Interlinking Multimedia

How to Apply Linked Data Principles to Multimedia Fragments

Michael Hausenblas
DERI, National University of Ireland, Galway
Ireland
michael.hausenblas@deri.org

Raphaël Troncy
CWI Amsterdam
The Netherlands
raphael.troncy@cwi.nl

Tobias Bürger
STI Innsbruck
Austria
tobias.buerger@sti2.at

Yves Raimond
BBC Audio & Music interactive
United Kingdom
yves.raimond@bbc.co.uk

ABSTRACT

In this paper, we introduce interlinking multimedia (iM), a pragmatic way to apply the linked data principles to fragments of multimedia items. We report on use cases showing the need for retrieving and describing multimedia fragments. We then introduce the principles for interlinking multimedia in the Web of Data, discussing potential solutions which sometimes highlight controversial debates regarding what the various representations of a Web resource span. We finally present methods for enabling a widespread use of interlinking multimedia.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous; I.7.2 [**Document Preparation**]: Languages and systems, Markup languages, Multi/mixed media, Standards; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Representation languages

General Terms

Languages, Standardization

Keywords

Linked Data, Media Fragments, Media Annotations

1. MOTIVATION

Multimedia content is easy to produce but rather hard to find and to reuse on the Web. Digital photographs can be easily uploaded, communicated and shared in community portals such as Flickr, Picasa and Riya, while video are available on portals such as YouTube, DailyMotion, Metacafe or Vimeo to name a few. These systems allow their users to manually tag, comment and annotate the digital content, but they lack a general support for fine-grained semantic descriptions and look-up, especially when talking about things "inside" multimedia content, such as an object in a video or a person depicted in a still image.

Figure 1 illustrates the problem. A photo is host and shared on a well-known photo portal. One can draw and tag a particular region ("notes" in Flickr) in a picture to state that this region actually depicts a certain person. Both the region definition and the annotation are represented in a proprietary format and locked-into the system. It can not be used by other applications across the Web.

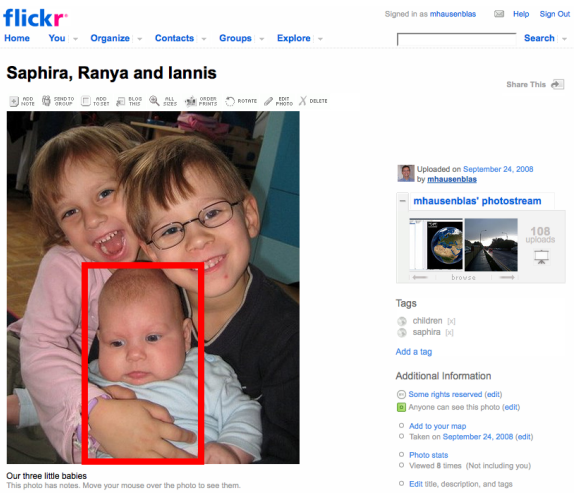

Figure 1: iM example

For addressing this issue, the multimedia semantics community has focused on developing so-called multimedia ontologies [2] where fragments are firstly defined before being used in semantic annotations. This approach has not yet proved it can scale the Web. Worst, it also means an indirection since the *aboutness* of a semantic description is a piece of XML document defining a multimedia fragment and not the fragment itself. On the other hand, the linked data community has successfully applied the linked data principles to publicly available datasets offering a light-weight,

LDOW2009, April 20, 2009, Madrid, Spain.
.

scalable solution for annotating web resources.

The aim of this paper is to discuss how to implement the linked data principles along with media fragments, yielding what we call "interlinking Multimedia" (iM)[1]. The contribution of this paper is a theoretical and practical framework of the required technologies and discussions of their suitability in order to be compliant with the Web architecture. We present a technology that enables to address multimedia fragments in URIs and we critically review some interlinking methods.

The remainder of the paper is structured as follows. In the Section 2, we discuss use cases and requirements for addressing and describing multimedia fragments. Based on these use cases, we state the interlinking multimedia principles as follows:

1. Apply **linked data principles** for fine-grained identification and description of spatio-temporal multimedia fragments (Section 3);

2. Deploy **legacy multimedia metadata** formats such as EXIF, ID3, XMP on the Web of Data (Section 4);

3. Discuss a set of **specialized interlinking methods for multimedia** (Section 5).

We reflect our work in the light of the existing work in Section 6 and finally conclude in Section 7.

