Multimedia Vocabularies on the Semantic Web

W3C Incubator Group Report 24 July 2007

Editor:
> Michael Hausenblas, JOANNEUM RESEARCH

Contributors:
> Susanne Boll, University of Oldenburg
> Tobias Bürger, Digital Enterprise Research Institute (DERI)
> Oscar Celma, Music Technology Group, Pompeu Fabra University
> Christian Halaschek-Wiener, University of Maryland
> Erik Mannens, IBBT-MMLab, University of Ghent
> Raphaël Troncy, Center for Mathematics and Computer Science (CWI Amsterdam)

> Also see Acknowledgements.

Abstract

This document gives an overview on the state-of-the-art of multimedia metadata formats. Initially, practical relevant vocabularies for developers of Semantic Web applications are listed according to their modality scope. In the second part of this document, the focus is set on the integration of the multimedia vocabularies into the Semantic Web, that is to say, formal representations of the vocabularies are discussed.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of Final Incubator Group Reports is available. See also the W3C technical reports index at http://www.w3.org/TR/.

This document was developed by the W3C Multimedia Semantics Incubator Group, part of the W3C Incubator Activity.

Publication of this document by W3C as part of the W3C Incubator Activity indicates no endorsement of its content by W3C, nor that W3C has, is, or will be allocating any resources to the issues addressed by it. Participation in Incubator Groups and publication of Incubator Group Reports at the W3C site are benefits of W3C Membership.

Incubator Groups have as a goal to produce work that can be implemented on a Royalty Free basis, as defined in the W3C Patent Policy. Participants in this Incubator Group have made no statements about whether they will offer licenses according to the licensing requirements of the W3C Patent Policy for portions of this Incubator Group Report that are subsequently incorporated in a W3C Recommendation.

Scope

This document targets Semantic Web developers that deal with multimedia. No prerequisites are assumed. The target audience may range from prosumers to professionals working with audio-visual archives, libraries, media productions, and broadcast industry.

After reading this document, readers may also be interested in related issues as presented in the tools and resources document.

Note: A living version of this document is maintained at the Multimedia Semantics Incubator Group Wiki page: http://www.w3.org/2005/Incubator/mmsem/wiki/Vocabularies .

Objectives

This document aims at:

- Giving an overview on the state-of-the-art of multimedia metadata formats and vocabularies, and
- Summarizing formalizations of multimedia metadata formats to be used on the Semantic Web.

Discussion of this document is invited on the public mailing list public-xg-mmsem@w3.org (public archives). Public comments should include "[MMSEM-Vocabulary]" as subject prefix .

Table of Contents

1. Introduction

This document gives an overview on the state-of-the-art of multimedia metadata formats. A special focus is set on the usability with respect to the Semantic Web, that is to say, formal representations of exiting vocabularies.

1.1 Declaration of Namespaces

The syntax for all RDF code snippets in this document is N3, the namespace used herein are listed in Table 1-1. Note that the choice of any namespace prefix is arbitrary, hence not significant semantically [*XML NS*].

Table 1-1. XML namespaces used in this document.

Prefix	URI
xsd	<"http://www.w3.org/2001/XMLSchema#">
rdf	<"http://www.w3.org/1999/02/22-rdf-syntax-ns#">
rdfs	<"http://www.w3.org/2000/01/rdf-schema#">
owl	<"http://www.w3.org/2002/07/owl#">
dc	<"http://purl.org/dc/elements/1.1/">

1.2 Related Pages

Complementary and related resources can be obtained at the following pages:

- The Tools and Resources Wiki page of the MMSEM-XG.
- The MPEG-7 and the Semantic Web document of the MMSEM-XG.

2. Types of Multimedia Metadata

Based on [Smith et al., 2006], the vocabularies in this document are described in terms of the following two tables. Table 2-1 lists the used discriminators, and Table 2-2 the categories. For the example column in both tables the *NewsML-G2* vocabulary is used.

Note that a discriminator is understood in terms of its possible values, that is, a list of comma-separated values. The possible values given in Table 2-1 should be understood as being exhaustive, thus the range for a value is defined by the content of the corresponding column. A category on the other side is a list of comma-separated items; the possible items are non-exhaustive. The range for an item of a category in Table 2-2 is

open; examples are listed in the content of the corresponding column.

Description of used discriminators:

- Representation: the primary (official) serialization format for the multimedia standard.
- Content Type: the type of media, a certain multimedia standard is capable to describe.

Table 2-1. Discriminators for multimedia metadata standards used in this document.

Discriminator	Permitted Values	Example
Representation	non-XML (nX), XML (X), RDF (R), OWL (O)	X, R
Content Type	still-image (SI), video (V), audio (A), text (T), general purpose (G)	G, T, SI, V

Description of used categories:

- Workflow: understood in terms of the Canonical Processes of Media Production, see [Hardman, 2005].
- Domain: the main domain in which a multimedia vocabulary is intended to be used in.
- Industry: the main branch of productive (commercial) usage.

Table 2-2. Categories for multimedia metadata standards
used in this document.

Category	Items	Example
Workflow	premeditation, production, publish, etc.	publish
Domain	entertainment, news, sports, etc.	news
Industry	broadcast, music, publishing, etc.	broadcast

3. Existing Multimedia Metadata Formats

This section introduces common existing metadata formats that are of importance for the description and usage of multimedia content. Each vocabulary starts with a table containing the responsible party, the specification (if available) and a list of discriminators and categories. The description of each vocabulary should enable the reader to get an idea of its capabilities, and its limitations.

3.1 Multimedia Metadata Formats For Describing Still Images

In the following, metadata formats are listed that deal with the description of still image content.

3.1.1. Visual Resource Association (VRA)

Responsible	Specification	Formal Representation
http://www.vraweb.org/	[VRA Core]	VRA - RDF/OWL

Representation	Content Type	Workflow	Domain	Industry
nX	SI	publish	culture	archives

The Visual Resource Association (VRA) is an organization consisting of over 600 active members, including many American Universities, galleries and art institutes. These often maintain large collections of (annotated) slides, images and other representations of works of art. The VRA has defined the VRA Core Categories to describe such collections. The VRA Core [VRA Core] is a set of metadata elements used to describe works of visual culture as well as the images that represent them.

