Why Real-World Multimedia Assets Fail to Enter the Semantic Web

Tobias Bürger*
Digital Enterprise Research Institute (DERI)
Technikerstrasse 21a
6020 Innsbruck, Austria
tobias.buerger@deri.at

Michael Hausenblas
Joanneum Research
Steyrergasse 17
8010 Graz, Austria
michael.hausenblas@joanneum.at

ABSTRACT

Making multimedia assets on the one hand first-class objects on the Semantic Web, while keeping them on the other hand conforming to existing multimedia standards is a non-trivial task. Most proprietary media asset formats are binary, optimized for streaming or storage. However, the semantics carried by the media assets are not accessible directly. In addition, multimedia description standards lack the expressiveness to gain a semantic understanding of the media assets. There exists an array of requirements regarding media assets and the Semantic Web, already. Based on a critical review of these requirements we investigate how ontology languages fit into the picture. We finally analyse the usefulness of formal accounts to describe spatio-temporal aspects of multimedia assets in a practical context.

Categories and Subject Descriptors

H.5.1 [**Information Systems**]: Multimedia Information Systems; I7.4 [**Document and Text Processing**]: Electronic Publishing

General Terms

Multimedia Semantics, Semantic Web

Keywords

Multimedia Semantics, Semantic Web, Multimedia Model, Requirements Analysis

1. INTRODUCTION

Today a huge explosion of content can be experienced on the Web generated by, and for the home users [27]: An increasing number of people produce media assets (as photos, video clips, etc.), and share them on popular sites as Flickr[1], and YouTube[2].

*Tobias Bürger is also affiliated with Salzburg Research
[1]http://www.flickr.org
[2]http://www.youtube.com

More recently, the popular attraction was guided away from image sharing to richer content sharing of videos. This can be seen by the launch of video portals like iFilm.com, Ziddio.com or the dozen of other portals that appeared recently to compete with YouTube[3].

Unsurprisingly there is already a portal called VideoRonk[4] trying to combine other portals by providing a MetaSearch interface, which is quite of an help as one does not want to search on ten or more different sites. However, what is missing is the link between the contents of all these sites, enabling distributed recommendations, cross-linking, etc.

Still, for example a cross-site search on the semantic level is close to impossible. The most obvious reason is due to a lack of metadata coming along with all the content. The power of providing metadata along with content on the Web can be seen at prospering mashups that not just combine APIs—provided by parties as Google[5]— but also trying to mashup things on a semantic level. This can be observed for example at Joost [34]. Having metadata about everything, as video content, blog posts, news feeds and the users of the system makes this new experience of watching TV through the Internet possible. To take this even one step further: Would every stream or video available on the Internet be described more detailed even content on the Internet could be matched with user profiles from applications like Joost and could be offered to watch.

As pointed out in [39, 29], high-quality metadata is essential for multimedia applications. Our recent work within initiatives [40] and research projects[6] has shown, there is a need for going beyond current metadata standards to annotate media assets. Current XML-based standards [21] are diverse, often proprietary and not ad hoc interoperable; cf. also [38]. In SALERO, for example, we are facing the problem to offer a semantic search facility over a diverse set of multimedia assets, e.g., image, videos, 3D objects or character animations. The same is true for the Austrian project GRISINO[7] where we aim to realize a semantic search facility for cultural heritage collections. Automating the handling of metadata for these collections and automating linkage between parts of these collections is hard as the vocabularies to

[3]http://www.youtube.com
[4]http://www.videoronk.com
[5]http://code.google.com/apis/
[6]as, e.g., EU project SALERO, http://www.salero.info
[7]http://www.grisino.at

describe them are mostly diverse and do not offer facilities to attach formal descriptions.

A Motivating Scenario. Imagine a person that wants to watch the recent clips similar to the ones of his favourite experimental artist. Tons of clips are potentially distributed on the Web, which makes searching for them sometimes time consuming and laborious. Thus a central facility to search for and negotiate content is needed. This facility should allow to formulate a search goal, including the characteristics, the subject matter, a maximum price, and the preferred encoding and file format of the clip. In a next step, all portal offerings will be scanned in order to retrieve and negotiate content that matches the users' intention. Note that also parts of a video may match his intention which means that videos need to be fine granular and sufficiently well enough described.