2. USE CASES AND REQUIREMENTS

In the previous section, we have made a case for addressing and describing a particular region of an image (anchor value in [7]). The current Web architecture [16] provides a means for uniquely identifying sub-parts of resources using URI fragment identifiers (e.g. for referring to a part of an HTML or XML document). However, for almost any other media type, the semantics of the fragment identifier has either not been specified or is not commonly accepted. Providing an agreed upon way to localize sub-parts of multimedia objects (e.g. sub-regions of images, temporal sequences of videos or tracking moving objects in space and in time) is fundamental [30]. In this section, we report on use cases for interlinking multimedia on the Web (Section 2.1) and we extract the requirements for iM (Section 2.2).

2.1 Use Cases

We observe a general demand in several communities for annotation tools enabling to specify links between compound objects or parts within these objects as well as the type of these relationships. For example, media researchers want to annotate and interrelate segments between books, screenplays or different film versions [29]. The following four use cases further demonstrate the need for interlinking multimedia in the Web of Data.

2.1.1 Sharing video clips on Twitter

Silvia is a big fan of Tim's research keynotes. She used to watch numerous videos starring Tim for following his research activities and often would like to share the highlight announcements with her collaborators. The company that she runs uses the Twitter[2] service to communicate and stay connected among business partners. Silvia is interested in TweeTube[3] that will allow her to share video directly on Twitter but she would like to point and reference only small temporal sequences of these longer videos. She would like to have a simple interface, similar to VideoSurf[4], to edit the start and end time points delimiting a particular sequence, and get back in return the media fragment URI to share with the rest of the world. She would also like to send her comments and (semantic) annotations about this particular video fragment.

2.1.2 Annotating faces in my personal photos

Michael is really enthusiastic with the new features of the iPhoto 2009 suite[5]. He went through the guide and he is happy to see that annotating people on his personal photos will be easier than ever. As soon as Michael uses the software, it learns who Michael's friends and relatives are, and suggests annotations as faces are automatically recognized in his pictures using some visual processing techniques. Bounding boxes around faces are therefore drawn on the photos and can be exported on Flickr for sharing. Michael is also a *Linked Data guru* and would like to tag his photos not with the name of his friends, but with the URI that identifies them on the Web of Data. Michael stores RDF annotations of all these spatial fragment URIs and hopes to create an artistic collage of his family.

2.1.3 Tracking your favorite artists on BBC Music

BBC Music[6] aims to provide a comprehensive guide to music content across the BBC. The service provides information about artists who appear on BBC programmes[7] or who have been covered in one of their reviews. Each artist is interlinked with biographical information (where available) supplied by Wikipedia, and with BBC programmes in which it has been played. Frank loves this service so he can quickly and easily find the kind of shows that might suit his taste. He issues a query for his two favorite German composers, Wagner and Strauss, and gets a list of media fragment URIs pointing to segments from tracklists of various performances broadcasted this week.

2.1.4 Watching named video clips on a mobile phone

Yves is a busy person. He doesn't have time to attend all meetings that he is supposed to. These are generally video recorded and podcasted on the internal Web site of his company together with a full text speech transcription aligned with the video. Yves often uses his mobile smart phone for accessing Web resources while traveling. He receives a daily digest email from the system containing a list of media fragment URIs pointing to video clips where his name has been pronounced, together with the term 'ACTION', during meetings. While on his next trip, Yves goes though his email backlog and watches the video clips by simply clicking on the links. The media server of his company dynamically composes a valid media resource from the URIs that Yves is requesting, so that Yves' video player just plays the right sequences where an action has been given to him.

[1] http://www.interlinkingmultimedia.info/
[2] http://twitter.com/
[3] http://www.tweetube.com/
[4] http://www.videosurf.com/
[5] http://www.apple.com/ilife/iphoto/
[6] http://www.bbc.co.uk/music/beta
[7] http://www.bbc.co.uk/programmes

2.2 Requirements

Several aspects and requirements for iM have already been discussed [5]. In this section, we argue that interlinking multimedia resources at a fine-grained level requires to deal with the addressing and the description of multimedia fragments.

2.2.1 Addressing Multimedia Fragments

We are working within the W3C Media Fragments Working Group[8] to provide a URI-based mechanism to address fragments of image, audio and video resources. Our first assumption is that an audio or a video resource has a single unified timeline.

The use cases described in the Section 2.1 highlight the need for being able to address parts of multimedia resources. By parts, we mean playable resources that can be extracted from the parent resources according to a number of dimensions, namely: time, space, track and names. The *time* dimension allows to address a temporal sequence of an audio or a video resource. The *space* dimension allows to address a rectangle bounding box of a frame or still image. The *track* dimension allows to address a particular track of multimodal resource if the container format exposes such a notion. Finally, the *name* dimension is a convenient shortcut for a combination of any of the three other dimensions that can be further referred by a name under the condition that the container format allows such a markup (e.g. a chapter name in a DVD).

