiation relation *rdf:type*.

The structure mapping is based on translating XML metadata instances to RDF ones that instantiate the corresponding constructs in OWL. The more basic translation is between relation instances, from *xsd:elements* and *xsd:attributes* to *rdf:Properties*. Concretely, owl:ObjectProperties for node to node relations and owl:DatatypeProperties for node to values relations. However, in some cases, it would be necessary to use *rdf: Properties* for *xsd:elements* that have both data type and object type values. Values are kept during the translation as simple types and RDF blank nodes are introduced in the RDF model in order to serve as source and destination for properties.

The resulting RDF graph model contains all that we can obtain from the XML tree. It is already semantically enriched, thanks to the *rdf:type* relation that connects each RDF property to the owl:ObjectProperty or owl:DatatypeProperty it instantiates. It can be enriched further if the blank nodes are related to the owl:Class that defines the package of properties and associated restrictions they contain, i.e. the corresponding *xsd:complexType*. This semantic decoration of the graph is formalised using *rdf:type* relations from blank nodes to the corresponding OWL classes.

At this point, we have obtained a semantics-enabled representation of the input metadata. The instantiation relations can now be used to apply OWL semantics to metadata. Therefore, the semantics derived from further enrichments of the ontologies, e.g. integration links between different ontologies or semantic rules, are automatically propagated to instance metadata, thanks to inference. We will show now how this mapping fits in the architecture for semantic multimedia metadata integration and retrieval.

However, before continuing to the next section, it is important to point out that these mappings have been validated in different ways. First, we have used OWL validators in order to check the resulting ontologies, not just the MPEG-7 ontology but also many others (García and Gil 2007; García, Gil and Delgado 2007). Second, our MPEG-7 ontology has been compared with handmade ontologies such as Jane Hunters' one (2001) and Tsinaraki et al.'s (2004b). This comparison has shown that our mapping captures all the semantics captured by these ontologies and even adds additional details not captured by them in order to get a full formalisation of the semantics in all the MPEG-7 XML schemas.

Finally, the two mappings have been tested in conjunction. Testing XML instances have been mapped to RDF, guided by the corresponding OWL ontologies from the used XML schemas, and then back to XML. Then, the original and derived XML instances have been compared using their canonical version in order to correct mapping problems.

2.6 Use Cases

Based on the previous XML to Semantic Web mapping, a system architecture that facilitates multimedia metadata integration and retrieval has been built. The architecture is sketched in Fig. 2.8. The MPEG-7 OWL ontology, generated by XSD2OWL, constitutes the basic ontological framework for semantic multimedia metadata integration and appears at the centre of the architecture. Other ontologies and XML schemas might be easily incorporated using the XSD2OWL module.

Semantic metadata can be directly fed into the system together with XML metadata, which is translated to semantic metadata using the XML2RDF module. XML MPEG-7 metadata has a great importance because it is commonly used for (automatically extracted) low-level metadata that constitutes the basic input of the system.

This framework has the persistence support of an RDF store, where metadata and ontologies reside. Once all metadata has been put together, the semantic integration can take place, as detailed in Section 2.6.1. Finally, from this

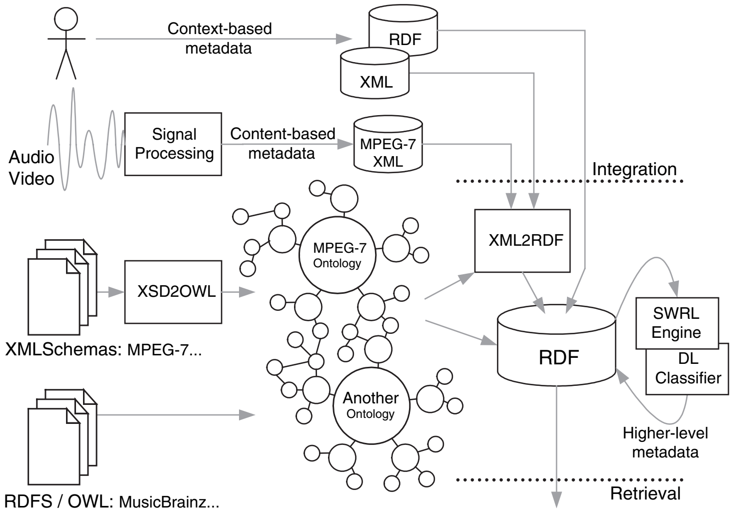

Fig. 2.8 Metadata integration and retrieval architecture

integrated space, higher-level metadata can be inferred and retrieved, as shown in Section 2.6.2.

2.6.1 Semantic Integration of Music Metadata

The problem of integrating heterogeneous data sources has grown in importance within the last years. One of the main reasons is the increasing availability of web-based data sources. Even within a single organisation, data from disparate sources must be integrated. Our approach to solve this problem is based on web ontologies. As we focus on the integration of multimedia assets, our base ontology is the MPEG-7 OWL ontology.

When multimedia metadata based on different schemes has to be integrated, the XML schemas are first mapped to OWL. Once this first step has been done, these schemas are easily integrated into the ontological framework using OWL semantic relations for equivalence and inclusion: *subClassOf*, *subPropertyOf*, *equivalentClass*, *equivalentProperty*, *sameIndividualAs*, etc. These relationships capture the semantics of the data integration. Then, once metadata is incorporated into the system and semantically enhanced, the integration is automatically performed by applying inference.