When the Dublin Core [Dublin Core] specifies a small and commonly used vocabulary for on-line resources in general, VRA Core defines a similar set targeted especially at visual resources. Dublin Core and VRA Core both refer to terms in their vocabularies as elements, and both use qualifiers to refine elements in similar way. The more general elements of VRA Core have direct mappings to comparable fields in Dublin Core. Furthermore, both vocabularies are defined in a way that abstracts from implementation issues and underlying serialization languages.

3.1.2 Exchangeable image file format (Exif)

Responsible	Specification	Formal Representation
http://www.jeita.or.jp/english/	[Exif]	Exif - RDF/OWL

Representation	Content Type	Workflow	Domain	Industry
nX	SI	capture-distribute	generic	digital camera

One of nowaday's commonly used metadata format for digital images is the Exchangeable Image File Format (Exif) [Exif]. The standard "specifies the formats to be used for images and sounds, and tags in digital still cameras and for other systems handling the image and sound files recorded by digital cameras." The so called Exif header carries the metadata for the captured image or sound.

The metadata tags which the Exif standard provides covers metadata related to the capture of the image and the context situation of the capturing. This includes metadata related to the image data structure (e.g., height, width, orientation), capturing information (e.g., rotation, exposure time, flash), recording offset (e.g., image data location, bytes per compressed strip), image data characteristics (e.g., transfer function, color space transformation), as well as general tags (e.g., image title, copyright holder, manufacturer). In these days new camera also write GPS information into the header. Lastly, we point out that metadata elements pertaining to the image are stored in the image file header and are marked identified by unique tags, which serve as an element identifier.

3.1.3 NISO Z39.87

Responsible	Specification
http://www.niso.org/	[NISO Z39.87]

Representation	Content Type	Workflow	Domain	Industry
X	SI	production	interoperability	image creation

The NISO Z39.87 standard [NISO Z39.87] defines a set of metadata elements for raster digital images to enable users to develop, exchange, and interpret digital image files.

Tags cover a wide spectrum of metadata: basic image parameters, image creation, imaging performance assessment, history. This standard is intended to facilitate the development of applications to validate, manage, migrate, and otherwise process images of enduring value. Such applications are viewed to be essential components of large-scale digital repositories and digital asset management systems.

The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management and continuing access to digital image collections.

3.1.4 DIG35

Responsible	Specification	Formal Representation
http://www.i3a.org/	[DIG35]	DIG35 - RDF/OWL

Representation	Content Type	Workflow	Domain	Industry
X	SI	publish	archives	consumer

The DIG35 specification [DIG35] includes a "standard set of metadata for digital images" which promotes interoperability and extensibility, as well as a "uniform underlying construct to support interoperability of metadata between various digital imaging devices."

The metadata properties are encoded within an XML Schema and cover:

- Basic Image Parameter (a general-purpose metadata standard);
- Image Creation (e.g. the camera and lens information);
- Content Description (who, what, when and where);
- History (partial information about how the image got to the present state);
- Intellectual Property Rights;
- Fundamental Metadata Types and Fields (define the format of the field defined in all metadata block).

Note: DIG35 Metadata Specification Version 1.1 is not free ($35).

3.1.5 PhotoRDF

Responsible	Specification
http://www.w3.org/	[PhotoRDF]

Representation	Content Type	Workflow	Domain	Industry

R	SI	capture-distribute	personal media	photo

PhotoRDF [PhotoRDF] is an attempt to standardize a set of categories and labels for personal photo collections. The standard has been proposed in early 2002 but did not develop since. The latest version is a W3C Note from 19 April 2002. The standard already works as a roof for different other standards that together should solve the "project for describing & retrieving (digitized) photos with (RDF) metadata". The metadata is separated into three different schemas, a Dublin Core, a technical schema and a content schema. As the standard aims to be short and simple it covers only a small set of properties. The Dublin Core schema is adopted for those parts of a photo that needs description for its creator, editor, title, date of publishing and so on. With regard to the technical aspects of a photo, however, the standard includes less properties than EXIF. For the actual description of the content, the content schema defines a very small set of keywords that shall be used in the "subject" field of the Dublin Core schema.

PhotoRDF addressed the demand for a small standard describing personal photos for personal media management as well as for publishing and exchanging photos between different tools. It covers the different aspects of a photo that range from the camera setting to the subject depicted on the photo. The standard fails, however, to cover the central aspects of photos as they are needed for interoperability of photo tools and photo services. For example, the place or position of a photo is not addressed as well as photographic information such as aperture. Also the content description property is limited by a small number of keywords. The trend for tagging has not been foreseen at the time of the development of the standard.

3.2 Multimedia Metadata Formats For Describing Audio Content

This section contains metadata for audio content, be it related to music, or speech.

3.2.1 ID3

Responsible	Specification
http://www.id3.org/	[ID3]

Representation	Content Type	Workflow	Domain	Industry
nX	A	distribute	generic	music

ID3 [ID3] is a metadata container used and embedded with an MP3 audio file format. It allows to state information about the title, artist, album, etc. about a song. The ID3 specification aims to address a broad spectrum of metadata (represented in so called 'frames') ranging from encryption, over involved people list, lyrics, band, relative volume adjustment to overownership, artist, and recording dates. Additionally user can define own properties. A list of 79 genres is defined (from Blues to Hard Rock).

3.2.2 MusicBrainz Metadata Initiative 2.1

Responsible	Specification

http://musicbrainz.org/ [MusicBrainz]

Representation	Content Type	Workflow	Domain	Industry
R	A	production	generic	music

MusicBrainz defines a RDF-S based vocabulary, including three namespaces [MusicBrainz]. The core set is capable of expressing basic music related metadata such as artist, album, track, etc.). Instances in RDF are being made available via a query language. The third namespace is reserved for future use in expressing extended music related metadata such as contributors, roles, lyrics, etc.

3.2.3 MusicXML

Responsible	Specification
http://www.recordare.com/	[MusicXML]

Representation	Content Type	Workflow	Domain	Industry
X	A	production	generic	music

Recordare has developed the MusicXML technology [MusicXML] to create an Internet-friendly method for publishing musical scores, enabling musicians and music fans to get more out of their online music.