In order for this scenario to work, the descriptions of (1) the goal formulation, (2) the description of the media content by all content owners and (3) the negotiation semantics have to be compatible. Three important focal points of these semantic descriptions are:

- *Expressivity for high level semantic descriptions of content* as typical users are not thinking in terms of colour histograms and spatial / temporal constructs. The characteristics of the media should be described detailed enough.
- The *need for rules*: To effectively identify the part of the content that matches the users' intention, rules are needed to map high level semantic concepts to spatial and temporal segments of the video (eg., because ratings and classifications could only apply to parts of the content, ie., a scene including crime is only suitable for adults)
- *Fine grain semantic descriptions* as of bandwidth, user effort, or cost reason to transfer the whole content is not possible. Thus parts of the content should be described detailed enough.

To reach out, we want to provide answers to the question: Why do we need rich semantic descriptions of media assets on the Web, and (why) is there a need to bundle these descriptions together with the multimedia assets? Simultaneous, we want to provide answers to the questions: How can descriptions be provided? Why are the metadata features of multimedia standards not enough?

2. REQUIREMENTS FOR THE DESCRIPTION OF MULTIMEDIA ASSETS

Requirements for multimedia content descriptions have been researched in a number of papers [15, 39, 29, 5] before and investigations of the combination of multimedia descriptions with features of the Semantic Web are yet numerous [22, 3, 36, 37, 2]. In the following, we give a summarisation of the proposed requirements and add two additional ones (Authoring & Consumption and Performance & Scalability).

Representational Issues. A basic prerequisite is the formal grounding and neutral representation of the format used to describe multimedia assets.

- *Neutral Representation:* The ideal multimedia metadata format has a platform and application independent representation, and is both human and machine processable;
- *Formal Grounding:* Knowledge about media assets must be represented in formal languages, as it must be interpretable by machines to allow for automation.

Extensibility & Reusability. It is requested that the format at hand is extensible, e.g., via an extension mechanism as found in MPEG-7. It should be possible to integrate or reference existing vocabularies [21].

Multimedia Characteristics and Linking. The format should reflect the characteristics of media assets, hence allow linking between data and annotations:

- *Description Structures.* The format should support description structures at various levels of detail, including a rich set of structural, cardinality, and multimedia data-typing constraints;
- *Granularity.* The language has to support the definition of the various spatial, temporal, and conceptual relationships between media assets in a commonly agreed-upon format;
- *Linking.* It has to facilitate a diverse set of linking mechanisms between the annotations and the data being described, including a way to segment temporal media.

Authoring & Consumption. A major drawback of existing metadata approaches is its lacking support for authors in creating annotations along with the lacking benefits of generated annotations.

- *Engineering support.* Appropriate tools are a prerequisite for uptake of new vocabularies. There is the need for at least authoring and consumption environments making use of the vocabularies to demonstrate their usefulness.
- *Deployment.* Multimedia Assets need to be exchangeable, and there must be ways to deploy descriptions along with the assets.

Performance & Scalability. The language should yield descriptions that can be stored, processed, exchanged and queried effectively and efficiently.

3. ENVIRONMENT ANALYSIS: THE SEMANTIC WEB

A good starting point for the analysis of our targeted hosting environment—the Semantic Web—is the Architecture of the World Wide Web [23], in which its three main building blocks are discussed: *identification*, *interaction*, and *data formats*. The Semantic Web, as an extension of the well-known Web roughly has the following characteristics:

- It is a *highly distributed system*. *Identification of resources* is based on URIs—for both data and services;
- There is no single, central "registry", viz. *authorities are decentralised*; data and metadata are under control of a lot of distinct individuals (companies, standardisation bodies, private, etc.)
- Alike in the Web a fundamental building block are *relations* between data, whereas the relations in the Semantic Web are *named*, may be of *any granularity* and allow the automatic interchange of data;
- *Contribuser*[8] inhabit it; each participant may play different roles at once: consuming content and contributing via comments, links, etc.
- Finally, there exists a number of standards. As, RDF allowing formal definitions of the intended meaning, SPARQL for querying, RDF(S), OWL or SKOS to classify content and OWL, WSML, or RIF for describing logical relationships.

Any multimedia metadata format that is after the successful application on the Semantic Web has to be in-line with the above listed characteristics. While some requirements, as formats (e.g. XML) are rather easy to meet, other can pose serious problems regarding the integration into the Semantic Web.