Numerous codec and container formats are used on the Web. A URI denoting a media fragment should be agnostic on these formats but bound to what they can expose in the compressed domain[9] (i.e. without any further transcoding operation). The HTTP protocol should at least be supported while we observe that any solution should be compatible with the notion of media fragment defined in the RTSP protocol[10].

Media fragments are really parts of a parent resource. The use of URI fragment [4] seems therefore appropriate to specify these media fragments. As for any URI fragment, access to the parent URI shall be possible in order to inspect its context.

2.2.2 Describing Multimedia Fragments

In order to be become part of the LOD cloud, iM must follow the linked data principles (see Section 3.1). Metadata descriptions have to be interoperable in order to reference and integrate parts of the described resources. The diversity of media content types, application scenarios and domains directly translates to the existence of a huge number of (partially) diverse metadata formats [15].

The integration of these formats, though often desirable [31] is difficult due to syntactic and semantic interoperability problems. Solutions should further take into account the characteristics of multimedia whose semantics – when interpreted by a user – are typically derived based on his/her experiences, culture and knowledge. Thus, solutions should consider *provenance information* and contextual information (e.g. who says what and when) when describing fragment of multimedia resources. These issues are in particular addressed within the W3C Media Annotations Working Group[11] and in other related forums.

[8] `http://www.w3.org/2008/WebVideo/Fragments/`

[9] The abilities and limitations of most of the multimedia formats are described in `http://www.w3.org/2008/WebVideo/Fragments/wiki/Types_of_Fragment_Addressing`

[10] `http://www.ietf.org/rfc/rfc2326.txt`

3. IM PRINCIPLES

As discussed above, the motivation for introducing iM stems from the fact that we currently do not have proper means to address and describe fragments of multimedia resources in the Web of data. We believe that we can overcome these limitations by defining a URI-based mechanism to address media fragments and by applying the linked data principles to those fragments. Our ultimate goal is to derive both semantic and audio-visual representations from multimedia resources on the Web.

In the context of the Web of Data, we deal with documents (e.g. a JPG file) and things (e.g. a person). The former is generally called an *Information Resource* while the later will be referred as a *Non-Information Resource*[12]. The W3C's "Architecture of the World Wide Web, Volume One" (AWW) [16] specifies that globally unique identifiers (URIs) are used to denote both things and documents. In the following, we first present the linked data principles (Section 3.1) before describing how iM can be deployed for both information resources (Section 3.2) and non-information resources (Section 3.3).

In both cases, we consider a related issue regarding the *addressing* versus the *description* of fragments of multimedia resources. The former case will refer to the ability of getting only the multimedia fragment served using the Web architecture while the latter will refer to the ability of getting semantic metadata about the media fragment.

3.1 Linked Data Principles

The basic idea of linked data was outlined by Tim Berners-Lee [3]:

1. All resources should be identified using *URIs*;
2. All URIs should be *dereferenceable*, that is HTTP URIs, as it allows looking up the resources identified;
3. When looking up an URI, it leads to more (useful) data about that resource;
4. Links to other URIs should be included in order to enable the discovery of more data.

We note that these four linked data principles are agnostic regarding the type of the resource. In the following, we revisit these four principles for interlinking multimedia. In particular, we discuss how looking up an URI to lead to more data can be realized.

3.2 iM for Information Resources

Let's imagine that Silvia would like to share with her colleagues a specific part of a recent podcast from Tim Berners Lee on the BBC. More particularly, she is interested in sending a link pointing to the sequence comprised between the seconds 15 and 45 of this podcast. Following the temporary Media Fragments URI syntax[13], Silvia will build the URI `http://www.example.org/myPodcast.mp3#t=15,45`.

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

[12] An ontology implementing the concepts discussed in the Generic URIs "Design Issues" note is available at `http://www.w3.org/2006/gen/ont.rdf`

[13] `http://www.w3.org/2008/WebVideo/Fragments/wiki/Syntax`

3.2.1 Retrieving Media Fragments

Conrad is working with Silvia. He is interested in listening to just this sequence, and would prefer to not download the one hour podcast. The following interaction could happen between his browser (the user agent) and the server. Conrad has a smart user agent that can further process requests containing a media fragment by stripping out the fragment part, but encoding it into a range header. The following GET command will therefore be issued:

```
GET http://www.example.org/myPodcast.mp3
Accept: application/mp3
Range: seconds=15-45
```

The server has a module for slicing on demand multimedia resources, that is, establishing the relationship between seconds and bytes, extract the bytes corresponding to the requested fragment, and add the new container headers in order to serve a playable resource. The server will then reply with the closest inclusive range in a 206 HTTP response:

```
HTTP/1.1 206 Partial Content
Accept-Ranges: bytes, seconds
Content-Length: 1201290
Content-Type: audio/mpeg
Content-Range: seconds 14.875-45.01/321
```

The user agent will then have to skip 0.125s to start playing the multimedia fragment as 15s. We observe that the relationship between bytes and seconds is in the general case unknown. The Media Fragment WG is considering a technical solution involving a second rountrip between the user agent and the server for establishing this mapping.

