

Lifting a Metadata Model to the Semantic Multimedia World

Gaëtan Martens*, Ruben Verborgh*, Chris Poppe* and Rik Van de Walle*

Abstract—This paper describes best-practices in lifting an image metadata standard to the Semantic Web. We provide guidelines on how an XML-based metadata format can be converted into an OWL ontology. Additionally, we discuss how this ontology can be mapped to the W3C's Media Ontology. This ontology is a standardization effort of the W3C to provide a core vocabulary for multimedia annotations. The approach presented here can be applied to other XML-based metadata standards.

Keywords—Multimedia, Metadata Annotation, Semantic Web Technologies

1. INTRODUCTION

In the last decade, digital imaging has experienced a worldwide revolution of growth in both the number of users and the range of applications that have replaced traditional film photography. However, despite its great value, multimedia metadata has yet to find its way into standard use [1]. Different image metadata standards currently exist. However, from a consumer perspective, most of them are too simple to satisfy needs or too complex for actual use.

The DIG35 standard, from an end-user perspective, is a metadata standard used to describe a set of public metadata for digital still images [2]. As this standard is modeled in XML, the semantic interpretation of the XML-tags can be subjective which interferes with exchanging and managing photo collections as well as reasoning about the annotations. Semantic Web technologies serve as an approach to overcome those difficulties as they permit reasoning about and relating to customized concepts.

In this paper, we describe how to lift the DIG35 standard into the Semantic Web and how to map it with current standardization efforts to the W3C [3, 4].

Section 2 gives an overview of existing image metadata standards and the DIG35 standard is compared with the other image metadata standards. Section 3 elaborates on the difficulties that arise when converting the XML Schema into an OWL DL ontology [5] and provides some solutions. In Section 4 the obtained ontology is presented. Section 5 will deal with the conversion of XML instances into OWL instances. Next, Section 6 holds a description of the MA Ontology. We will show that there is some overlap between this ontology and DIG35 and discuss some of the relationships in more detail in Section 7. Section 8 evaluates the

‡ The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

Manuscript received August 20, 2010; accepted September 14, 2010.

Corresponding Author: Chris Poppe

* Multimedia Lab, Dept. of Electronics and Information Systems, IBBT, Ghent University, Belgium ({gaetan.martens, ruben.verborgh, chris.poppe, rik.vandewalle}@ugent.be)

mappings. Finally, a use case and some conclusions appear in Section 9 and Section 10, respectively.

2. OVERVIEW OF IMAGE STANDARDS

The Exchangeable Image File Format (**EXIF** [6]) is the specification used by almost all digital cameras nowadays. The metadata tags provided by the EXIF standard cover metadata related to the capturing process of the image. Recently, there have been efforts to represent the EXIF metadata tags in an RDF Schema [7] ontology [8, 9].

MPEG-7, developed by the Motion Pictures Expert Group (MPEG), offers a comprehensive set of audiovisual description tools [10]. MPEG-7 provides rich and 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 it difficult to cope with the interpretation of high-level descriptions.

The International Press Telecommunications Council (**IPTC**) Core Schema is a metadata set primarily for photographers' use and aligns with the IPTC Headers¹. These IPTC photo metadata categorize the metadata fields into four groups regarding the semantics they describe, namely, Administrative, Content Descriptive, Rights and Technical metadata [11].

Other standards for describing metadata of digital images using RDF, are **Dublin Core** (DC [12]) and **PhotoRDF** [13]. DC consists of a rudimentary set of 15 elements describing common properties of resources, such as title and creator. PhotoRDF is an attempt to standardize a set of categories and labels for personal photo collections using the DC schema as well as an additional schema for technical and content data. The content schema contains 10 fixed keywords to be used in the subject property of the DC schema, such as "Portrait" and "Baby."

VRA Core is a set of metadata elements used to describe works of visual culture as well as the images which they represent. In the context of VRA Core, a work is a physical entity that exists, has existed at some time in the past, or that could exist in the future.