4. MULTIMEDIA ASSETS ON THE SEMANTIC WEB

Firstly, addressing the environmental requirements together with an efficient layering of the semantic descriptions on top of the existing metadata (sub-symbolic level - symbolic level - semantic level) is a necessary prerequisite for multimedia assets to enter the Semantic Web successfully. Secondly, from the requirements gathered in section 2 and the environmental analysis done in section 3 we deduce the following characteristics for multimedia assets on the Semantic Web:

Formality of Descriptions. Formal descriptions are the basic building blocks of the Semantic Web. To enable automatic handling like retrieval, and negotiation of multimedia assets formality of descriptions is a pre-requisite.
Three different (semantic) levels of multimedia metadata can be identified [15]: (1) At the *subsymbolic layer* covering the raw multimedia information typically binary formats are used which are optimized for storage or streaming and which mostly do not provide metadata. (2) The *symbolical layer* provides an additional structural layer for the binary essence stream. For this level standards like MPEG-7, Dublin Core or MPEG-21 can be used. The semantics of the information encoded with these standards are only specified within each standards framework. (3) Therefore the *semantic and logical layer* is needed to provide the semantics for the symbolical layer. This layer should be formally described.

[8]a portmanteau word; contributor and user

Efficient layering and referencing of descriptions. It is necessary to support different levels of meaning attached to multimedia assets, i.e., meaning at the bit-level, *traditional metadata* and semantic (high-level) information. As there are already widely adopted standards available for the description of multimedia assets, the semantic layer must be efficiently put upon those traditional description layers and should not aim to replace it. Furthermore *semantic* descriptions from these traditional layers shall be re-used. As content, parts of content, and traditional and semantic descriptions may be distributed efficient referencing mechanisms for multimedia content must be present.

Interoperability among descriptions. Many formats used in various communities cause interoperability problems when dealing with multimedia content. To overcome this, an RDF based semantic layer should be added on top of these numerous formats to ease their semantic and syntactic integration. However, there are some open problems resp. the integration of existing annotation standards and semantic approaches [39, 29]: The stack of Semantic Web languages and technologies provided by the W3C is well suited to the formal, semantic descriptions of the terms in a multimedia document's annotation. But, as also pointed out in [35], the Semantic Web based languages lack the structural advantages of the XML-based approaches. Additionally, there is a huge amount of work already done on multimedia document annotation within the framework of other standards. This is why a combination of the existing standards is the most promising path for multimedia document description in the near future.

Subjectivity and granularity of descriptions. Opinions and views of content differ among users because of their personal background, culture or previous experiences. As many users are potential contributors to descriptions of assets, opinions may differ. Many of these opinions sometimes do not serve to a unique whole opinion. This is why it should be possible to separately attach these opinions to multimedia assets and keep them separate.

Trust and IPR issues. The Web consists of decentralized authorities and a huge number of contribusers. As descriptions of content—especially in the new changing Web 2.0 environment—are subject to vandalism, there need to be ways to guarantee the validity of the descriptions and to secure descriptions that are just read-only for a user group. Popular portals like Flickr or YouTube show that there is no need to own content in order to annotate it. Furthermore copyright is critical when dealing with multimedia content.

Functional Descriptions. Sometimes the fact that metadata is created to support some specific function is forgotten when summarizing the requirements for a metadata schema. For the metadata creator it should be clear beforehand for what purpose the metadata will be used and what benefits he gains from it [28], ie., using this part of the metadata scheme enhances retrieval, raises social attention or helps you protect your assets.
This in turn also applies to the consumer of the metadata, functional descriptions of what type of information can be inferred from the attached metadata or what type of actions can be performed on the content are essential: this is especially true for information that is obfuscated prior to a possible negotiation phase of the content.