3.2.2 Describing Media Fragments

Conrad is also interested in retrieving the semantic description of this fragment to feed his semantic web agent. He could issue a similar request changing simply the accept header:

```
GET http://www.example.org/myPodcast.mp3
Accept: application/rdf+xml
Range: seconds=15-45
```

Providing an adequate configuration, the server could return an RDF file containing the semantic annotations of this media fragment. The additional "HTTP Link: header" proposal[14] could further establish the relationship between the mp3 file and the RDF file using the `rdfs:seeAlso` property, the resource referenced by the request URI being then the subject of the assertion.

```
HTTP/1.1 200 OK
Accept-Ranges: bytes, seconds
Content-Length: 1088
Content-Location: http://www.example.org/myPodcast.rdf
Content-Type: application/rdf+xml
Link: <http://www.example.org/myPodcast.mp3>;
 rel="http://www.w3.org/2000/01/rdf-schema#seeAlso";
Vary: accept
```

[14] http://tools.ietf.org/html/draft-nottingham-http-link-header-04

3.2.3 Discussion

In this scenario, we have considered that the podcast resource is available in several different representations. Therefore, content negotiation can be used to serve alternatively an excerpt of the audio file, or a semantic description of this fragment depending on the HTTP accept header. We observe for example that this is how the jigsaw web server [18] is configured. As benefits, it works right away with all text-based browsers (lynx, emacs with emacsspeak, etc.) and the output can be rendered directly by selecting, e.g., the transcript of the audio file contained in the description from the RDF file. An RDF crawler will be able to get all the descriptions of a collection of audio files to create a knowledge database, just by asking for the right MIME type.

However, this solution boils down to say that the RDF description of the podcast and the podcast itself are both representations of the same information resource. On one hand, we observe that the Web Accessibility Guidelines[15] define two *equivalent* content when both fulfill essentially the same function or purpose upon presentation to the user. For example, the text "Tim Berners Lee promoting the Web of Data" might convey the same information as an excerpt of an audio podcast when presented to (deaf) users. On the other hand, some voices[16] in the Web of Data community consider that a description of a multimedia resource is not the same as the resource itself since it cannot convey all its perceptual and cognitive effect. Consequently, following this way of thought, content negotiating between these two resources would be just plain wrong. Embedding the multimedia resource within an HTML document on one side, and providing RDF metadata within this document on the other side would, however, work since metadata about the resource would convey the same information as the resource and thus can be subject to content negotiation. We conclude that discovery and HTTP[17] is a controversial issue for which we advocate a TAG resolution.

The use of a URI fragment for addressing a media fragment is also problematic. The URI RFC states [4]:

> "The semantics of a fragment identifier are defined by the set of representations that might result from a retrieval action on the primary resource. The fragment's format and resolution is therefore dependent on the media type [RFC2046] [6] of a potentially retrieved representation, even though such a retrieval is only performed if the URI is dereferenced. If no such representation exists, then the semantics of the fragment are considered unknown and are effectively unconstrained. Fragment identifier semantics are independent of the URI scheme and thus cannot be redefined by scheme specifications."

Therefore, it might be necessary to register new media-types defining the semantics of a fragment for each media formats using for example sub-class/class hierarchies provided by the IANA registry.

[15] http://www.w3.org/TR/WAI-WEBCONTENT-TECHS/#glossary

[16] http://chatlogs.planetrdf.com/swig/2009-02-09.html#T15-09-20

[17] http://www.hueniverse.com/hueniverse/2008/09/discovery-and-h.html

3.3 iM for General Resources

Let's now imagine that Michael would like to annotate some specific parts of his personal photos. More precisely, he would like to highlight the face of his new born daughter and send a media fragment URI to his family and relatives (Figure 2). Following the temporary Media Fragments URI

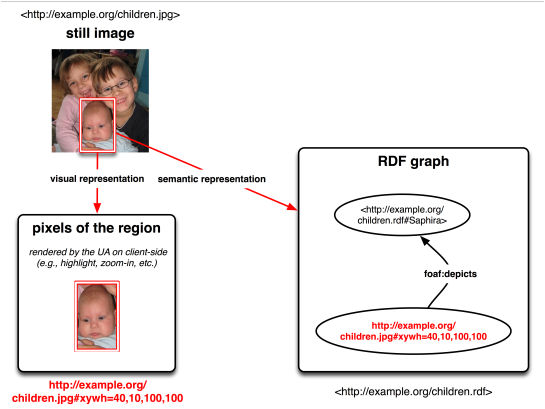

Figure 2: Interlinking Spatial Media Fragments

syntax, Michael will build the URI `http://www.example.org/children#xywh=40,10,100,100`. He can then further produce the annotation depicted in the listing 1:

```
<http://www.example.org/children#xywh=40,10,100,100>
    foaf:depicts :Saphira .
```

Listing 1: Description of a spatial region depicting a person.