MusicXML is a universal translator for common Western musical notation from the 17th century onwards. It is designed as an interchange format for notation, analysis, and retrieval for music notation nd digital sheet music applications. The MusicXML format is open for use by anyone under a royalty-free license, and is supported by over 75 applications.

3.3 Multimedia Metadata Formats For Describing Audio-Visual Content

In this section, multimedia metadata formats for describing audio-visual content in general are described.

3.3.1 Multimedia Content Description Interface (MPEG-7)

Responsible	Specification	Formal Representation
http://www.iso.org/iso/en/prods-services/popstds/mpeg.html	[MPEG-7]	MPEG-7 - RDF/OWL

Representation	Content Type	Workflow	Domain	Industry
X, nX	SI, V, A	archive-publish	generic	generic

The MPEG-7 standard [MPEG-7], formally named "Multimedia Content Description" aims

to be an overall for describing any multimedia content. MPEG-7 standardizes so-called "description tools" for multimedia content: Descriptors (Ds), Description Schemes (DSs) and the relationships between them. Descriptors are used to represent specific features of the content, generally low-level features such as visual (e.g. texture, camera motion) or audio (e.g. melody), while description schemes refer to more abstract description entities (usually a set of related descriptors). These description tools as well as their relationships are represented using the Description Definition Language (DDL), a core part of the language. The W3C XML Schema recommendation has been adopted as the most appropriate schema for the MPEG-7 DDL, adding a few extensions (array and matrix datatypes) in order to satisfy specific MPEG-7 requirements. MPEG-7 descriptions can be serialized as XML or in a binary format defined in the standard.

MPEG-7's comprehensiveness results from the fact that the standard has been designed for a broad range of applications and thus employs very general and widely applicable concepts. The standard contains a large set of tools for diverse types of annotations on different semantic levels (the set of MPEG-7 XML Schemas define 1182 elements, 417 attributes and 377 complex types). The flexibility is very much based on the structuring tools and allows the description to be modular and on different levels of abstraction. MPEG-7 supports fine grained description, and it provides the possibility to attach descriptors to arbitrary segments on any level of detail of the description. The possibility to extend MPEG-7 according to the conformance guidelines defined in part 7 provides further flexibility. Two main problems arise in the practical use of MPEG 7 from its flexibility and comprehensiveness: complexity and limited interoperability. The complexity is a result of the use of generic concepts, which allow deep hierarchical structures, the high number of different descriptors and description schemes, and their flexible inner structure, i.e. the variability concerning types of descriptors and their cardinalities. This causes sometimes hesitance in using the standard. The interoperability problem is a result of the ambiguities that exist because of the flexible definition of many elements in the standard (e.g. the generic structuring tools). There can be several options to structure and organize descriptions which are similar or even identical in terms of content, and they result in conformant, yet incompatible descriptions. The description tools are defined using DDL. Their semantics is descibed textually in the standard documents.

Due to the wide application, the semantics of the description tools are often very general. Several works have already pointed out the lack of formal semantics of the standard that could extend the traditional text descriptions into machine understandable ones. These attempts that aim to bridge the gap between the multimedia community and the Semantic Web, either for the whole standard, or just one of its part, are detailed below.

MPEG-7 Profiles and Levels

Profiles and levels have been proposed as a means to reduce the complexity of MPEG-7 descriptions [MPEG-7 Profiles]. Like in other MPEG standards, profiles are subsets of the standard that cover certain functionalities, while levels are flavours of profiles with different complexity. In MPEG-7, profiles are subsets of description tools for certain application areas, levels have not yet been used. The proposed process of the definition of a profile consists of three steps:

- Selection of tools supported in the profile, i.e. the subset of descriptors and

description schemes that are used in description that conform to the profile.
- Definition of constraints on these tools, such as restrictions on the cardinality of elements and on the use of attributes.
- Definition of constraints on the semantics of the tools, which describe their use in the profile more precisely.

The result of tool selection and the definition of tool constraints are formalized using the MPEG-7 DDL and result in an XML schema like the full standard. Several profiles have been under consideration for standardization and three profiles have been standardized (they constitute part 9 of the standard, with their XML schemas being defined in part 11):

- *Simple Metadata Profile (SMP).* Allows describing single instances of multimedia content or simple collections. The profile contains tools for global metadata in textual form only. The proposed Simple Bibliographic Profile is a subset of SMP. Mappings from ID3, 3GPP and EXIF to SMP have been defined.
- *User Description Profile (UDP).* Its functionality consists of tools for describing user preferences and usage history for the personalization of multimedia content delivery.
- *Core Description Profile (CDP).* Allows describing image, audio, video and audiovisual content as well as collections of multimedia content. Tools for the description of relationships between content, media information, creation information, usage information and semantic information are included. The CDP does not include the visual and audio description tools defined in parts 3 and 4.

The adopted profiles will not be sufficient for a number of applications. If an application requires additional description tools, a new profile must be specified. It will thus be necessary to define further profiles for specific application areas. For interoperability it is crucial, that the definitions of these profiles are published, to check conformance to a certain profile and define mappings between the profiles. It has to be noted, that all of the adopted profiles just define the subset of description tools to be included and some tool constraints; none of the profile definitions includes constraints on the semantics of the tools that clarify how they are to be used in the profile.

Apart from the standardized ones, a profile for the detailed description of single audiovisual content entities called Detailed Audiovisual Profile (DAVP) [DAVP] has been proposed. The profile includes many of the MDS tools, such as a wide range of structuring tools, as well as tools for the description of media, creation and production information and textual and semantic annotation, and for summarization. In contrast to the adopted profiles, DAVP includes the tools for audio and visual feature description, which was one motivation for the definition of the profile. The other motivation was to define a profile the supports interoperability between systems using MPEG-7 by avoiding possible ambiguities and clarifying the use of the description tools in the profile. The DAVP definition thus includes a set of semantic constraints, which play a crucial role in the profile definition. Due to the lack of formal semantics in DDL, these constraints are only described in textual form in the profile definition.