Engineering Support. The presence of metadata is a prerequisite to make multimedia assets accessible, and deployable on the Semantic Web, hence to enable their automated processing. From a developers perspective, there must be tools and standards enabling an integrated authoring, testing, and deployment of multimedia assets along with their associated metadata. In the following the most important areas of engineering support are listed:

- *Edit & Visualise.* To aid the engineer in handling the annotations, editor tools, and IDEs[9] are needed. These may include validator services[10], converter or mapper, and visualisation modules.
- *Libraries & Applications.* When developing applications, the availability of APIs is a core requirement. In special for Semantic Web applications, interface and mapping issues are of importance [17].
- *Deployment* Multimedia containers as HTML, SMIL, etc. require the metadata either being referenced from within the media assets, or being embedded into it. As the data model needs to be RDF—in contrast to existing, flat (tags, etc.) technologies—upcoming approaches as RDFa [1] need to be utilised thoroughly.

5. FORMAL DESCRIPTIONS OF MULTIMEDIA ASSETS

In this part ontology languages which are thought to be used for the advanced requirements which were identified in the sections before. In its core it comprises a a comparison of two families of ontology languages against the requirements postulated in section 4.

The reader is invited to note that not all of the existing languages have the same expressiveness and not all have the same inferential capabilities. Further, the underlying knowledge representation paradigms can differ (eg., Description Logics, Logic Programming, etc.). Corcho and Gomez-Perez [18] present a framework that allows for analysing and comparing the expressiveness and reasoning capabilities of ontology languages, which can be used in the decision process. The process of choosing and selecting the appropriate ontology language includes questions like:

- What expressiveness does an ontology language have?
- What are the inference mechanisms (reasoning capabilities) of it?
- Are there any supporting tools for that language?
- Is the language appropriate for exchanging ontologies between applications?
- Are there translators that transform the ontology implemented in a source into a target language (to enhance reusability, exchangeability or interoperability)?

We are going to take these questions into consideration and simultaneously verify if the languages meet the requirements discussed in section 4.

5.1 Ontology Languages

A number of logical languages have been used for the description of different kinds of knowledge (i.e. ontologies and rules) on the Semantic Web: First Order Logic, Description Logics, Logic Programming and Frame-based Logics. Each of which allow the description of different statements and each imply different complexity results for certain reasoning tasks with these languages.

In this section we want to introduce two of the most promising ontology language families, ie., the OWL- and the WSML-family of languages. The OWL family of languages is a standardisation effort of the W3C and the WSML family of languages is an effort of the WSMO working group, whereas WSML is a formal language for the description of ontologies and Semantic Web Services.

5.1.1 Web Ontology Language (OWL) Family

The Web Ontology Language (OWL) family was designed in a W3C standardisation process because of the need for an ontology language that can be used to formally describe the meaning of terminology used in Web documents, thus, making it easier for machines to automatically process and integrate information available on the Web. This language should be layered on top of XML and RDF (W3C's Resource Description Framework[11]) in order to build on XML's ability to define customized tagging schemes and RDF's approach to representing data.

Currently OWL 1.1[12] is under development; it extends OWL DL in several ways: the underlying DL now is is $\mathcal{SROIQ}$, which provides increased expressive power with respect to properties and cardinality restrictions. Further, OWL 1.1 has user-defined datatypes and restrictions involving datatype predicates, and a weak form of meta-modelling known as punning.

The usage of rules in combination with DL has been investigated for some time [12, 19]—in the Semantic Web stack, it is expected that a rule language will complement the ontology layer.

[9] as for example http://www.topbraidcomposer.com/
[10] http://phoebus.cs.man.ac.uk:9999/OWL/Validator
[11] http://www.w3.org/TR/rdfprimer/
[12] http://owl1_1.cs.manchester.ac.uk/owl_specification.html

5.1.2 *The WSML family of languages*

The activities of the WSMO Working group[13] have yielded proposals of new ontology languages, namely WSML (WSML-Core, WSML-DL, WSML-Flight, WSML-Rule, WSML-Full), OWL- ("OWL minus") [6] and OWL Flight [8]. In [14] unique key features of WSML in comparison of other language proposals are presented. Compared to OWL key features include (1) WSML offers one syntactic framework for a set of layered languages, and (2) it separates between conceptual and logical modelling. The different variants of the WSML framework, all having different expressiveness [26], are:

- **WSML-Core** corresponds with the intersection of DL and Horn Logic (without function symbols and without equality), extended with datatype support in order to be useful in practical applications. WSML-Core is fully compliant with a subset of OWL.
- **WSML-DL** extends WSML-Core to an expressive Description Logic, namely, SHIQ, thereby covering that part of OWL which is efficiently implementable.
- **WSML-Flight** extends WSML-Core in the direction of Logic Programming. WSML-Flight has a rich set of modelling primitives for different aspects of attributes, such as value and integrity constraints. Furthermore, WSML-Flight incorporates a rule language, while still allowing efficient decidable reasoning.
- **WSML-Rule** extends WSML-Flight to a fully-fledged Logic Programming language, by allowing function symbols and unsafe rules.
- **WSML-Full** unifies all WSML variants under a common First-Order umbrella with non-monotonic extensions which allow to capture non-monotonic negation of WSML-Rule.