Our study on metadata integration is based on three different schemas: MusicBrainz schema, *Foafing the Music* ontology, and a music vocabulary to describe performances. MusicBrainz is a community music metadatabase that attempts to create a comprehensive music information site. MusicBrainz schema is written in RDF and describes all the tracks, albums, and artists available in their

Table 2.3 MusicBrainz to MPEG-7 OWL ontology mappings

musicbrainz:Artist	⊆	mpeg7:CreatorType
Musicbrainz:Album	⊆	mpeg7:CollectionType
Musicbrainz:Track	⊆	mpeg7:AudioSegmentType
dc:author	⊆	mpeg7:Creator
Dc:title	⊆	mpeg7:Title
musicbrainz:sortName	⊆	mpeg7:Name
musicbrainz:duration	≡	mpeg7:MediaDuration

music repository. Their mappings to the MPEG-7 OWL ontology are shown in Table 2.3.

The *foafing the music* ontology describes (low-level) content-based descriptors extracted automatically from the audio itself. The mappings of this schema to the MPEG-7 OWL ontology are summarised in Table 2.4. An artist is defined as a subclass of the MPEG-7 creator type, a track is defined as a subclass of the MPEG-7 AudioSegment and the audio descriptor class describes the content-based properties of a track. This descriptor is linked with the MPEG-7 AudioDS type. Thus, all *foafing the music* descriptors' subclasses inherit the properties from the MPEG-7 audio descriptor scheme. To characterise the descriptors related with the tonality of a song, the *Foafing the Music* ontology defines some properties, such as mode and key. Finally, the ontology defines rhythm descriptors to describe the rhythm component of a track, e.g. metre and tempo.

The last of the three schemas, a music vocabulary to describe performances, is linked as well with the MPEG-7 OWL (see Table 2.5). This schema models – for example, in the classical music world – a concert with the conductor, performers, the whole programme, time schedule, etc. The most general class related with a music piece is the Musical_Unit, from which all types of performances are derived (e.g. an opera performance, a symphony, a movement of the symphony).

Decomposition of a musical unit is achieved by defining its sections, and we link it with the MPEG-7 AudioSegment. Finally, there is an Artist class, the superclass for all the agents of the performances, e.g. director, musician, singer. Therefore, we link the Artist class with MPEG-7 OWL, and, automatically (transitivity property of rdfs:subClassOf) all the subclasses are linked with the MPEG-7 OWL ontology.

Table 2.4 Foafing the music ontology to MPEG-7 OWL ontology mappings

foafingthemusic:Artist	⊆	mpeg7:CreatorType
foafingthemusic:name	≡	mpeg7:GivenName
foafingthemusic:Track	⊆	mpeg7:AudioSegmentType
foafingthemusic:title	≡	mpeg7:Title
foafingthemusic:duration	≡	mpeg7:MediaDuration
foafingthemusic:Descriptor	≡	mpeg7:AudioDSType
foafingthemusic:mode	≡	mpeg7:Scale
foafingthemusic:key	≡	mpeg7:Key
foafingthemusic:tempo	≡	mpeg7:Beat
foafingthemusic:meter	≡	mpeg7:Meter

Table 2.5 Music vocabulary ontology to MPEG-7 OWL ontology mappings

music:Music_Unit	$\subseteq$	mpeg7:AudioSegmentType
music:sections	$\equiv$	mpeg7:AudioSegment
music:Artist	$\subseteq$	mpeg7:CreatorType
music:key	$\equiv$	mpeg7:Key
music:meter	$\equiv$	mpeg7:Meter

Once these mappings are done, all the multimedia assets are integrated into the ontological framework; that is the MPEG-7 OWL linked with all the schemas. Now, querying the system for audio segments will retrieve information from all the different sources, transparently to the user.

2.6.2 Semantic Retrieval of Music Metadata

Retrieving multimedia assets in the proposed architecture can be easily achieved by using semantic query languages such as the SPARQL query language (Prud'hommeaux and Seaborne 2007). SPARQL can benefit from the semantics made explicit by the XSD2OWL and XML2RDF mappings. It can, as well, exploit the results of semantic rules for metadata integration in order to retrieve all the related multimedia information for a given query. In our case, SPARQL queries use the MPEG-7 OWL ontology "vocabulary" in order to integrate all data sources. Using the mappings explained in the previous section, a SPARQL query can acquire information from *MusicBrainz*, *Foafing the Music*, the classical music ontology, etc.

A typical scenario that shows the usefulness of the architecture proposed could be the following: an Internet crawler is looking for audio data (we may assume that it is searching for MP3 files) and it downloads all the files. Getting editorial and related information for these audio files can be achieved by reading the information stored in the ID3 tag. Unfortunately, sometimes there is no basic editorial information such as the title of the track, or the performer.

However, content-based low-level descriptors can be computed for these files, including its MusicBrainz fingerprint, a string that uniquely identifies each audio file based on its content. The example in Table 2.6 shows an RDF/N3 description for a track with the calculated tempo and fingerprint.

On the other hand, the MusicBrainz database has the editorial metadata – as well as the fingerprint already calculated – for more than 3 million tracks. For example, the RDF description of the song "Blowin' in the wind" composed by Bob Dylan in Table 2.7.