The media fragment URI denotes here a part of a thing. Let's have a look at the communication between a (smart) user agent and a server with such a URI request. In case the accept header is image/jpeg:

```
GET http://www.example.org/children
Accept: image/jpeg
Range: pixels=40,10,100,100
```

The server will answer with a 307 Temporary Redirect[18] response indicating the location of the image file.

```
HTTP/1.1 307 Temporary Redirect
Location: http://www.example.org/children.jpg
Content-Type: image/jpeg
```

This response code requires that the request should be repeated with another URI, for example the one specified in the Location header. Similarly, a GET with a different accept header (e.g. application/rdf+xml) could return a 307 response code with the Location header pointing to the location of the RDF file. We conclude that apply iM for general resources might imply one more round-trip due to the fact that the URI denotes a non-information resource.

[18]The Web of Data community tends to use the 303 See Other response code. However, this HTTP response has originally been defined for changing the HTTP (verb) method, and the 307 code seems to be more appropriate in our case.

3.4 Summary

Let us review the adequacy of the four general linked data principles against the iM principles stated above. We acknowledge that the first and the second principle has been respected as, for example, `http://example.org/children.jpg#xywh=40,10,100,100` is a valid HTTP URI and the second principle can easily be done substituting e.g., `:Saphira` with a DBpedia resource.

The problem comes down to the third principle: "When looking up an URI, it leads to more (useful) data about that resource". Hence, when dereferencing `http://www.example.org/myPodcast.mp3#t=15,45`, what we want to have is both an audio and a semantic representation of the resource, i.e. the bytes corresponding to a particular playable sequence of an audio file and the RDF triples that might describe its transcript. We have argued that content negotiation could technically be used for this purpose but that it is unclear whether it is appropriate with the spirit of the Web architecture. We have also argued that the link header proposal is a viable alternative.

We observe also that there are cases where multimedia resources "embed" semantic annotations. For example, an image can have some XMP metadata represented in RDF stored in its header. We finally note that the whole RDF file is systematically returned when the user agent requests it. The problem is that it is hard to map the fragment of a multimedia resource with the corresponding piece in the RDF description of this resource.

4. LEGACY MULTIMEDIA METADATA

We have described in the previous section a technology that realizes the first iM principle: how to identify in a URI a fragment of a multimedia resource. The second principle states that legacy multimedia metadata formats should be deployed. Actually, descriptions about multimedia resources must be interoperable in order to enable the interlinking of the described resources. Semantic technologies have already been considered as a viable solution to leverage these interoperability issues [31]. For the description of multimedia resources, a plethora of metadata formats are in use, causing interoperability issues. Further, a multitude of multimedia ontologies have been identified [15].

Describing content using multimedia ontologies, however, also causes some problems as different ontologies are most often not aligned with each others. For example, different names are used for its elements which again hinders their integration. This problem of capturing the semantics of the various multimedia formats and of aligning them is currently tackled by the W3C Media Annotations Working Group[19].

Alternatively, we have recently proposed ramm.x[20] ("RDFa enabled multimedia metadata"), a proposal to integrate various descriptions based on different metadata standards [12]. The basic idea of ramm.x is to define a light-weight deployment description vocabulary allowing—deployed with RDFa [1]—a Semantic Web agent to determine the formalisation steps in order to process the native multimedia metadata format. This, in turn, allows a Semantic Web agent to determine what a multimedia object is about which enables

[19]A mapping table between numerous multimedia formats is available at `http://dev.w3.org/2008/video/mediaann/mediaont-1.0/mapping_table_090223_common.htm`

[20]`http://sw.joanneum.at/rammx/spec/`

him to set links at least semi-automatically.

5. METHODS FOR INTERLINKING MULTIMEDIA

Finally, the third iM principle states that specialized interlinking methods can be used to effectively create interlinking between multimedia resources at a fine-grained level. In this section, we critically review various methods and tools generally used for interlinking resources in the Web of Data.

5.1 Manual Methods

We have recently introduced *User Contributed Interlinking* (UCI) [8, 14], a manual interlinking methodology which relies on the end user as a source of qualitative information. UCI has been applied to enrich the Eurostat dataset [8]. A recent proposal, called CaMiCatzee [13] implements UCI for multimedia. CaMiCatzee allows people to semantically annotate picture on Flickr and to query for person's using their FOAF documents, URIs or person names.