5.1.3 *The Relation of WSML to OWL*

The relation of WSML to OWL is presented in [7]: WSML-Core is a semantic subset of OWL Lite and WSML-DL is semantically equivalent to OWL DL. A major difference between ontology modelling in WSML and ontology modelling in OWL is that WSML separates between conceptual modelling for the non-expert users and logical modelling for the expert user as it—unlike OWL—uses an epistemology which abstracts from the underlying logical language. WSML-Flight and WSML-Rule are based on the Logic Programming paradigm, rather than the Description Logic paradigm. Thus, their expressiveness is quite different from OWL. On the one hand, WSML-Flight/Rule allow chaining over predicates and non-monotonic negation, but do not allow classical negation and full disjunction and existential quantification. With WSML Logic Programming and the Description Logics paradigm are captured in one coherent framework whereas interaction between the paradigms is achieved through a common subset, WSML-Core.

5.2 Rules

Due to the manifold availability of rule systems, harmonisation efforts have not been successful so far. A relatively new W3C initiative, the Rule Interchange Format Working Group, is now after defining a core rules language for exchanging rules. This Rule Interchange Format Core[14] (RIF Core) language aims at achieving maximum interoperability while preserving rule semantics; from a theoretical perspective, RIF Core corresponds to the language of definite Horn rules. As standardisation is still in its infancy, we will not go further into detail regarding rules, but one has to note that the careful integration of ontology languages is an issue to be addressed; for example the usage of DL concepts in a rule has to be well-defined.

5.3 Comparing Formal Descriptions Regarding the Requirements

In the following a high-level comparison of formal description paradigms for multimedia assets is performed. We chose OWL+RIF on the one side, and WSML/OWL-Flight on the other to achieve a somehow realistic scenario; the result can be found in Table 1[15].

Requirement	OWL 1.1 + RIF	WSML/OWL-Flight
Formal Description	++	++
Layering of Descriptions	+	+
Interoperability	++	+
Granularity	+	+
Trust & IPR issues	-	-
Functional Descriptions	-	+
Engineering Support	++	+
Datatype Support	+	++

Table 1: Comparison of Formal Descriptions for Media Assets.

In the following, we elaborate in detail on each of the items in Table 1, and argue therefore our findings regarding the comparison of OWL 1.1 + RIF vs. WSML/OWL-Flight.

5.3.1 *Formal Description*

Both OWL and WSML provide a framework for the formal (machine-processable) descriptions of ontologies. An ontology in WSML consists of the elements concept, relation, instance, relationInstance and axiom. The primary elements of an OWL ontology concern classes and their instances, properties, and relationships between these instances. The formality of the descriptions is based on logics that allow machines to reason on the information. Whereas OWL is based on Description Logics, the WSML family members are based on different logic languages (ie. Description Logics, Logic Programming or First Order Logic).

Despite the fact, that OWL is more widely adopted and used we believe that WSML with its layered framework is conceptually superior to OWL. A major difference between ontology modelling in WSML and ontology modelling in OWL is that WSML separates conceptual modelling for the non-expert users, and logical modelling for the expert user as it—unlike OWL—uses an epistemology, which abstracts

[13] `http://www.wsmo.org`

[14] `http://www.w3.org/TR/rif-core/`

[15] ++ . . . good support, + . . . available , - . . . not supported

from the underlying logical language making the *surface syntax* nicer. Even if an application later requires OWL, one is able to use WSML tools to convert ontologies that reside in popular logic/language fragments automatically into equivalent OWL ontologies. Furthermore the WSML family framework enables one to choose exactly which language with the needed expressiveness is intend to be used, and later allows an easy switch to another family member because of its common grounding. WSML Rule and WSML Flight also include rule-support. Thus, unlike with OWL, no additional rule language is needed.