Table 2.6 Content-based metadata, tempo, and fingerprint

```
<http://example.org/track#1> a foafingthemusic:Track;
   foafingthemusic:tempo "122";
   musicbrainz:trmid "e3c41bc1-4fdc-4ccd-a471-243a0596518f".
```

Table 2.7 Editorial metadata, title, and author, plus fingerprint

```
<http://example.org/track#2> a musicbrainz:Track;
  dc:title "Blowin' in the wind";
  dc:author [musicbrainz:sortName "Bob Dylan"];
  musicbrainz:trmid "e3c41bc1-4fdc-4ccd-a471-243a0596518f".
```

A closer look at both examples should highlight that the two resources are sharing the same MusicBrainz's fingerprint. Therefore, it is clear that, using a simple rule (1), one can assert that both audio files are actually the same file, that is to say the same instance in terms of OWL, owl:sameIndividualAs.

```
mpeg7:AudioType(track1) ∧ mpeg7:AudioType(track2) ∧
musicbrainz:trmid(track1, trm1) ∧
musicbrainz:trmid(track2, trm2) ∧ (trm1 = trm2)        (1)
⇒ owl:sameIndividualAs (track1, track2)
```

From now on, we have merged the metadata from both sources and we have deduced that the metadata related with both tracks is in fact referring to the same track. This data integration (at the instance level) is very powerful as it can combine and merge context-based data (editorial, cultural, etc.) with content-based data (extracted from the audio itself).

Finally, issuing a SPARQL query that searches for all the songs composed by Bob Dylan that have a fast tempo retrieves a list of songs, including "Blowin' in the wind". Moreover, there is no need for metadata provenance awareness at the end-user level. As the example in Table 2.8 shows, all query terms are referred only to the MPEG-7 ontology namespace.

Table 2.8 SPARQL query for integrated metadata retrieval

```
PREFIX mpeg7:<http://rhizomik.net/ontologies/2005/03/Mpeg7-2001.
owl#>
SELECT ?title
WHERE {
  ?track  a                mpeg7:AudioSegmentType;
          mpeg7:Title      ?title;
          mpeg7:Beat       ?tempo;
          mpeg7:Creator    ?author.
  ?author mpeg7:Name       "Bob Dylan".
  FILTER (?tempo >= 60) }
ORDER BY ASC(?title)
```

2.7 An Integrated Ontological Infrastructure for the Semantic Description of Multimedia Content

In this section, we present the *DS-MIRF ontological infrastructure*, an integrated ontological infrastructure for the semantic description of multimedia content that allows for the systematic integration of domain knowledge within the MPEG-7 semantics. This infrastructure was developed in the context of the *DS-MIRF framework* (Tsinaraki et al. 2007), which facilitates the development of knowledge-based multimedia content services based on the MPEG-7/21 standards. The DS-MIRF ontological infrastructure can support different usage scenarios, which fall into two main categories:

- The usage scenarios where the DS-MIRF ontological infrastructure is used in order to guide MPEG-7-based semantic multimedia content annotation and/or semantic multimedia service provision on top of an OWL/RDF repository. In this case, the OWL/RDF semantic multimedia annotations are produced (manually, automatically, or semi-automatically), possibly after being enriched through the application of rule-based reasoning, and stored in the repository.
- The usage scenarios where the DS-MIRF ontological infrastructure is used in order to guide MPEG-7-based semantic multimedia content annotation and/or semantic multimedia service provision on top of a pure MPEG-7 repository. In this case, the OWL/RDF semantic multimedia annotations that are produced, possibly after being enriched through reasoning, are transformed into pure MPEG-7 descriptions and are then stored in the repository. This category of usage scenarios is extremely useful both for groups using pure MPEG-7 and for groups sharing pure MPEG-7 descriptions with their partners. Full support for this category of usage scenarios is provided by the DS-MIRF framework.

The ontological infrastructure of the DS-MIRF framework (depicted in Fig. 2.9) includes an *OWL-DL upper ontology*, *OWL-DL application ontologies*, and *OWL-DL domain ontologies*.

The *OWL-DL upper ontology* fully captures the semantics of the MPEG-7 MDS and the *MPEG-21 DIA architecture* (ISO/IEC 2004) and the parts of the MPEG-7 visual and audio that are necessary for the complete representation of the MPEG-7 MDS. This ontology includes the MPEG-7 OWL-DL ontology described in Section 2.4, extended with the MPEG-21 DIA architecture semantics in order to better support multimedia content personalisation and adaptation.

The *OWL application ontologies* either enhance, using OWL-DL syntax, the semantics of MPEG-7/21 so that the users find it easier to use MPEG-7/21 or allow using advanced multimedia content services that cannot be directly supported by MPEG-7/21. The application ontologies provide general purpose constructs that either are not available in MPEG-7/21 (for example, semantic user preferences) or are implied in the text of MPEG-7/21 but lacking in their syntax (for example, typed relationships).