The **DIG35** standard, a specification of the International Imaging Association (I3A), defines a set of public metadata for digital images covering a broad spectrum of metadata fields [2]. The standard consists of five logical blocks. The first metadata block covers basic image parameters to specify generic information about the image, such as file name, format and size. The creation metadata is grouped in a second block. It defines the metadata that are relevant to the creation of the digital image data. Many fields defined in this block are actually the same as those defined by EXIF. The third container comprises the content description of an image. While many image metadata standards only provide limited support to describe the content, e.g., by keywords or a free text description, DIG35 offers fine grained description methods to model and relate depicted concepts within an image. The history metadata block holds partial information about how the image came to be in its present state. Finally, the last block defines the intellectual property rights metadata (IPR) which are designed to protect the contents of an image file from misuse and must preserve both moral rights and copyrights. Table 1 lists a comparison between DIG35 and other image metadata standards regarding the different metadata types they cover.

¹ IPTC Metadata, <http://www.iptc.org/>

Table 1. Comparison of image metadata standards

	PhotoRDF	Exif	VRA Core	IPTC Photo	MPEG-7	DIG35
Low level	-	-	-	-	✓	-
Creational	limited	✓	limited	✓	✓	✓
Technical	-	✓	-	✓	-	✓
Content	<i>keywords</i>	<i>title</i>	✓	✓	✓	✓
History	-	-	-	-	✓	✓
IPR	<i>copyright</i>	<i>copyright + holder</i>	<i>description &rights</i>	✓	✓	✓

3. ONTOLOGY CREATION

As part of the plethora of metadata formats, DIG35 consists of an XML Schema accompanied with plain text. Not all semantic information can be deduced from the XML structure; the accompanying plain text describes the semantical meaning of the XML and even some supplementary semantics. Another negative consequence related to the use of XML is its interoperability problems between different metadata standards [7]. A possible solution is to use formal representations by applying Semantic Web Technologies. This solution is also proposed by working groups of the W3C, such as the W3C Multimedia Semantics Incubator Group² and the W3C Media Annotations WG. In the remainder of this section, we give an overview of some common difficulties that can arise when converting an XML Schema into an OWL ontology and propose some solutions which we applied during the modeling of the DIG35 ontology.

3.1 Identifiers

According to the DIG35 specifications, many concepts can have an associated identifier. Expressing the concept of an identifier (cf. a key in a relational database system) in OWL can be accomplished using the *OWL:InverseFunctionalProperty* [14]. However, there is a restriction in that the latter property type is only applicable to properties of objects. Consequently, if one wants to use a literal value as an identifier (ID), an additional ID-class needs to be created. This ID-class holds a data type property *uid* to refer to the value of the ID. A drawback of this approach is that some extra complexity is added since one ends up with the definitions for 2 properties and an extra class in order to model one single concept.

3.2 Order of Appearance

The ‘Image Processing Hints’ in the History metadata section of DIG35 specifies a list of the operations applied when editing an image. Consequently, the order in which the operations are listed is of great importance. RDF provides two container mechanisms to encapsulate data in a user defined order: the *rdf:Seq* class which is used for representing ordered lists of literals or resources and *rdf:List* which is a class for representing a closed list of items. However, those predefined classes are not OWL DL compatible. To formalize the order in which an object or literal appears in a list we propose the creation of an item-class with two properties representing the order and the value. The first property, a data type property (order), defines the order of the

² World Wide Web Consortium (W3C) Multimedia Semantics Incubator Group, <http://www.w3.org/2005/Incubator/mmsen/>

object and the second property (value) refers to the object itself. The item-class is to be referred to by the *OWL:ObjectProperty* with no cardinality restrictions. However, the constraint that the value of the order property must be unique within the scope of the list cannot be expressed in OWL. So, in OWL DL only a partial solution for this problem can be devised.