5.3.2 *Layering of Descriptions*

An array of existing multimedia metadata formats have been used for years in diverse application areas. However, when one aims at using these formats (as MPEG-7, ID3, etc.) in the context of the Semantic Web, the options are limited. Hence, to enable an efficient layering of RDF-based vocabularies on top of existing multimedia metadata, one may use hybrid techniques.

As a result of our works in the media semantics area, we recently proposed the RDFa-deployed Multimedia Metadata (ramm.x) specification [20]. ramm.x is a light-weight framework allowing existing multimedia metadata to hook into the Semantic Web using RDFa [1].

A different but as well Web compatible approach is described in [25]. There, the authors propose the concept of semantic documents; semantic documents include any information regarding the document and its relationships to other documents. The concept is realised by including XMP descriptions in PDF documents which can be rendered in any browser with available plugins. XMP is a format for embedding metadata in documents using RDF.

5.3.3 *Interoperability*

To adhere to the architecture of the WWW, OWL uses (1) URIs for naming and (2) RDF to provide extensible descriptions. (3) OWL builds on RDF and RDF Schema and adds additional vocabulary for describing properties and classes. (4) The datatype support for OWL is grounded on XML Schema.

WSML has a number of features which allow to integrate it seamlessly in the Web: (1) WSML uses IRIs[16] [13] for the identification of resources. (2) WSML adopts the namespace mechanism of XML, and WSML and XML Schema datatypes are compatible. (3), WSML has an XML- and RDF based syntax for exchange over the Web.
To reach compatiability between WSML and OWL, WSML has a set of defined translators between OWL and WSML [9, 10].

5.3.4 *Granularity*

As stated above, when referring to granularity, we understand the support of the definition of various spatial, temporal, and conceptual relationships regarding annotations. In this sense, OWL and WSML meet the minimal requirements, but do not explicitly address this issue. Depending on the granularity, obviously scalability and performance issues come along. In this respect, again, OWL and WSML can be perceived comparable.

[16]IRIs are the successors of URIs

5.3.5 *Trust and IPR*

In an interdependent, interconnected environment as the Semantic Web, two important aspects immediately arise: data provenance and trust [4]. Requirements regarding trust issues gathered from [30, 16] contain costs and benefits w.r.t. implementation, technology-driven vs. social networking, etc.

Both WSML and OWL do not have explicit provisions for handling trust and IPR issues, respectively.
However, as WSML also is a language capable of describing different aspects of a Web service, it is capable of describing so called 'non-functional properties'. Non-functional properties typically are used to constrain functional—i.e. the formal specification of what a service can do—and behavioural aspects, namely how the functionality can be achieved in terms of interaction of a Web services; they may also be utilised to specify trust and IPR properties [31].

5.3.6 *Functional Descriptions*

WSML is a language for the specification of ontologies and different aspects of Web services. As such it not only provides means for modeling and description of ontologies but also functional (service) descriptions, i.e. the description of a service capability by means of precondition, assumptions, postconditions and effects [24].
OWL does not have support for such kind of descriptions.

5.3.7 *Engineering Support*

Tool Support for WSML and especially OWL is constantly growing. However, the amount of tools available for OWL [41] and WSML [11] can drastically not be compared. As OWL is a W3C Recommendation, the support for it is huge.

5.3.8 *Data Type Support*

The reader is invited to note that both OWL and WSML ground their datatype support on XML Schema. In WSML, XML Schema primitive datatypes, simple types and XML Schema derived datatypes are supported [33]; OWL adopts the RDF(S) specification of datatypes [32], though some XML Schema built-ins are problematic.

6. CONCLUSIONS

The first question we kept open is "What are real-world multimedia assets"? Real-world multimedia assets are multimedia objects which can be currently found embedded in HTML pages on the Web, as images, videos, etc. We see three main reasons why media assets fail to enter the Semantic Web:

1. There is a lack of the critical mass of annotated content which is mainly due to the large scale automation of (semantic) visual analysis has not gone that far. This is why the user is the central person in the process in order to provide manual annotations. Motivating user to attach complex annotations to content is not easy to achieve.