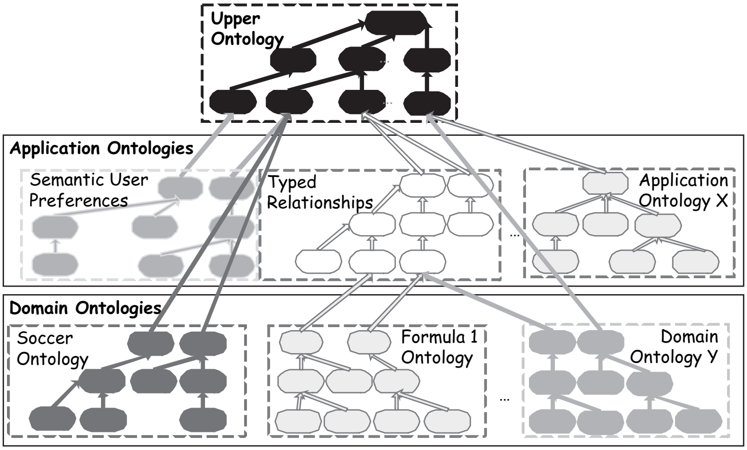

Fig. 2.9 The ontological infrastructure of the DS-MIRF framework

The *domain ontologies* systematically extend the upper ontology and the application ontologies with domain knowledge (for example, sports ontologies that extend the abstract semantic description capabilities of the MPEG-7 MDS).

In the rest of this section, we present the application ontologies that have been already integrated in the DS-MIRF ontological infrastructure (in Section 2.7.1), the methodology followed for domain knowledge representation in the form of OWL domain ontologies and their integration with the MPEG-7 semantics (in Section 2.7.2), and the DS-MIRF framework and the support it provides for pure MPEG-7 applications (in Section 2.7.3).

2.7.1 Application Ontologies

We outline in this section the application ontologies that have already been integrated in the DS-MIRF framework. These include (a) a *typed relationship* application ontology, which extends the MPEG-7 MDS in order to allow the full and systematic representation of typed relationships that are literally described in the MPEG-7 MDS text but their features are not fully captured in the MPEG-7 MDS syntax and (b) a *semantic user preference* application ontology that supports the semantic-based description of the desired multimedia content, which is not allowed in the MPEG-7 user preferences. The application ontologies are described in the following paragraphs.

2.7.1.1 Typed Relationship Application Ontology

The *typed relationship* application ontology assists the semantic multimedia content description. It extends the MPEG-7 MDS in order to allow the full and systematic representation of typed relationships that are literally described in the MPEG-7

MDS text but their features are not fully represented in the MPEG-7 MDS syntax. The typed relationship ontology is an application ontology that can greatly facilitate application development by the users in the large majority of cases. The users are not forced to use this ontology, but if they do so, the definition of relationships in MPEG-7 metadata descriptions becomes much easier.

The semantics of the typed relationships are partially covered in the MPEG-7 MDS syntax in the *GraphRelation, SpatialRelation, SemanticRelation, BaseRelation*, and *TemporalRelation* classification schemes. The representation of the relationship types in the form of classification scheme terms does not allow for expressing formally whether a relationship is directed and, if so, which is its inverse relationship; this information is available only in the textual description of the relationship type.

The typed relationship ontology, depicted in Fig. 2.10, extends the upper ontology with an OWL class hierarchy rooted in the *TypedRelationType* (which is a subclass of the *RelationType* class of the upper ontology that represents relationships). The direct subclasses of *TypedRelationType* are homonyms of the classification schemes where the relationship types are defined. Each of the subclasses of *TypedRelationType* has a number of subclasses, which correspond to the relationship types defined in the homonym classification scheme, together with the information literally described about them in the MPEG-7 MDS text. This information includes the type of the relationship, if it is directed or not and, in the latter case, its inverse relationship. The annotator that uses the typed relationship application ontology does not have to be aware of the textual description of the MPEG-7 MDS, since all the information is captured in the ontology.

The OWL classes of the typed relationship ontology formally capture all the information about the typed relationships that exist in the MPEG-7 MDS text. In fact, they express formally the semantics that exist in the textual descriptions of the different relationship types (for example, that the *before* relationship is directed and that the *after* relationship is its inverse).

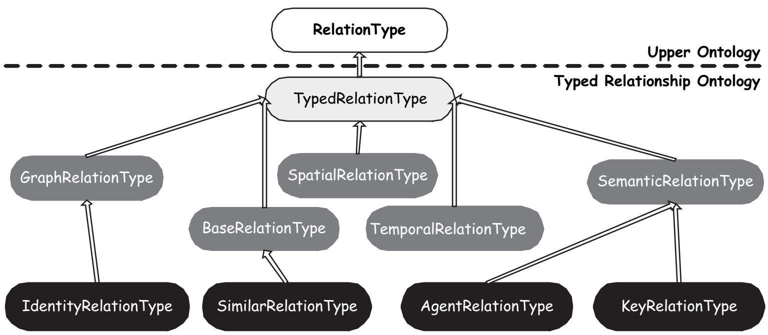

Fig. 2.10 The typed relationship ontology

2.7.1.2 Semantic User Preference Application Ontology

The *semantic user preference* application ontology allows the semantic-based description of the desired multimedia content in the user preferences. Such an extension of MPEG-7/21 is needed because the MPEG-7/21 user preference descriptions allow keyword-only descriptions of the semantics of the preferred content. As an example, consider a user who wishes to receive all the images that contain a teacher who gives a book to a student, as soon as such images are available. The current MPEG-7 search and filtering preference descriptions allow the users to describe the desired images using the keywords *teacher*, *student*, *gives*, and *book*. These user preference descriptions will provide, together with the images that contain a teacher who gives a book to a student, images that contain a student who gives a book to a teacher.