Controlled vocabularies in MPEG-7

Annotation of content often contains references to semantic entities such as objects, events, states, places, and times. In order to ensure consistent descriptions (e.g. make

sure that persons are always referenced with the same name) some kind of controlled vocabulary should be used in these cases. MPEG-7 provides a generic mechanism for referencing terms defined in controlled vocabularies. The only requirement is that the controlled vocabulary is identified by a URI, so that a specific term in a specific controlled vocabulary can be referenced unambiguously. In the simplest case, the controlled vocabulary is just a list of possible values of a property in the content description, without any structure. The list of values can be defined in a file accessed by the application or can be taken from some external source, for example the list of countries defined in ISO 3166. The mechanism can also be used to reference terms from other external vocabularies, such as thesauri or ontologies.

Classification schemes (CSs) are a MPEG-7 description tool that allows to describe a set of terms using MPEG-7 description schemes and descriptors. It allows to define hierarchies of terms and simple relations between them, and allows the term names and definitions to be multilingual. Part 5 of the MPEG-7 standard already defines a number of classification schemes, and new ones can be added. The CSs defined in the standard are for those description tools, which require or encourage the use of controlled vocabularies, such as

- Technical media information: encoding, physical media types, file formats, defects;
- Content classification: genre, format, rating;
- Other: affection, role of creator, dissemination format

Note: Further descriptions of MPEG-7 will be available in the XGR MPEG-7 and the Semantic Web.

3.3.2 Advanced Authoring Format (AAF)

Responsible	Specification
http://www.aafassociation.org/	[AAF]

Representation	Content Type	Workflow	Domain	Industry
nX	SI, V, A	production	content creation	broadcast

The Advanced Authoring Format (AAF) [AAF] is a cross-platform file format that allows the interchange of data between multimedia authoring tools. AAF supports the encapsulation of both metadata and essence, but its primary purpose involves the description of authoring information. The object-oriented AAF object model allows for extensive timeline-based modeling of compositions (i.e. motion picture montages), including transitions between clips and the application of effects (e.g. dissolves, wipes, flipping). Hence, the application domain of AAF is within the post production phase of an audiovisual product and it can be employed in specialized video work centers. Among the structural metadata contained for clips and compositions, AAF also supports storing event-related information (e.g. time-based user annotations and remarks) or specific authoring instructions.

AAF files are fully agnostic as to how essence is coded and serve as a wrapper for any kind of essence coding specification. In addition to describe the current location and characteristics of essence clips, AAF also supports descriptions of the entire derivation

chain for a piece of essence, from its current state to the original storage medium, possibly a tape (identified by tape number and time code), or a film (identified by an edge code for example).

The AAF data model and essence are independent of the specificities of how AAF files are stored on disk. The most common storage specification used for AAF files is the Microsoft Structured Storage format, but other storage formats (e.g. XML) can be used.

The AAF metadata specifications and object model are fully extensible (e.g. subclassing existing objects) and the extensions are fully contained in a metadata dictionary, stored in the AAF file. In order in order to achieve predictable interoperability between implementations created by different developers, due to the format's flexibility and use of proprietary extensions, the Edit Protocol was established. The Edit Protocol combines a number of best practices and constraints as to how an Edit Protocol-compatible AAF implementation must function and which subset of the AAF specification can be used in Edit Protocol-compliant AAF files.

3.3.3 Material Exchange Format (MXF)

Responsible	Specification
http://www.smpte.org/	[MXF]

Representation	Content Type	Workflow	Domain	Industry
nX	SI, V, A	production	content creation	broadcast

The Material Exchange Format (MXF) [MXF] is a streamable file format optimized for the interchange of material for the content creation industries. MXF is a wrapper/container format intended to encapsulate and accurately describe one or more 'clips' of audiovisual essence (video, sound, pictures, etc.). This file format is essence-agnostic, which means it should be independent of the underlying audio and video coding specifications in the file. In order to process such a file, its header contains data about the essence. An MXF file contains enough structural header information to allow applications to interchange essence without any a priori information. The MXF metadata allows applications to know the duration of the file, what essence codecs are required, what timeline complexity is involved and other key points to allow interchange.

There exists a 'Zero Divergence' doctrine, which states that any areas in which AAF and MXF overlap must be technologically identical. As such, MXF and AAF share a common data model. This means that they use the same model to represent timelines, clips, descriptions of essence, and metadata. The major difference between the two is that MXF has chosen not to include transition and layering functionality. This makes MXF the favorable file format in embedded systems, such as VRTs or cameras, where resources can be scare. Essentially, this creates an environment in which raw essence can be created in MXF, it can be post produced in AAF, and then the finished content can be generated as an MXF file.

MXF uses KLV coding throughout the file structure. This KLV is a data interchange format defined by the simple data construct: *Key-Length-Value*, where the *Key* identifies the data meaning, the *Length* gives the data length, and the *Value* is the data itself. This principle

allows a decoder to identify each component by its key and skip any component it cannot recognize using the length value to continue decoding data types with recognized key values. KLV coding allows any kind of information to be coded. It is essentially a machine-friendly coding construct that is datacentric and is not dependent on human language. Additionally, the KLV structure of MXF allows this file format to be streamable.

Structural Metadata is the way in which MXF describes different essence types and their relationship along a timeline. The structural metadata defines the synchronization of different tracks along a timeline. It also defines picture size, picture rate, aspect ratio, audio sampling, and other essence description parameters. The MXF structural metadata is derived from the AAF data model. Next to the structural metadata described above, MXF files may contain descriptive and *dark* metadata.

MXF descriptive metadata comprises information in addition to the structure of the MXF file. Descriptive metadata is metadata created during production or planning of production. Possible information can be about the production, the clip (e.g. which type of camera was used) or a scene (e.g. the actors in it). **DMS-1** (Descriptive Metadata Scheme 1) [MXF-DMS-1] is an attempt to standardize such information within the MXF format. Furthermore DMS-1 is able to interwork as far as practical with other metadata schemes such as MPEG-7, TV-Anytime, P/meta and Dublin Core. The **SMPTE Metadata Dictionary** [MXF-RP210] is a thematically structured list of metadata elements, defined by a key, the size of the value and its semantics.