2. Current traditional and Web 2.0 based approaches to multimedia annotation are not useful to achieve the goals of the Semantic Web: The most important aspects that the Semantic Web intends to solve are (i) Annotation, (ie., how to associate metadata to a resource), (ii) Information Integration (ie., how to integrate information about resources), and (iii) Inference (ie., reasoning over known facts to unleash hidden facts).
   Existing multimedia metadata standards as MPEG-7 can be used to annotate but keep a certain amount of ambiguity amongst these annotations. As it is a standard it allows easy integration based on it (a requirement for that is that everyone adheres to this standard!) but inference is not possible with the information attachable to a MPEG-7 file. The problem with tagging is manifold: There are big open problems amongst tagging: (a) How can you guarantee consistency among tags of different users? (b) How do you reconcile tags? (c) How do you associate tags with parts of the tagged content? This huge amount of uncertainty will not allow reliable information integration, nor allow to reason on it.

3. As we argued in this paper, more requirements have to be fulfilled, which can not be solely solved by traditional or Web 2.0 based approaches and which make more formalized descriptions of content necessary. However, before not being able to attach these directly to the media being described, multimedia assets will not be able to enter the Semantic Web.

Acknowledgements

The research leading to this paper was partially supported by the European Commission under contract FP6-027026, "Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space" and SALERO (contract number FP6-027122).

7. REFERENCES

[1] B. Adida and M. Birbek. RDFa Primer 1.0 - Embedding RDF in XHTML. W3C Working Draft, W3C RDF in XHTML Taskforce, 2007.

[2] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *Proceedings of the 6th International Semantic Web Conference (ISWC'2007), Busan, Korea, November 11-15, 2007, (forthcoming)*, 2007.

[3] T. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis, and Y. Kompatsiaris. Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content. Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05), Galway, Ireland, November 2005, 2005.

[4] C. Bizer and R. Oldakowski. Using context- and content-based trust policies on the Semantic Web. In *Proceedings of the 132th international World Wide Web conference on Alternate track papers & posters*, pages 228–229. ACM Press, 2004.

[5] T. Bürger and R. Westenthaler. Mind the gap - requirements for the combination of content and knowledge. In *Poster Proceedings of the SAMT 2006 Conference, Athens, Greece*, 2006.

[6] J. de Bruijn and A. P. (eds.). OWL^{-}. WSML Deliverable D20.1v0.2 WSML Working Draft 05-15-2005, http://www.wsmo.org/TR/d20/d20.1/v0.2/, 2005.

[7] J. de Bruijn, H. Lausen, A. Polleres, and D. Fensel. The Web Service Modeling Language WSML: An Overview. In *ESWC*, pages 590–604, 2006.

[8] J. de Bruijn (ed.). OWL Flight. D20.3v0.1 OWL Flight WSML Working Draft 23-08-2004, http://www.wsmo.org/2004/d20/d20.3/v0.1/, 2004.

[9] DERI. OWL - WSML Translator v1.0. http://tools.deri.org/wsml/owl2wsml-translator/v0.1/, 2007.

[10] DERI. WSML - OWL Translator v1.0. http://tools.deri.org/wsml/wsml2owl-translator/v0.1/, 2007.

[11] DERI. WSML Tools. http://tools.deri.org/wsml/, 2007.

[12] F. M. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. AL-log: Integrating Datalog and Description Logics. *Journal of Intelligent Information Systems*, 10(3):227–252, 1998.

[13] M. Duerst and M. Suignard. Internationalized Resource Identifiers (IRIs). IETF RFC 3987, 2005. http://www.ietf.org/rfc/rfc3987.txt.

[14] D. Fensel, H. Lausen, A. Polleres, J. de Bruijn, M. Stollberg, D. Roman, and J. Domingue. *Enabling Semantic Web Services: The Web Service Modeling Ontology*. Springer, 11 2006.

[15] J. Geurts, J. van Ossenbruggen, and L. Hardman. Requirements for practical multimedia annotation. In *Proceedings of the Workshop on Multimedia and the Semantic Web, May 2005, Heraklion, Crete*, pages 4–1, 2005.

[16] J. Golbeck, B. Parsia, and J. Hendler. Trust Networks on the Semantic Web. In *Proceedings of Cooperative Intelligent Agents 2003*, 2003.

[17] N. M. Goldman. Ontology-Oriented Programming: Static Typing for the Inconsistent Programmer. In *Proceedings of the Second International Semantic Web Conference - ISWC 2003*, pages 850–865, 2003.