Manual method for interlinking multimedia could be combined with incentives such as *Game Based Interlinking* (GBI), following the principles set forward by Louis van Ahn with his *games with a purpose*[21] [32]. One approach is to make the *interlinking* of resources *fun* and to hide the task of interlinking behind games. This is related to UCI but with the main difference that the user is not aware of him contributing links as his task is hidden behind a game.

GBI seems to be a promising direction for multimedia interlinking. The most interesting examples to build on are Ahn's ESP games in which users are asked to describe images, or Squigl[22] in which users are asked to trace objects in pictures. Another interesting approach is followed by OntoGame whose general aim is to find shared conceptualizations of a domain. OntoGame players are asked to describe images, audio or video files. Users are awarded if they describe content in the same way. Further exemplary games are OntoTube, Peekaboom, or ListenGame which hide the complexity of the annotation process of videos, images or audio files respectively, behind entertaining games. These approaches together with appropriate browsing interfaces for multimedia assets could be a promising starting point to let users draw meaningful relations between objects and their parts.

5.2 Collaborative Interlinking

Collaborative approach to interlinking of resources could be followed using Semantic Wikis. Semantic Wikis extend the principles of traditional Wikis such as collaboration, easy use, linking and versioning with means to type links and articles via semantic annotations [28]. Some of the systems support the annotation of multimedia objects including Semantic Wikis with dedicated multimedia support such as Ylvi [24], MultiMakna [21]. Most of these systems however treat a multimedia object as part of an article in which they appear. Thus, they do not allow specific annotations of it or treat them in the same manner like articles which can be only annotated globally. MultiMakna allows to assign annotations to temporal segments in videos through the use of an *appliesTo*-relation. While annotations may be constrained to its temporal context, to the best of our knowledge, links can only be established between articles and not segments.

Another Semantic Wiki with multimedia support is MetaVidWiki (MVW)[23] which enables community engagement with audio/visual media assets and associative temporal metadata. MVW extends the popular Semantic MediaWiki [17] with media specific features such as streaming, temporal metadata, and viewing and editing of video sequences. MVW supports the addressing and linking between temporal fragments. Segments of videos can be treated like "articles", referenced via URIs which support time intervals according to the temporalURI specification [22] and metadata about them can be exported in CMML [23].

5.3 Semi-Automatic Methods

Semi-automatic interlinking methods consist in combining multimedia-analysis techniques with human feedback. Analysis techniques can process the content itself or the context surrounding the content such as the user profile in order to suggest potential interlinking. The user would need to accept, reject, modify or ignore those suggestions. Inspiration for this type of approach can be found in the area of semi-automatic multimedia annotation.

Emergent Interlinking (EI) is another approach based on the principles of Emergent Semantics whose aim is to discover semantics through observing how multimedia information is used [5]. This can be essentially accomplished by putting multimedia resources in context-rich environments being able to monitor the user and his behavior. In these environments, two different types of context are present: (i) static or structural context, which is derived from the way how the content is placed in the environment (e.g. a Web page) and (ii) dynamic context, which is derived from the interactions of the user in the environment (e.g. his browsing behavior, which links he follows, or on which object he zooms). The assumption is that in appropriate environments, the browsing path of a user is semantically coherent and thus allows to derive links between objects which are semantically close to each other.

5.4 Automatic Methods

Finally, automatic interlinking of fragments of a multimedia resource can be achieved by purely analyzing its content. For example, in the case of such a musical audio content, the audio signal can be analyzed in order to derive a temporal segmentation. The resulting segments can be automatically linked to musically relevant concepts, e.g. keys, chords, beats or notes. The media fragment URI specification described in the Section 3.2 can then be used to relate these different segments to fragments of audio content.

The Music Ontology [25] provides a framework for the temporal annotation of audio signals. An audio file encodes an audio signal. This audio signal is linked to its timeline[24], i.e. its temporal backbone. The Event ontology[25] can then be used to classify particular regions of such timelines. For example, we define here two classifiers, capturing two chorus and two verses.

```
@prefix ps: <http://purl.org/ontology/pop-structure/>.
@prefix event: <http://purl.org/NET/c4dm/event.owl#>.
```

[21] http://www.gwap.com/

[22] http://www.gwap.com/gwap/gamesPreview/squigl/

[23] http://metavid.org/wiki/

[24] http://purl.org/NET/c4dm/timeline.owl

[25] http://purl.org/NET/c4dm/event.owl

```
ps:Chorus rdfs:subClassOf event:Event .
ps:Verse rdfs:subClassOf event:Event .
```