The application ontology is based on the semantic user preference model proposed in Tsinaraki and Christodoulakis (2007), which is also compatible with the MP7QL and allows for the explicit specification of the Boolean operators to be used in the different phases of multimedia content search and filtering. The semantic user preferences structured according to this model allow the accurate expression of the user preferences of a user who wishes to receive all the images that contain a teacher who gives a book to a student.

2.7.2 Domain Knowledge Representation and Integration with MPEG-7

The multimedia content description approaches, which have been implemented in MPEG-7 and have been described in Section 2.2, are general purpose and can be applied in any domain. In particular, the general purpose semantic description capabilities of MPEG-7 distinguish only events, agents (people, person groups, and organisations), places, states, times, objects, and concepts. On the other hand, the systematic integration of domain knowledge in the multimedia content descriptions has been shown to enhance the retrieval effectiveness of the multimedia content retrieval and filtering services built on top of them. We outline in this section a methodology for domain knowledge representation in OWL and its integration with MPEG-7 semantics (Tsinaraki et al. 2007; Tsinaraki, Polydoros and Christodoulakis 2004a). This methodology has been developed in the DS-MIRF framework for the definition and integration of domain ontologies in the DS-MIRF ontological infrastructure. Thanks to this methodology, OWL/RDF multimedia content descriptions can be defined that are structured according to MPEG-7 semantics and are also enhanced with domain knowledge.

According to this methodology, the domain-specific entities are represented as domain ontology classes. These classes are (direct or indirect) subclasses of the OWL classes that represent the subtypes of *SemanticBaseType* (*EventType, ObjectType, AgentObjectType, SemanticPlaceType, SemanticTimeType, Semantic-StateType*, and *ConceptType*) in the upper OWL-DL ontology defined in Tsinaraki

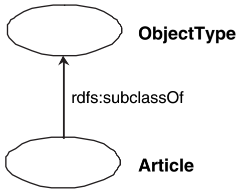

Fig. 2.11 RDF graph showing the *Article* class, which represents articles

et al. (2004b). This way, the knowledge captured in the domain ontologies is integrated with the MPEG-7 semantic model. As an example, the *Article* class (shown in Fig. 2.11), which represents the articles, should be defined as a subclass of the *ObjectType* class.

Features that are not present in the upper ontology class are represented as additional object or data type properties in its domain-specific subclass. For example, the number of pages of an article should be represented as a data type property of non-negative integer type in the domain of the *Article* class.

Additional constraints may be applied on the properties inherited from the ancestor classes, in order to guide the indexers to produce valid metadata (for example, the author of an article should have a name).

In addition, properties may be defined that permit the attachment of relationships to the allowed domain-specific entities only (for example, only persons are allowed to be related with articles as authors). These properties are subproperties of the *Relation* property of the *SemanticBaseType* class, which links semantic entities with relationships. The properties have as domain the union of the classes to which belong individuals that are capable of being sources of a typed relationship and the typed relationship class as range. The inverse property of the one defined previously is defined in the domain of the classes the individuals of which are capable of being targets of the typed relationship.

The methodology described above can be also used in order to integrate existing OWL domain ontologies in the MPEG-7 semantics. It has been tested in the DS-MIRF framework through the definition of domain ontologies for soccer and Formula 1 and their integration with the DS-MIRF ontological infrastructure.

2.7.3 The DS-MIRF Framework

The architecture of the DS-MIRF framework and the information flow between its components are depicted in Fig. 2.12.

The multimedia content *annotator* is a special type of user in the DS-MIRF framework that is responsible for the semantic annotation of multimedia documents. He uses a *multimedia annotation interface* that makes use of the ontological infrastructure of the DS-MIRF framework in order to support ontology-based semantic annotation of the multimedia content. The DS-MIRF framework ontologies are expressed in OWL; thus the result of the annotation process is an OWL description

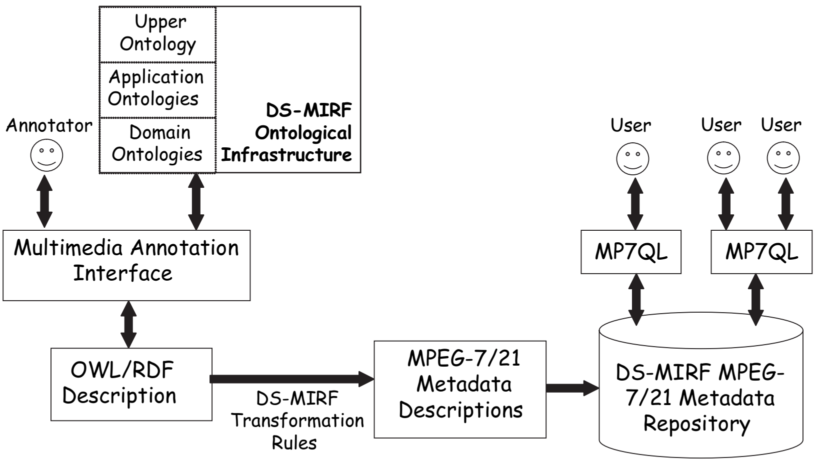

Fig. 2.12 The DS-MIRF framework – architecture and information flow

of the multimedia content. The OWL descriptions are then transformed, using the *DS-MIRF transformation rules* to standard *MPEG-7/21 metadata descriptions*. The MPEG-7/21 metadata are stored in the *DS-MIRF MPEG-7/21 metadata repository*, which is accessed by the end-users through application interfaces that are based on the *MPEG-7 query language* (*MP7QL*) (Tsinaraki and Christodoulakis 2007), a query language that has been developed in the context of the DS-MIRF framework for querying MPEG-7 multimedia descriptions.