[18] Gomez-Perez, Fernandez-Lopez, and Corcho-Garcia. *Ontological Engineering*. Springer, Berlin, 2004.

[19] B. Grosof, I. Horrocks, R. Volz, and S. Decker. Description Logic Programs: Combining Logic Programs with Description Logics. In *12th International World Wide Web Conference (WWW'03)*, Budapest, Hungary, 2003.

[20] M. Hausenblas, W. Bailer, and T. Bürger. Deploying Multimedia Metadata on the Semantic Web - RDFa-deployed Multimedia Metadata (ramm.x). Specification, ramm.x Working Group, 2007.

[21] M. Hausenblas, S. Boll, T. Bürger, O. Celma, C. Halaschek-Wiener, E. Mannens, and R. Troncy. Multimedia Vocabularies on the Semantic Web. W3C Incubator Group Report, W3C Multimedia Semantics Incubator Group, 2007.

[22] J. Hunter. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *First International Semantic Web Working Symposium (SWWS'01)*, Stanford, California, USA, 2001.

[23] I. Jacobs and N. Walsh. Architecture of the World Wide Web, Volume One. http://www.w3.org/TR/webarch/, 2004.
[24] U. Keller, H. Lausen, and M. Stollberg. On the Semantics of Functional Descriptions of Web Services. In *The Semantic Web: Research and Applications (Proceedings of ESWC 2006)*, pages 605–619, 2006.
[25] H. Kim, H. Kim, J. H. Choi, and S. Decker. Translating Documents into Semantic Documents using Semantic Web and Web 2.0. In *Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW2006)*, 2006.
[26] H. Lausen, J. de Bruijn, A. Polleres, and D. Fensel. WSML - a Language Framework for Semantic Web Services. In *Proceedings of the W3C Workshop on Rule Languages for Interoperability*, 2005.
[27] J. Markoff. Web content by and for the masses. New York Times Online, June 2005.
[28] A. Morgan and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM Press.
[29] F. Nack, J. van Ossenbruggen, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). *IEEE Multimedia*, 12(1), 2005.
[30] K. O'Hara, H. Alani, Y. Kalfoglou, and N. Shadbolt. Trust Strategies for the Semantic Web. In *ISWC Workshop on Trust, Security, and Reputation on the Semantic Web*, 2004.
[31] J. O'Sullivan, D. Edmond, and A. H. ter Hofstede. Formal description of non-functional service properties. technical report. Technical report, Queensland University of Technology, Brisbane., 2005.
[32] J. Z. Pan. *Description Logics: Reasoning Support for the Semantic Web*. PhD thesis, School of Computer Science, The University of Manchester, 2004.
[33] D. Roman, H. Lausen, and U. Keller. Web Service Modeling Ontology (WSMO), WSMO Deliverable D2v1.0., WSMO Working Draft 20 September 2004, September 2004.
[34] L. Simons. RDF at the Venice Project. http://www.leosimons.com/2006/rdf-at-the-venice-project.html, 2006. Blog Post.
[35] G. Stamou, J. van Ossenbruggen, J. Z. Pan, and G. Schreiber. Multimedia annotations on the semantic web. *IEEE MultiMedia*, 13(1):86–90, 2006.
[36] R. Troncy. Integrating Structure and Semantics into Audio-visual Documents. In *Proceedings of the 2nd International Semantic Web Conference (ISWC'03)*, volume LNCS 2870, pages 566–581, 2003.
[37] C. Tsinaraki, P. Polydoros, and S. Christodoulakis. Integration of OWL ontologies in MPEG-7 and TVAnytime compliant Semantic Indexing. In *Proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAiSE)*, 2004.
[38] V. Tzouvaras (ed.). Multimedia Annotation Interoperability Framework; MMSEM XG Report. http://www.w3.org/2005/Incubator/mmsem/wiki/Semantic_Interoperability, 2007.
[39] J. van Ossenbruggen, F. Nack, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). *IEEE Multimedia*, 11(4), 2004.
[40] W3C. Multimedia Semantics Incubator Group. http://www.w3.org/2005/Incubator/mmsem/, 2007.
[41] W3C. Semantic Web Development Tools. http://esw.w3.org/topic/SemanticWebTools, 2007.