Dark Metadata is the term given to metadata that is unknown by an application. This metadata may be privately defined and generated, it may be new properties added or it may be standard MXF metadata not relevant to the application processing this MXF file. There are rules in the MXF standard on the use of dark metadata to prevent numerical or namespace clashes when private metadata is added to a file already containing dark metadata.

3.4 Multimedia Metadata Formats For Describing Multimedia Presentations

The formats listed in this section deal with multimedia presentations with appropriate support for metadata.

3.4.1 Synchronized Multimedia Integration Language (SMIL)

Responsible	Specification
http://www.w3.org/	[SMIL]

Representation	Content Type	Workflow	Domain	Industry
X	G	publish, distribution, presentation, interaction	generic	Web, mobile applications

The Synchronized Multimedia Integration Language (SMIL) [SMIL] is an XML-based 2-dimensional graphics language enabling simple authoring of interactive audiovisual presentations. SMIL is used to describe scenes with streaming audio, streaming video,

still images, text or any other media type. SMIL can be integrated with other web technologies such as XML, DOM, SVG, CSS and XHTML.

Next to media, a SMIL scene also consists of a spatial and temporal layout and supports animation and interactivity. SMIL also has a timing mechanism to control animations and for synchronization. SMIL is based on the download-and-play concept; it has also a mobile specification, SMIL Basic.

The SMIL 2.1 Metainformation module contains elements and attributes that allow description of SMIL documents. It allows authors to describe documents with a very basic vocabulary (meta element; inherited from SMIL 1.0), and in its recent version the specification introduces new capabilities for describing metadata using RDF.

Note: SMIL 3.0 is a Last Call Working Draft at time of publishing this XGR.

3.4.2 Scalable Vector Graphics (SVG)

Responsible	Specification
http://www.w3.org/	[SVG]

Representation	Content Type	Workflow	Domain	Industry
X	G	publish, presentation	generic	Web, mobile applications

Scalable Vector Graphics (SVG) [SVG] is a language for describing two-dimensional vector and mixed vector/raster graphics in XML. It allows for describing scenes with vector shapes (e.g. paths consisting of straight lines, curves), text, and multimedia (e.g. still images, video, audio). These objects can be grouped, transformed, styled and composited into previously rendered objects.

SVG files are compact and provide high-quality graphics on the Web, in print, and on resource-limited handheld devices. In addition, SVG supports scripting and animation, so SVG is ideal for interactive, data-driven, personalized graphics. SVG is based on the download-and-play concept. SVG has also a mobile specification, SVG Tiny, which is a subset of SVG.

Metadata which is included with SVG content is specified within the metadata elements, with contents from other XML namespaces such as Dublin Core or RDF.

3.5 Multimedia Metadata Formats For Describing Specific Domains Or Workflows

Metadata formats listed in this section focus on a specific domain (e.g. news) or are concerned with workflow issues such as MPEG-21.

3.5.1 NewsML-G2

Responsible	Specification

http://www.iptc.org/NAR/ [NewsML-G2]

Representation	Content Type	Workflow	Domain	Industry
X	G	publish	news	news agencies

For easing the exchange of news, the International Press Telecommunication Council (IPTC) has developed the News Architecture for G2-Standards [NewsML-G2] whose goal is *to provide a single generic model for exchanging all kinds of newsworthy information, thus providing a framework for a future family of IPTC news exchange standards*. This family includes NewsML-G2, SportsML-G2, EventsML-G2, ProgramGuideML-G2 or a future WeatherML. All are XML-based languages used for describing not only the news content (traditional metadata), but also their management, packaging, or related to the exchange itself (transportation, routing).

3.5.2 TVAnytime

Responsible	Specification
http://www.tv-anytime.org/	[TVAnytime]

Representation	Content Type	Workflow	Domain	Industry
X	G	distribute	Electronic Program Guides (EPG)	broadcast

The TV Anytime Forum is an association of organizations which seeks to develop specifications to provide value-added interactive services, such as the electronic program guide, in the context of TV digital broadcasting. The forum identified the metadata [TVAnytime] as one of the key technologies enabling their vision and have adopted MPEG-7 as the description language. They have extended the MPEG-7 vocabulary with higher-level descriptors, such as, for example, the intended audience of a program or its broadcast conditions.

3.5.3 MPEG-21

Responsible	Specification
http://www.iso.org/iso/en/prods-services/popstds/mpeg.html (ISO/MPEG)	[MPEG-21]

Representation	Content Type	Workflow	Domain	Industry
nX, X	G	annotate, publish, distribute	generic	generic

The MPEG-21 [MPEG-21] standard aims at defining a framework for multimedia delivery and consumption which supports a variety of businesses engaged in the trading of digital objects. MPEG-21 is quite different to its predecessors, as it is not focused on the representation and coding of content like MPEG-1 to MPEG-7 do, but instead focusing on filling the gaps in the multimedia delivery chain. MPEG-21 was developed with the vision in mind that it should offer users transparent and interoperable consumption and delivery of rich multimedia content. The MPEG-21 standard consists of a set of tools and builds on

its previous coding and metadata standards like MPEG-1, -2, -4 and -7, i.e., it links them together to produce a protectable universal package for collecting, relating, referencing and structuring multimedia content for the consumption by users (the Digital Item). The vision of MPEG-21 is to enable transparent and augmented use of multimedia resources (e.g. music tracks, videos, text documents or physical objects) contained in digital items across a wide range of networks and devices.

The two central concepts of MPEG-21 are *Digital Items*, a fundamental unit of distribution and transaction, and the concept of *Users* interacting with Digital Items: A User is any entity that interacts in the MPEG-21 environment or makes use of a Digital Item, and a Digital Item is a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. This entity is also the fundamental unit of distribution and transaction within this framework. In other words, the Digital Item groups multimedia resources (e.g. audio, video, image, text) and metadata (such as identifiers, licenses, content-related and processing-related information) within a standardized structure enabling interoperability among vendors and manufacturers.