In the following paragraphs, we will focus on the MP7QL query language and on the support for interoperation with applications using pure MPEG-7.

2.7.3.1 The MP7QL Query Language

The MP7QL query language has MPEG-7 as data model and allows for the querying of every aspect of an MPEG-7 multimedia content description, including semantics, low-level features, and media-related aspects. It also allows for the exploitation of domain knowledge encoded using pure MPEG-7 constructs. In addition, it allows the explicit specification of Boolean operators and/or preference values. The MP7QL queries may utilise the users' *filtering and search preferences* (*FASP*) and *usage history* as context, thus allowing for personalised multimedia content retrieval. The MP7QL has been expressed both in XML schema and in OWL-based syntax, in order to be applicable to all usage scenarios and working environments. The XML schema-based syntax of the MP7QL is used in the current implementation of the DS-MIRF framework.

General purpose languages, such as XQuery in the pure MPEG-7 environment and SPARQL in the Semantic Web environment, do not take into account the following peculiarities of the MPEG-7 description elements: (a) the MPEG-7 semantic model is expressed in a rather complex way; (b) the domain knowledge integrated in the semantic MPEG-7 descriptions is expressed in the document level; and (c) the low-level visual and audio features should be evaluated using specialised functions. Thus, in order to fully exploit the semantics of the MPEG-7 descriptions, a query language for querying MPEG-7 descriptions is needed, with clear, MPEG-7 specific semantics (instead of the generic semantics of XQuery and SPARQL). These semantics will also allow the optimisers to effectively perform consistency checking and first-level optimisation. The MP7QL fulfils the requirement for MPEG-7 semantics, as it has MPEG-7 as its data model.

2.7.3.2 Support for Interoperation with Applications Using Pure MPEG-7

Interoperation of the multimedia content descriptions with applications using pure MPEG-7 is achieved through the DS-MIRF transformation rules that allow the transformation of domain ontologies and semantic content descriptions to valid MPEG-7 descriptions. In particular, they allow the transformation of (a) domain ontologies defined according to the methodology described in Tsinaraki et al. (2005) into abstract MPEG-7/21 semantic descriptions; (b) OWL individuals that belong to the domain ontology classes into MPEG-7/21 semantic descriptions. The descriptions which are produced are valid MPEG-7/21 (parts of) documents.

During the metadata transformation from OWL to MPEG-7/21, the individuals representing MPEG-7/21 constructs are transformed into XML elements. The object properties are transformed into elements and the data-type properties are transformed into the constructs they represent in the original MPEG-7/21 schemas (attributes, elements, or simple values). In order to produce valid MPEG-7/21 descriptions, information regarding the MPEG-7/21 XML element order, the default values and the original MPEG-7/21 representation of the data type properties are needed. This information is kept in a mapping ontology and is utilised during both ontology and metadata transformations, as shown in Fig. 2.13.

The classes of the OWL domain ontologies and the OWL individuals belonging to them are both transformed into instances of the subtypes of *SemanticBaseType*. This way, the domain knowledge is represented in a way compatible with the domain knowledge expressed according to the methodology presented in Tsinaraki et al. (2005). The *AbstractionLevel* element of the *SemanticBaseType* and the MPEG-7 semantic relationships are used to capture the ontology semantics.

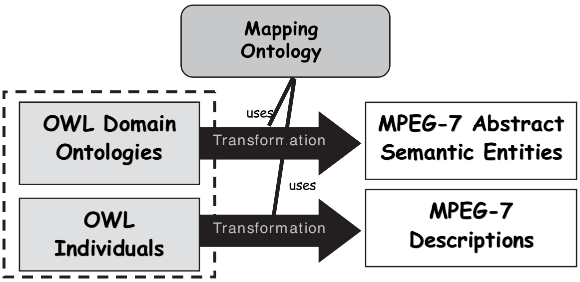

Fig. 2.13 OWL-MPEG-7 transformations in the DS-MIRF framework

An abstract semantic entity that represents a domain-specific class has a non-zero *AbstractionLevel.Dimension* and is related with the semantic entities that represent its subclasses through (a) a relationship of type *generalizes*, which has as source the semantic entity that represents the class and as target the semantic entity that represents the subclass, and (b) a relationship of type *specializes*, which has as source the semantic entity that represents the subclass and as target the semantic entity that represents the class. In addition, an abstract semantic entity that represents a class is related with each of the semantic entities representing the class individuals through pairs of *exemplifies/exemplifiedBy* relationships.

The data-type properties of the classes of the domain ontologies are transformed into *Property* elements and the object properties into pairs of *property/propertyOf* relationships.