3.3 Reusing existing Ontologies

Many existing ontologies are OWL Full which renders them difficult to reuse and import in existing ontologies [15]. Examples of some popular OWL Full ontologies are Friend of a Friend (FOAF [16]), DC and Simple Knowledge Organization Schema (SKOS [17]). In most cases, ontologies are rather OWL Full due to syntactic errors or accidental misuse of the vocabulary, such as the use of RDF-properties instead of OWL-properties. On the other hand for SKOS, the restrictions on OWL DL prevent treating SKOS concepts as OWL classes. Since a SKOS concept is defined as an OWL class, an instance of a concept should also be an OWL instance, but according to SKOS, an instance must be treated as a class. The latter statement is only supported by OWL Full. Meta-modeling, i.e., the treatment of classes, properties and other entities as individuals is partially allowed by OWL 2.0: a name can be used for any or all individuals, classes, or properties. For DC, there is an OWL DL version available on the Protégé website. However, a DC property is modeled using an *owl:AnnotationProperty*. The range of the latter must be an individual, literal, or URI which makes it impossible to refine an existing OWL DL ontology using this version of DC. For example, it is not possible to express that the datatype property *fileFormat* is an equivalent property or sub-property of *dc:format* or that the range of *dc:creator* is the class *Person*. As a result of the OWL Fullness of much ontology we did not reuse any existing ontologies during the modeling process in order to keep the DIG35 ontology OWL DL.

4. THE ONTOLOGY

The obtained DIG35 ontology³ (called SemDIG35) is created manually and permits deduction of semantic relationships that are not defined in the provided XML Schema. According to the different parts that are covered by the metadata specifications (see Sect. 2.6) and the semantics they describe, SemDIG35 consists of five subontologies: Basic Image Parameters, Image Creation, Content Description, Image History and IPR. Additionally, a number of fundamental ontologies are created to represent basic concepts: Address, Audio stream, Date & Time, Direction (describes the direction to specify a 3D heading), Email & Web Address, Event, Location, Organization, Person, Phone number, Position (of an object within an image), Product details (hardware or software) and Tangible thing. These ontologies are domain independent and can be reused for different purposes. The legitimacy of the produced OWL ontologies is validated with the WonderWeb OWL Ontology Validator which is recommended by W3C. The metrics of SemDIG are listed in Table 2. Note that the ontology that was constructed holds all information present in the current DIG35 standard. However, many annotations are available that still use the

XML format. Consequently, a mechanism is needed that converts the XML instances into instances of the ontology. This will be discussed in the next section.

³ The ontology is available at <http://multimedialab.elis.ugent.be/users/gmartens/Ontologies/DIG35>

Table 2. SemDIG35 Metrics

Number of classes	152
Number of data type properties	151
Number of object properties	146
Number of restrictions	185
Expressivity	<i>ALUOIN(D)</i>

5. XML TO RDF

Today, many metadata standards are expressed in XML Schema and there are numerous multimedia documents with XML-based annotations. Consequently, a conversion tool is needed to automatically generate the RDF/OWL triples out of the XML fragments. For this purpose, we created a generic XML to RDF convertor [18]. This convertor uses an XML document as a mapping document that defines specific mapping rules between an XML instance document and the resulting RDF document. A generic XMLtoRDF tool takes this mapping document and the used ontology as input and then automatically transforms corresponding XML documents into RDF instances. To foster re-use of our ontology, it should be coupled with other multimedia initiatives in the Semantic Web. This is the topic of the following sections.

6. MEDIA ANNOTATIONS

The W3C Media Annotations Working Group (MAWG) has the goal of improving the interoperability between media metadata schemas. The proposed approach is to provide Interlingua ontology and an API designed to facilitate cross-community data integration of information related to media resources in the web, such as video, audio and images. At first an ontology was created, Media Ontology 1.0 (MA ontology), with the goal of finding common properties in multimedia annotations. Additionally, an API is being defined that allows uniform access to the properties. DIG35 holds many correspondences with the MA ontology since both are targeting multimedia resources. In the next section, we will provide ways to map both ontologies together, allowing an automated conversion of DIG35 instances into instances of the MA ontology.