The MPEG-21 standard consists of 18 parts of which the following are the most relevant for the scope of the MMSEM-XG:

- Part 2, Digital Item Declaration (DID), provides an abstract model and an XML-based representation thereof which is used to define Digital Items. The DID Model defines digital items, containers, fragments or complete resources, assertions, statements, choices/selections, and annotations on digital items.
- Part 3, Digital Item Identification and Description (DII), is concerned with the ability to identify and refer to complete or partial Digital Items.
- Part 5, Rights Expression Language (REL), provides a machine-readable language to declare rights and permissions using the terms as defined in the Rights Data Dictionary.
- Part 17, Fragment Identification for MPEG Media Types, specifies a syntax for identifying parts (e.g., track of a CD/DVD) of MPEG resources via Uniform Resource Identifiers (URIs).

MPEG-21 identifies and defines the mechanisms and elements needed to support the multimedia delivery chain as described above, as well as the relationships between and the operations supported by them. Within the parts of MPEG-21, these elements are elaborated by defining the syntax and semantics of their characteristics, such as interfaces to the elements.

Note: For an overview on MPEG-21, see also MPEG-21 Overview v.5 via Leonardo Chiariglione.

3.5.4 EBU P/Meta

Responsible	Specification
http://www.ebu.ch/ European Broadcasting Union (EBU)	[EBU P/Meta]

Representation	Content Type	Workflow	Domain	Industry
nX, X	A, V	publish	generic	broadcast

The EBU P/Meta working group has designed this standard as a metadata vocabulary [EBU P/Meta] for programme exchange in the professional broadcast industry. It is not intended as an internal representation of a broadcaster's system. P/Meta has been designed as metadata format in a business-to-business scenario to exchange broadcast programme related metadata between content producers, content distributors and archives. The P/Meta definition uses a three-layer model: the *definition layer* (i.e. the semantic of the description), the *technology layer* defines the encoding used for exchange (currently KLV — key, length, value — and XML representations are specified), and the lowest layer, the *data interchange layer*, which is out of scope of the specification. P/Meta consists of a number of attributes (some of them with a controlled list of values), which are organized into sets. The standard covers the following types of metadata:

- Identification
- Technical metadata
- Programme description and classification
- Creation and production information
- Rights and contract information
- Publication information

Note: it is worth noting that EBU is working on replacing P/Meta by NewsML-G2.

3.6 Other Multimedia Metadata Related Formats

3.6.1 Dublin Core (DC)

Responsible	Specification
http://dublincore.org/	[Dublin Core]

Representation	Content Type	Workflow	Domain	Industry
X, R	G	publish	generic	generic

The Dublin Core Metadata Initiative (DCMI) has defined a set of elements [Dublin Core] for cross-domain information resource description. The set consists of a flat list of 15 elements describing common properties of resources, such as title, creator etc. Dublin Core recommends using controlled vocabularies for providing the values for these elements.

3.6.2 XMP and IPTC Metadata for XMP

Responsible	Specification
http://www.adobe.com/	[XMP]

Representation	Content Type	Workflow	Domain	Industry
X, R	G	annotate, publish, distribute	generic	generic

The main goal of XMP [XMP] is to attach more powerful metadata to media assets in order to enable a better management of multimedia content, and better ways to search

and retrieve content in order to improve consumption of these multimedia assets. Furthermore XMP aims to enhance reuse and repurposing of content and to improve interoperability between different vendors and systems.

The Adobe XMP specification standardizes the definition, creation, and processing of metadata by providing a data model, storage model (serialization of the metadata as a stream of XML), and formal schema definitions (predefined sets of metadata property definitions that are relevant for a wide range of applications). XMP makes use of RDF in order to represent the metadata properties associated with a document.

With XMP, Adobe provides a method and format for expressing and embedding metadata in various multimedia file formats. It provides a basic data model as well as metadata schemas for storing metadata in RDF, and provides storage mechanism and a basic set of schemas for managing multimedia content like versioning support. The most important components of the specification are the data model and the pre-defined (and extensible) schemas:

- XMP Data Model is derived from RDF and is a subset of the RDF data model. It provides support for metadata properties to attach metadata to a resource. Properties have property values, which can be structured (structured properties) or simple types or arrays. Properties may also have properties (property qualifiers) which may provide additional information about the property value.
- XMP Schemas consist of predefined sets of metadata property definitions. Schemas are essentially collections of statements about resources which are expressed using RDF. It is possible to define new external schemas, to extend the existing ones or to add some if necessary. There are some predefined schemas included in the specification like a Dublin Core Schema, a basic rights schema or a media management schema.

There is a growing number of commercial applications that already support XMP. For example, the International Press and Telecommunications Council (IPTC) has integrated XMP in its Image Metadata specifications and almost every Adobe application like Photoshop or In-Design supports XMP. *IPTC Metadata for XMP* can be considered as a multimedia metadata format for describing still images and could actually be soon the most used one.

4. Multimedia Ontologies

This section discusses some known approaches for converting existing multimedia metadata into RDF [RDF Primer] / OWL [OWL Guide] for the purpose of interoperability, reasoning, etc.

Note: The formalizations presented in the following are subsumed under the more common term **Multimedia Ontology**, hence the title.

4.1 VRA - RDF/OWL

At the time of writing, there exists no commonly accepted mapping from VRA Core to RDF/OWL. However, at least two conversions have been proposed:

  - [RDF/OWL Representation of VRA](#) by Mark van Assem, Vrije Universiteit Amsterdam, and
  - [RDF/OWL VRA ontology](#) from SIMILE.

4.2 Exif - RDF/OWL

Recently, there has been efforts to represent the Exif metadata tags in an RDF-S ontology. The two approaches presented here are semantically very similar, yet are both described for completeness:

  - The [Kanzaki Exif RDF Schema](#) provides an encoding of the basic Exif metadata tags in RDF Schema. We also note here that relevant domains and ranges are used as well. Kanzaki Exif additionally provides an Exif conversion service, Exif-to-RDF, which extracts Exif metadata from images and automatically maps it to the RDF encoding.
  - The [Norm Walsh Exif RDF Schema](#) provides another encoding of the basic Exif metadata tags in RDF Schema. Walsh Exif additionally provides JPEGRDF, which is a Java application that provides an API to read and manipulate Exif metadata stored in JPEG images. Currently, JPEGRDF can extract, query, and augment the Exif/RDF data stored in the file headers. In particular, we note that the API can be used to convert existing Exif metadata in file headers to the schema defined in Walsh Exif.