The transformations outlined above allow the representation of the OWL class hierarchy and the preservation of the class properties. Several OWL axioms, especially the restrictions and the set operations, cannot be expressed in the pure MPEG-7 syntax. This feature does not allow the use of reasoning on top of the pure MPEG-7 descriptions, but it allows the systematic use of the domain knowledge that is expressed using MPEG-7 constructs. As a consequence, queries of the form "Give me the multimedia objects that show a teacher who gives a book to a student" can be expressed accurately, instead of searching in the textual parts of the MPEG-7 description elements (including the semantic ones) for the keywords *teacher*, *student*, *gives*, and *book*. The latter query is also ambiguous, as it will also return the multimedia objects where a student gives a book to a teacher. In addition, it may evaluate as teachers semantic entities that represent people who have worked as teachers for a while and the keyword *teacher* exists in the textual annotation of the semantic entities.

Another approach for domain knowledge representation is the definition of subtypes of the MPEG-7 types that represent semantic entities in order to represent domain-specific classes. The advantage of the utilisation of abstract semantic entities instead of subtypes of the semantic entity types for the representation of domain-specific classes is that this way, full compatibility with MPEG-7 is maintained so that all the tools and the applications that use pure MPEG-7 still work transparently with the MPEG-7 descriptions which are produced.

The ontological infrastructure of the DS-MIRF framework and the mechanisms that have been developed in the context of DS-MIRF for the support of interoperability between OWL domain ontologies and MPEG-7/21 support the semantic multimedia content description, which in turn allows the provision of advanced (semantic) multimedia content services. In particular, advanced retrieval services can be supported on top of the semantic multimedia annotations, which allow more accurate multimedia content retrieval. Accurate retrieval results in the better support of advanced services built on top of it, such as filtering and content-based personalisation. Such services are offered in the DS-MIRF framework based on the MP7QL language that supports personalised semantic retrieval and filtering on top of MPEG-7 multimedia content descriptions.

2.8 Conclusions

In this chapter, we have introduced the need for representing MPEG-7 constructs using Semantic Web languages. First, we presented in Section 2.2 the MPEG-7 standard and described the well-accepted general purpose approaches for multimedia content description that are supported by the MPEG-7 standard and their implementation in MPEG-7. Then, we have presented the Semantic Web languages in Section 2.3. The research efforts towards the expression of MPEG-7 using Semantic Web languages have been outlined in Section 2.4, followed by the mapping of the MPEG-7 constructs to OWL constructs in Section 2.5. The application of two of the MPEG-7 ontologies in real application environments have been presented next: (a) a case study for the music domain has been presented in Section 2.6, which has introduced the problems of annotating multimedia assets, integrating data from different sources, and retrieving music-related descriptors, and (b) an integrated ontological infrastructure for the semantic description of multimedia content which allows for combining the general purpose MPEG-7 constructs with domain and application-specific knowledge has been described in Section 2.7.

Both application scenarios show the benefits of the MPEG-7 formal semantics. MPEG-7 is a big standard, difficult to deal with, but the availability of some formal semantics facilitates the development of more advanced tools capable of dealing with its complexity.

References

The aceMedia project (2007) *Harmonization of Multimedia Ontologies Activity,* in aceMedia annual public report 2006.

Antoniou, G. and Harmelen, F. van. (2004) *A Semantic Web Primer*. The MIT Press, April.

Amann, B., Beer, C., Fundulak, I. and Scholl, M. (2002) *Ontology-Based Integration of XML Web Resources*. In Proceedings of the 1st International Semantic Web Conference (ISWC).

Berners-Lee, T. (1999) *Weaving the Web*. Texere Publishing Ltd., November.

Berners-Lee, T, Hendler, J. and Lassila, O. (2001) The semantic web. *Scientific American*, May.

Brickley, D. and Guha, R.V. (eds.) (2004) *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C Recommendation.

Celma, O., Gómez, E., Janer, J., Gouyon, F., Herrera, P. and García, D. (2004) *Tools for Content-Based Retrieval and Transformation of Audio Using MPEG-7: The SPOffline and the MDTools*. Proceedings of 25th International AES Conference London, UK.

Celma O., Dasiopoulou S., Hausenblas M., Little S., Tsinaraki C. and Troncy R. (2007) *MPEG-7 and the Semantic Web*. W3C Multimedia Semantics Incubator Group (MMSEM-XG), XG deliverable.

Cruz, I., Xiao, H. and Hsu, F. (2004) *An Ontology-Based Framework for XML Semantic Integration*. In Proceedings of the Eighth International Database Engineering and Applications Symposium. Coimbra, Portugal (IDEAS’04).

Fallside, D. (2001) *XML Schema Part 0: Primer*. W3C Recommendation.

García, R. and Celma, O. (2005) *Semantic Integration and Retrieval of Multimedia Metadata*. In the Proceedings of the Knowledge Markup and Semantic Annotation Workshop, Semannot’05.

García, R. (2006) *A Semantic Web Approach to Digital Rights Management*. PhD thesis, Technologies Department, Universitat Pompeu Fabra.

García, R. and Gil, R. (2007) *Facilitating Business Interoperability from the Semantic Web*. In Proceedings of the 10th International Conference on Business Information Systems, BIS'07. Springer-Verlag, Lecture Notes in Computer Science, (4439):220–232.