7. MAPPINGS

A mapping is needed to relate the concepts and properties of the SemDIG35 ontology to those of the MA ontology. Using the mappings, DIG35 instance data can consequently be transformed to instance data of the MA ontology, which can be accessed through the API.

The first step is to identify those concepts that correspond with the MA properties. For this purpose, the ontology documents for the Media Ontology and the textual specifications of DIG3 need to be searched for similar concepts. For DIG35 we created a so called mapping table [3]. This table holds the specific relationships between the concepts of the MA ontology and DIG35. Special care needs to be taken among the actual relationships, since concepts can have different semantics, or syntactic features.

Once these concepts are found, a way to define these relations or mappings is needed. We propose using OWL to express direct mappings between the Media Ontology and SemDIG35 ontology [19]. The mappings expressed in OWL are in fact by themselves an ontology, called a mapping ontology, consisting of basic OWL or RDFS constructs (e.g., `owl:equivalentClass` and `rdfs:subPropertyOf`). The following example defines a formal semantic equivalence between the title property defined in SemDIG35 and in the Media Ontology:

```
dig:descriptiontitle owl:equivalentProperty ma:title
```

These constructs can be used for all properties that have the same semantic and structural characteristics. Note that, rules are needed to create advanced conditional relationships, e.g., to declare instance equivalence when certain properties match, or to do type conversion and transformation of values (e.g., convert bps to kbps). The following example expresses in SWRL [20] that the value of an `ma:frameSize` property can be filled from the values of SemDIG35 properties:

```
[r1: (?res rdf:type dig:BasicParam) ^ (?res dig:imageWidth ?width) ^ (?res dig:imageHeight ?height)
  => (?size1 rdf:type ma:frameSize) ^ (?size1 ma:width ?width) ^ (?size1 ma:height ?height) ^
  (?size1 ma:unit "pixels")]
```

First the rule searches for triples describing the basic parameters of SemDIG35. Then it stores the image width and height in two variables. Next, it creates an instance of the `ma:Size` class and fills in the width and height. Another example of the use of rules is string manipulation through built-in functions:

```
[r2: dig:Person(?p1) ^ dig:imageCreator(?x2,?p1) ^ dig:givenName(?p1,?y1) ^ dig:familyName(?p1,?y2)
  => strcat(?name, ?y1, "", ?y2) ^ ma:creator(?resource,?p1) ^ foaf:name(?p1,?name)]
```

In this example, we first search for an instance of the Person class in SemDIG35; more specifically the person should be the creator of the image. Next, the given and family names are stored in two variables. These variables are used to create the name by concatenation. Additionally, it is stated that the person is the creator of the media resource using the `ma:creator` property. Currently, the creator is represented as an identifier which might correspond with a FOAF Person. As such, by creating a `foaf:name` we can state that the person is a FOAF person and fill in the entire name.

8. EVALUATION

A number of different approaches can be taken for creating the mappings. Cruz et al. use a proprietary mapping table that links an upper ontology to local RDF ontologies [21]. A query upon this upper ontology is therefore first translated, using the mapping table, to different queries on the local RDF ontologies. Moreover, according to their architecture, these queries are translated into XML queries that are executed upon the XML sources. This introduces much overhead compared to the proposed system where only one query is executed on the RDF data

to retrieve the same results.

Garcia et al. used an automatically generated MPEG-7 ontology and presented architecture to achieve semantic multimedia metadata integration and retrieval [22]. They use XSD2OWL to automatically create an ontology based on an existing XML schema. However, this tool is made to allow the automatic conversion from the MPEG-7 XML schema to an OWL ontology (and only results for MPEG-7 conversions were presented). We noticed that this tool is not usable for the conversion of the DIG35 XML schema. For example, when an XSD2OWL translation is used that translates an XML sequence into an intersection of classes (denoted by the *OWL:intersectionOf* construct). Within the DIG35 schema the element *ContentDescription* is defined as a sequence of different elements (Caption, Location, Person, Thing, Comment, etc.). Some of these would be translated into *owl:Class* constructs, others into *owl:DatatypeProperty* constructs, which would invalidate the intersection. Moreover, Garcia et al. do not define ways to map different ontologies to each other. By using the XSD2OWL conversions, making a mapping would only be possible if the different metadata schemes use the same names for the same concepts, which is obviously not always the case.