4.3 DIG35 - RDF/OWL

The [DIG35 ontology](#), developed by the [IBBT Multimedia Lab](#) (University of Ghent) in the context of the W3C Multimedia Semantics Incubator Group, provides an OWL Schema covering the entire DIG35 specification. For the formal representation of DIG35, no other ontologies have been used. However, relations with other ontologies such as Exif, FOAF, etc. will be created to give the DIG35 ontology a broader semantic range. The DIG35 ontology is an OWL Full ontology.

4.4 MPEG-7 - RDF/OWL

For MPEG-7, there is no commonly agreed upon mapping to RDF/OWL. However, this section lists existing approaches regarding the translation of (parts of) [MPEG-7](#) into RDF/OWL.

4.4.1 MPEG-7 Upper MDS Ontology by Hunter

Ontology Source	Description
[http://metadata.net/mpeg7](#)	[[Hunter, 2001](#)]

Chronologically the first one, this MPEG-7 ontology was firstly developed in RDFS, then converted into DAML+OIL, and is now available in OWL-Full. The ontology covers the upper part of the Multimedia Description Scheme (MDS) part of the MPEG-7 standard. It comprises about 60 classes and 40 properties.

4.4.2 MPEG-7 MDS Ontology by Tsinaraki

Ontology Source	Description
http://elikonas.ced.tuc.gr/ontologies/av_semantics.zip	[Tsinaraki et.al., 2004]

Starting from the ontology developed by Hunter [Hunter, 2001] this MPEG-7 ontology covers the full Multimedia Description Scheme (MDS) part of the MPEG-7 standard. It contains 420 classes and 175 properties. This is an OWL DL ontology.

4.4.3 MPEG-7 Ontology by Rhizomik

Ontology Source	Description
http://rhizomik.net/ontologies/mpeg7ontos	[Garcia et.al., 2005]

This MPEG-7 ontology has been produced fully automatically from the MPEG-7 standard in order to give it a formal semantics. For such a purpose, a generic mapping XSD2OWL has been implemented. The definitions of the XML Schema types and elements of the ISO standard have been converted into OWL definitions according to the table given in [Garcia et.al., 2005]. This ontology could then serve as a top ontology thus easing the integration of other more specific ontologies such as MusicBrainz. The authors have also proposed to transform automatically the XML data (instances of MPEG-7) into RDF triples (instances of this top ontology).

This ontology aims to cover the whole standard and it thus the most complete one (with respect to the previous mentioned). It contains finally 2372 classes and 975 properties. This is an OWL Full ontology since it employs the rdf:Property construct to cope with the fact that there are properties that have both datatype and object type ranges.

4.4.4 Core Ontology for Multimedia (COMM)

Ontology Source	Description
http://multimedia.semanticweb.org/COMM/	[Arndt et.al., 2007]

The Core Ontology for Multimedia (COMM) [Arndt et.al., 2007] is based on both the MPEG-7 standard and the DOLCE [Masolo et.al., 2002] foundational ontology. COMM is an OWL DL ontology. It is composed of multimedia patterns specializing the DOLCE design patterns for Descriptions & Situations and Information Objects. The ontology covers a very large part of the MPEG-7 standard. The explicit representation of algorithms in the multimedia patterns allows also to describe the multimedia analysis steps, something that is not possible in MPEG-7.

4.4.5 aceMedia Visual Descriptor Ontology

Ontology Source	Description
http://www.acemedia.org/aceMedia/files/software/m-ontomat/acemedia-visual-descriptor-ontology-v09.rdfs	[VDO]

The Visual Descriptor Ontology (VDO) developed within the aceMedia project for semantic multimedia content analysis and reasoning, contains representations of MPEG-7 visual descriptors and models Concepts and Properties that describe visual characteristics of objects. The term descriptor refers to a specific representation of a visual feature (color, shape, texture etc) that defines the syntax and the semantics of a specific aspect of the feature. For example, the dominant color descriptor specifies among others, the number and value of dominant colors that are present in a region of interest and the percentage of pixels that each associated color value has. Although the construction of the VDO is tightly coupled with the specification of the MPEG-7 Visual Part, several modifications were carried out in order to adapt to the XML Schema provided by MPEG-7 to an ontology and the data type representations available in RDF Schema.

4.5 Mindswap Image Region Ontology

Ontology Source	Description
http://www.mindswap.org/2005/owl/digital-media	[Halaschek-Wiener et.al., 2005]

The Mindswap digital-media is an OWL ontology which models concepts and relations covering various aspects of the digital media domain. The main purpose of the ontology is to provide the expressiveness to assert what is depicted within various types of digital media, including image and videos. The ontology defines concepts including image, video, video frame, region, as well as relations such as depicts, regionOf, etc. Using these concepts and their associated properties, it is therefore possible to assert that an image/imageRegion depicts some instance, etc.

4.6 Audio Ontologies

The audio community if quite active in disseminating Semantic Web technologies; the known formalisations in the realm of audio (mainly music) are:

- Music Ontology Specification by Frederick Giasson et Yves Raimond (Zitgist). The Music Ontology Specification provides main concepts and properties fo describing music (i.e. artists, albums and tracks) on the Semantic Web. Based on (or inspired by) the MusicBrainz MusicBrainz editorial metadata.
- Kanzaki's music vocabulary. A vocabulary to describe classical music and performances. Classes (categories) for musical work, event, instrument and performers, as well as related properties are defined.
- Music Recommendation by Oscar Celma, Universitat Pompeu Fabra. Foafing the Music system [Celma, 2006] uses the Friend of a Friend (FOAF) and RDF Site Summary (RSS) vocabularies for recommending music to a user, depending on the user's musical tastes and listening habits. It comprises a simple OWL-DL ontology that defines basic information of artists (and their relationships), and songs. It includes some descriptors automatically extracted from the audio (beats per minute, key and mode, intensity, etc.).

References

[AAF]

Advanced Media Workflow Association (formerly AAF Association), AAF Specifications

[Arndt et.al., 2007]

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

[Celma, 2006]

O. Celma. Foafing the Music: Bridging the Semantic Gap in Music Recommendation . Semantic Web Challenge 2006.

[DAVP]

W. Bailer and P. Schallauer The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems. In Proc. of 12th International Multi-Media Modeling Conference, Beijing, CN, 2006.