We can now use these two classifiers for annotating fragments of an audio signal. In the following, we describe the first chorus and the first verse in the the Beatles "Can't buy me love". We use two events, classifying two regions of the corresponding audio signal's timeline.

```
@prefix mo: <http://purl.org/ontology/mo/>.
@prefix event: <http://purl.org/NET/c4dm/event.owl#>.
@prefix tl: <http://purl.org/NET/c4dm/timeline.owl#>.
@prefix ps: <http://purl.org/ontology/pop-structure/>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix : <#>.

:signal owl:sameAs
  <http://zitgist.com/music/signal/eb20ee61-414f-4eee-
8dce-190db516a466>.

:signal mo:time [
     tl:duration "PT2M14S";
     tl:timeline :tl;
  ].
:chorus1 a ps:Chorus;
     rdfs:label "First chorus";
     event:time [
        tl:timeline :tl;
        tl:start "PT0S";
        tl:duration "PT9S";
     ].
:verse1 a ps:Verse;
     rdfs:label "First verse";
     event:time [
        tl:timeline :tl;
        tl:start "PT9S";
        tl:duration "PT33S";
     ].
```

An application of automatic interlinking of media fragments in the music domain is Henry[26] [26]. Henry aggregates music processing workflows available on the Web and applies them on audio signals to dynamically derive temporal segmentations and interlink these different segments with Web identifiers for music-related concepts.

For example, the following SPARQL query issued to Henry will dynamically generate annotations corresponding to changes of musical keys , described in the same way as in the above example. When processing the query, Henry accesses the audio file `http://dbtune.org/audio/Both-Axel.ogg` and applies a key extraction workflow to derive the temporal annotations. Each annotation is linked to a Web identifier corresponding to a musical key. Henry then binds the times at which the key change to the variable `?start`, and the Web identifiers for musical keys to the variable `?key`.

```
select ?start ?key where {
<http://dbtune.org/audio/Both-Axel.ogg> mo:encodes ?sig.
?sig mo:time ?time.
?time tl:timeline ?tl.
_:evt a af:KeyChange;
        event:time [tl:at ?start; tl:timeline ?tl] ;
        af:new_key ?key }
```

[26] `http://dbtune.org/henry`

5.5 Discussion

Currently, most interlinks between web datasets are generated entirely automatically, using heuristics to determine when two resources in two web datasets identify the same object [27]. However, this interlinking is done on the basis of existing RDF descriptions, and cannot directly be applied on multimedia objects. Such automated interlinking algorithms can only be applied if some RDF statements about the multimedia object (e.g. the performer involved in a particular recording) exist. These automated interlinking algorithms can be adapted to include an analysis step, extracting some information from the content of a multimedia object before deriving interlinks. The Henry software described in Section 5.4 is one example of automated interlinking performed by analyzing the multimedia object itself. However, the accuracy of the derived links is debatable, and heavily depends on the underlying analysis algorithm.

Another large source of interlinks on the current data Web comes from the work of large communities. For example, the Musicbrainz community created the links between artists in Musicbrainz and the corresponding artists in DBpedia. However, manual interlinking of media fragments is tedious, as lots of different annotations can be done. For example, a musical track could be described by its structural segments, by notes being played, by performers playing, by beats, etc. Therefore, large communities need to be involved in that process for it to be successful, perhaps using the collaborative annotation frameworks mentioned in Section 5.2 and the emergent interlinking mentioned in Section 5.3.

A possible solution to make such annotations scale would be to combine both approaches. Automated interlinking algorithms could post the resulting interlinks to a Semantic Wiki, where these links could be reviewed and modified by the community. Automated interlinking algorithm would then kickstart the interlinking process, and the resulting interlinks would gradually become more accurate.

6. RELATED WORK

In this section, we discuss related work and previous attempts for defining URI-based mechanisms for defining media fragments.

6.1 Hypermedia Links

Related work can be traced back to hypermedia research. An hypermedia document [9] refers to a collection of information units including information about synchronization between these units and about references between them. Temporal and spatial dimensions are typically included, whereas references can be made between parts in both dimensions. The issue of linking in hypermedia is discussed in [10, 19, 11]. Linking within multimedia presentations, within and among linear and non-linear multimedia presentations is discussed in [11]. [10] discusses links in time-based presentations and proposes a method to specify the context of links, i.e., what happens with the source or destination presentation of a link when it is traversed.