García, R., Gil, R. and Delgado, J. (2007) *A Web Ontologies Framework for Digital Rights Management*. Journal of Artificial Intelligence and Law, 15(2):online first

Gruber, T. R. (1993) Towards principles for the design of ontologies used for knowledge sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, Kluwer Academic Publishers.

Halevy, A., Ives, Z., Mork, P. and Tatarinov, I. (2003) *Piazza: Data Management Infrastructure for Semantic Web Applications*. In Proceedings of the 12th World Wide Web Conference. Budapest, Hungary, 556–567.

Hunter, J., (1999) *A proposal for an MPEG-7 description definition language*. Technical report, MPEG-7 AHG Test and Evaluation Meeting, Lancaster, UK

Hunter, J., (2001) *Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology*. In the Proceedings of the International Semantic Web Working Symposium (SWWS).

ISO/IEC (2001a) 15938-3:2001: Information Technology – Multimedia content description interface – Part 3 visual. Version 1.

ISO/IEC (2001b) 15938-4:2001: Information Technology – Multimedia content description interface – Part 4 audio. Version 1.

ISO/IEC (2003) 15938-5:2003: Information Technology – Multimedia content description interface – Part 5: Multimedia description schemes. First Edition.

ISO/IEC (2004) 21000-7:2004, Information Technology – Multimedia Framework (MPEG-21) – Part 7: Digital Item Adaptation.

Klein, M.C.A. (2002) *Interpreting xml Documents via an rdf Schema Ontology*. In Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA).

Lakshmanan, L. and Sadri, F. (2003) *Interoperability on XML Data*. In Proceedings of the 2nd International Semantic Web Conference (ICSW 03), pp. 146–163.

Manola, F. and Milles, E. (eds.) (2004) *RDF Primer*. W3C Recommendation.

Martínez, J. (ed.) (2004), *MPEG-7 Overview*, ISO/IEC JTC1/SC29/WG11 N6828, Palma de Mallorca, October.

McGuinness, D. L. and van Harmelen, F. (eds.) (2004) *OWL Web Ontology Language: Overview*. W3C Recommendation.

McGuinness, D.L., Fikes, R., Hendler, J. and Stein. L.A. (2002) *DAML+OIL: An Ontology Language for the Semantic Web*. IEEE Intelligent Systems, September–October, 17(5):72–80

Patel-Schneider, P. and Simeon, J. (2002) *The Yin/Yang Web: XML Syntax and RDF semantics*. In Proceedings of the 11th International World Wide Web Conference (WWW02), Honolulu, Hawaii, USA, 443–453.

Prud'hommeaux, R. and Seaborne A. (2007). *SPARQL Query Language for RDF*. W3C Working Draft, 26 March.

Salembier, P., Manjunath, B.S. and Sikora T. (eds.) (2002). *Introduction to MPEG 7: Multimedia Content Description Language*. Ed. Wiley.

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) *The Semantic Web Revisited*. IEEE Intelligent Systems, 21(3):96–101.

Siméon, J., Chamberlin, D., Fernández, M., Boag, S., Florescu, D. and Robie, J. (Eds) (2007) *XQuery 1.0: An XML Query Language*. W3C Recommendation, 23 January.

Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G. and Kollias, S. (2005) *A Visual Descriptor Ontology for Multimedia Reasoning*. In the Proceedings of the Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS).

Tous, R., García, R., Rodríguez, E. and Delgado, J. (2005) *Arquitecture of a Semantic XPath Processor. Application to Digital Rights Management (DRM)*. In the Proceedings of E-Commerce and Web Technologies: 6th International Conference, EC-Web. Springer-Verlag, LNCS (3590):1–10

Tsinaraki, C. and Christodoulakis, S. (2007) *An MPEG-7 Query Language and a User Preference Model that allow Semantic Retrieval and Filtering of Multimedia Content*. ACM-Springer Multimedia Systems Journal, in Special Issue on Semantic Multimedia Adaptation and Personalization, (to appear)

Tsinaraki, C., Polydoros, P. and Christodoulakis, S. (2007) *Interoperability Support Between MPEG-7/21 and OWL in DS-MIRF*. Transactions on Knowledge and Data Engineering (TKDE), Special Issue on the Semantic Web Era, pp. 219–232.

Tsinaraki, C., Polydoros, P. and Christodoulakis, S. (2004a) *Integration of OWL Ontologies in MPEG-7 and TV-Anytime Compliant Semantic Indexing*. In the Proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAISE), pp. 398–413.

Tsinaraki, C., Polydoros, P. and Christodoulakis, S. (2004b) *Interoperability Support for Ontology-Based Video Retrieval Applications*. In the Proceedings of the Conference on Image and Video Retrieval (CIVR), pp. 582–591.

Tsinaraki, C., Polydoros, P. Kazasis, F. and Christodoulakis, S. (2005) *Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content*. In Multimedia Tools and Application Journal (MTAP), Special Issue of on Video Segmentation for Semantic Annotation and Transcoding, 26, pp. 299–325, August.

Tzouvaras, V. and Pan, J. (2007) *Multimedia Annotation Interoperability Framework*. W3C Multimedia Semantics Incubator Group (MMSEM-XG) XG Deliverable.