9. USE CASE

The MA ontology is foreseen as being accessible through a standardized API. Browsers will implement the API so that Web developers can request the metadata of multimedia resources. This allows for the creation of richer search engines, news aggregators and so on. Using the guidelines described above, a photo collection with annotations in DIG35 metadata can be lifted to the MA-space. Web-sites can consequently more easily extract information from the collection and show for instance all pictures created by a specific person. A prototype⁴ has been built including a client-side implementation of the API (in JavaScript). Upon requests by a user, the API interrogates with the MA ontology. Underlying this ontology is a number of different metadata formats (e.g., EXIF, MPEG-7, Dublin Core, and DIG35). The underlying information is represented as XML, and can be queried using the uniform interface.

10. CONCLUSION

In this paper, we have presented guidelines to lift an XML-based metadata standard to the Semantic Web. In our first step, we compared different multimedia metadata standards and have chosen DIG35 as the underlying format. We have discussed different modeling issues that arise when creating ontology based on an existing metadata scheme. Ways have been presented to automatically convert instances of the XML-based standard into instances of the OWL-based ontology. Furthermore, the created ontology is mapped upon the MA ontology, an ongoing initiative of the W3C. We have shown how Semantic Web Technologies can be used to accomplish this mapping.

Future work consists of analyzing how our proposed solution can be used for other metadata standards that are not necessarily focusing on multimedia (e.g., in the cultural heritage or library sectors).

4 <http://multimedialab.elis.ugent.be/users/chpoppe/MediaAnnotations.html>

REFERENCES

- [1] J. R. Smith, P. Schirling, Metadata Standards Roundup, *IEEE Multimedia*, Vol.13, 2006, pp.84-88.
- [2] Digital Imaging Group (DIG), *DIG35 Specification - Metadata for Digital Images - Version 1.1* 2001 April 16, available at <http://www.i3a.org/i dig35.html>
- [3] W. Lee, T. Bürger, F. Sasaki, V. Malaisé, F. Stegmaier, and J. Söderberg. "Ontology for Media Resource" 1.0, *W3C Working Draft, 2009*. <http://www.w3.org/TR/mediaont-10/>
- [4] W. Lee, F. Stegmaier, and C. Poppe. "API for Media Resource 1.0." *W3C Working Draft, 2010*. <http://www.w3.org/TR/mediaont-api-1.0>
- [5] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. Andrea Stein, *OWL Web Ontology Language Reference*, <http://www.w3.org/TR/owl-ref/>
- [6] JEIDA, "Digital still camera image file format standard (exchangeable image file format for digital still cameras: Exif) version 2.1," Technical report, 1998.
- [7] S. Decker et al, The semantic web - on the respective roles of XML and RDF, *IEEE Internet Computing*, 2000.
- [8] M. Kanzaki, "The Kanzaki Exif RDF Schema", <http://www.kanzaki.com/ns/exif>
- [9] The Norm Walsh Exif RDF Schema, <http://sourceforge.net/projects/jpegrdf>
- [10] MPEG-7 Overview, "International Organization for Standardisation," *Klagenfurt ISO/IEC JTC1/SC29/WG11*, July 2002, <http://mpeg.telecomitalia.com/standards/mpeg-7/mpeg-7.htm>
- [11] H. Löffler, W. Baranger, *Photo Metadata White Paper 2007*, IPTC.
- [12] Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set, version 1.1: Reference Description*, Tech. Rep., 2004\
- [13] Y. Lafon, B. Bos, *Describing and Retrieving Photos using RDF and HTTP*, W3C Note, 2002.
- [14] OWL 1.1 Web Ontology Language Structural Specification and Functional-Style Syntax, <http://www.webont.org/owl/1.1/>
- [15] L. S. Bechhofer, R. Volz, "Patching Syntax in OWL ontologies," *International Semantic Web Conference 2004*, pp.668-682.
- [16] L. D. Brickley, L. Miller, "FOAF Vocabulary Specification," 2006, <http://xmlns.com/foaf/0.1/>
- [17] A. J. Miles, N. Rogers, and D. Beckett, "SKOS Core RDF Vocabulary," 2004, <http://www.w3.org/2004/02/skos/core>
- [18] D. Van Deursen, C. Poppe, G. Martens, E. Mannens, R. Van de Walle, "XML to RDF Conversion: a generic approach," in: *Proceedings of the Fourth International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*, 2008.
- [19] Chris Poppe, Gaëtan Martens, Erik Mannens, and Rik Van de Walle. "Personal Content Management System: A Semantic Approach," *Journal of Visual Communication and Image Representation*, Vol.20, No.2, 2009, pp.131-144.
- [20] I. Horrocks, P. Patel-Schneider, H. Boley, S. Tabet, B. Grosz, M. Dean, "SWRL: a semantic web rule language-combining owl and rule ML," *W3C Member Submission*, 21 May, 2004, Available from: <http://www.w3.org/Submission/SWRL>
- [21] I.F. Cruz, H. Xiao, F. Hsu, "An ontology-based framework for XML semantic integration," in: *Proceedings of the International Database Engineering and Applications Symposium*, 2004, pp.217-226.
- [22] R. Garcia, O. Celma, "Semantic integration and retrieval of multimedia metadata," in: *Proceedings of the Fifth Knowledge Markup and Semantic Annotation Workshop*, 2006, pp.69-80.