[DIG35]

Digital Imaging Group (DIG), DIG35 Specification - Metadata for Digital Images - Version 1.0 August 30, 2000

[Dublin Core]

The Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, Version 1.1: Reference Description (2006-12-18)

[EBU P/Meta]

European Broadcasting Union, EBU Tech 3295: The EBU Metadata Exchange Scheme version 1.2 - Publication Release

[Exif]

Standard of Japan Electronics and Information Technology Industries Association, Exchangeable image file format for digital still cameras: Exif Version 2.2

[Garcia et.al., 2005]

R. Garcia and O. Celma. Semantic Integration and Retrieval of Multimedia Metadata . In Proc. of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005), Galway, Ireland, 7 November 2005.

[Halaschek-Wiener et.al., 2005]

C. Halaschek-Wiener, A. Schain, J. Golbeck, M. Grove, B. Parsia and J. Hendler. A Flexible Approach for Managing Digital Images on the Semantic Web . In Proc. of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005), Galway, Ireland, 7 November 2005.

[Hardman, 2005]

Lynda Hardman. Canonical Processes of Media Production . In Proc. of the ACM workshop on Multimedia for human communication. ACM Press, 2005.

[Hunter, 2001]

J. Hunter. Adding Multimedia to the Semantic Web — Building an MPEG-7 Ontology . In International Semantic Web Working Symposium (SWWS 2001) , Stanford University, California, USA, July 30 - August 1, 2001

[ID3]

Martin Nilsson et. al., ID3v2 documents

[Masolo et.al., 2002]

C. Masolo and S. Borgo and A. Gangemi and N. Guarino and A. Oltramari and L. Schneider. The WonderWeb Library of Foundational Ontologies (WFOL). Technical Report, WonderWeb Deliverable 17, 2002.

[MPEG-7]

Information Technology - Multimedia Content Description Interface (MPEG-7). Standard No. ISO/IEC 15938:2001, International Organization for Standardization(ISO), 2001

[MPEG-21]

Information Technology - Multimedia framework (MPEG-21). Standard ISO/IEC TR 21000-1:2004, International Organization for Standardization(ISO), 2004

[MPEG-7-Profiles]

Information Technology - Multimedia Content Description Interface -- Part 9: Profiles and levels. Standard No. ISO/IEC 15938-9:2005, International Organization for Standardization(ISO), 2005

[MusicBrainz]

MusicBrainz (MetaBrainz Foundation), MusicBrainz Metadata Initiative 2.1

[MusicXML]

Recordare, MusicXML Definition Version 2.0

[MXF]

SMPTE, Material Exchange Format (MXF) - File Format Specification (Standard). SMPTE 377M, 2004.

[MXF-DMS-1]

SMPTE, Material Exchange Format (MXF) - Descriptive Metadata Scheme-1. SMPTE 380M, 2004.

[MXF-RP210]

SMPTE, Metadata Dictionary Registry of Metadata Element Descriptions. SMPTE RP210.8, 2004.

[Ossenbruggen, 2004]

J. van Ossenbruggen, F. Nack, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part I). In: IEEE Multimedia 11(4), pp. 38-48 October-December 2004

[Nack, 2005]

F. Nack, J. van Ossenbruggen, and L. Hardman. That Obscure Object of Desire: Multimedia Metadata on the Web (Part II). In: IEEE Multimedia 12(1), pp. 54-63 January-March 2005

[NISO Z39.87]

American National Standards Institute, ANSI/NISO Z39.87-2006: Data Dictionary - Technical Metadata for Digital Still Images

[NewsML-G2]

IPTC, News Architecture (NAR) for G2-Standards Specifications (released 30th May, 2007)

[OWL Guide]

OWL Web Ontology Language Guide, Michael K. Smith, Chris Welty, and Deborah L. McGuinness, Editors, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/owl-guide/

[OWL Semantics and Abstract Syntax]

OWL Web Ontology Language Semantics and Abstract Syntax, Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks, Editors, W3C Recommendation 10 February 2004, http://www.w3.org/TR/owl-semantics/

[PhotoRDF]

W3C Note 19 April 2002, Describing and retrieving photos using RDF and HTTP

[PhotoStuff]

PhotoStuff Project, http://www.mindswap.org/2003/PhotoStuff/

[RDF Primer]

 RDF Primer, F. Manola, E. Miller, Editors, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/rdf-primer/

[RDF Syntax]

 RDF/XML Syntax Specification (Revised) , Dave Beckett, Editor, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/rdf-syntax-grammar/

[SMIL]

 W3C Recommendation 13 December 2005, Synchronized Multimedia Integration Language (SMIL 2.1) - Chapter 8. The SMIL 2.1 Metainformation Module

[Smith et.al., 2006]

 J. R. Smith and P. Schirling. Metadata Standards Roundup. IEEE MultiMedia, vol. 13, no. 2, pp. 84-88, Apr-Jun, 2006.

[SVG]

 W3C Recommendation 14 January 2003, Scalable Vector Graphics (SVG) 1.1 Specification - Chapter 21. Metadata

[Tsinaraki et.al., 2004]

 C. Tsinaraki, P. Polydoros and S. Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In Proc. of 3rd International Conference on Image and Video Retrieval (CIVR 2004), Dublin, Ireland, 21-23 July 2004.

[TVAnytime]

 IPTC, WG Metadata - Important Documents

[VDO]

 aceMedia Visual Descriptor Ontology, http://www.acemedia.org/aceMedia/reference/resource/index.html

[VRA Core]

 Visual Resources Association Data Standards Committee, VRA Core Categories, Version 4.0, http://www.vraweb.org/projects/vracore4/index.html

[XML NS]

 Namespaces in XML, Bray T., Hollander D., Layman A. (Editors), World Wide Web Consortium, 14 January 1999, http://www.w3.org/TR/REC-xml-names/

[XMP]

 Adobe, XMP Specification

Acknowledgments

$Id: Overview.html,v 1.5 2007/08/09 09:53:58 rtroncy Exp $