Gaëtan Martens

Gaëtan Martens received his M.Sc. in Computer Science from Ghent University in 1999 and obtained his PhD degree in computer science engineering in 2011 at the same university. He is a senior researcher at the Multimedia Lab of the Department of Electronics and Information Systems, Ghent University and the Interdisciplinary Institute for Broadband Technology (IBBT). He was a member of the Institute of Psychoacoustics and Electronic Music (IPEM) and of the Department of Applied Mathematics and Computer Science at Ghent University. His current research interests include machine learning, computer vision, musical audio mining, and semantic web applications.



Ruben Verborgh

Ruben Verborgh received his Master's degree in Computer Science Engineering from Ghent University, Belgium in 2010. He is a researcher with the Multimedia Lab — Ghent University, Department of Electronics and Information Systems (ELIS) and the Interdisciplinary institute for Broadband Technology (IBBT). His interests include Semantic Web technologies, multimedia annotation, artificial intelligence, and their relation to multimedia processing. More information about his work is available at <http://multimedialab.elis.ugent.be/organon/arseco>.



Chris Poppe

Chris Poppe received his Master's degree in Industrial Sciences from KaHo Sint-Lieven, Belgium, in 2002 and received his Master's degree in Computer Science from Ghent University, Belgium, in 2004. He joined the Multimedia Lab, Department of Electronics and Information Systems (ELIS), Interdisciplinary Institute for Broadband Technology (IBBT), where he obtained a Ph.D. degree in 2009. His research interests include video coding technologies, video analysis, and multimedia metadata extraction, processing and representation, with a strong focus on standardization processes.



Rik Van de Walle

Rik Van de Walle received his MSc and PhD degrees in engineering from Ghent University, Belgium, in 1994 and 1998, respectively. After a visiting scholarship at the University of Arizona Tucson, he returned to Ghent University, where he became a professor of multimedia systems and applications, and head of the Multimedia Lab. His current research interests include multimedia content delivery, presentation and archiving, coding and description of multimedia data, content adaptation, and interactive mobile multimedia applications.