Hypermedia presentations consist of both static and dynamic media objects which are grouped together in so-called composite entities. Parts of these entities, identified via an-

chors that provide hooks for links, can be linked with each other and the behaviour of source and destination entities can be defined (e.g. shall the source video be paused or replaced). The ideas discussed above were implemented in the "Synchronized Internet Markup Langauge" (SMIL), a W3C recommendation that enables the integration of independent multimedia objects such as text, audio, graphics or videos into a synchronized multimedia presentation. Within this presentation, an author can specify the temporal coordination of the playback, the layout of the presentation and hyperlinks for single multimedia components. The latest version of SMIL provides a rich MetaInformation module which allows the description of all elements of a SMIL document using RDF. Media Fragments URI as defined in the previous section can be legally used in conjunction with SMIL documents.

6.2 Addressing Multimedia Fragments

Providing a standardized way to localize spatial and temporal sub-parts of any non-textual media content has been recognized as urgently needed to make video a first class citizen on the Web [30].

Previous attempts include non-URI based mechanisms. For images, one can use either MPEG-7 or SVG snippet code to define the bounding box coordinates of specific regions. Assuming a simple multimedia ontology available (designated with "mm:") the following listing 2 provides a semantic annotation of a region within an image: However,

```
<http://example.org/myRegion> foaf:depicts :Saphira
    ;
    rdf:type mm:ImageFragment ;
    mm:topX "40px" ;
    mm:topY "10px" ;
    mm:width "100px" ;
    mm:height "100px" ;
    mm:hasSource <http://example.org/children.jpg> .
```

Listing 2: Description of a spatial region depicting a person using a dedicated multimedia ontology.

the identification and the description of the region is intertwined and one needs to parse and understand the multimedia ontology in order to access the multimedia fragment.

URI-based mechanisms for addressing media fragments have also been proposed. MPEG-21 specifies a normative syntax to be used in URIs for addressing parts of any resource but whose media type is restricted to MPEG [20]. The temporalURI RFC [27] defines fragment of multimedia resources using the query parameter ('?') thus creating a new resource. YouTube launched a first facility[28] to annotate parts of videos spatio-temporally and to link to particular time points in videos. It uses the URI fragment ('#') but the whole resource is still sent to the user agent that just perform a seek in the media file. In contrast, the solution we are advocating allows to send only the bytes corresponding to media fragments while being still able to cache them.

7. CONCLUSION AND FUTURE WORK

In this paper, we have described the principles behind interlinking multimedia (iM), a pragmatic way to apply the linked data principles to fragments of multimedia resources. We have presented a URI-based mechanism for addressing parts of a multimedia resources following four dimensions (time, space, track and name). Furthermore, we have shown how these URIs can be used in the linked data context. We have pointed out that the use of content negotiation to serve alternatively media resource or description of these resources is debatable, though some implementation exists. We have stressed the importance of having mechanism to deploy multimedia metadata using light-weight approach such as ramm.x. Finally, we have presented various methods that can be used to actually generate interlinks between multimedia resources.

The presentation of these technologies left a number of challenging problems unsolved. It is unclear what content negotiation (in spirit if not technically) should do in this context. The semantics of media fragments is currently undefined. Retrieving partial content from a video resource given the definition of a temporal media fragment makes sense while it is hard to find use cases for its spatial counterpart. The Media Fragments WG is currently tackling these issues. We finally plan to work on a general framework for establishing the mapping between a media fragment and its RDF description in the general case.

[27] http://www.annodex.net/TR/URI_fragments.html
[28] http://youtube.com/watch?v=UxnopxbOdic

Acknowledgements

The authors would like to thank Richard Cyganiak, Yves Lafon, Ivan Herman, Silvia Pfeiffer, Tim Berners Lee and all the participants of the W3C Media Fragments Working Group for their willingness to discuss the linked data principles, the definition of media fragments and more generally their adequacy within the Web architecture.

The work of Michael Hausenblas has partly been supported by the European Commission under Grant No. 231335, FP7/ICT-2007.4.4 project "Intelligent metadata-driven processing and distribution of audiovisual media" (iMP).

8. REFERENCES

[1] B. Adida and M. Birbeck. RDFa Primer 1.0, W3C Working Group Note. http://www.w3.org/TR/xhtml-rdfa-primer/, 14 October 2008.

[2] 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'07)*, pages 30–43, Busan, Korea, 2007.

[3] T. Berners-Lee. Linked Data, Design Issues. http://www.w3.org/DesignIssues/LinkedData.html, 27 July 2006.

[4] T. Berners-Lee, R. Fielding, and L. Masinter. Uniform Resource Identifier (URI): Generic Syntax, RFC3986. IETF Network Working Group, 2005. http://www.ietf.org/rfc/rfc3986.txt.

[5] T. Bürger and M. Hausenblas. Interlinking Multimedia - Principles and Requirements. In *International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08)*, pages 31–36, Koblenz, Germany, 2008.

[6] N. Freed and N. Borenstein. Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types,

RFC2046. IETF Network Working Group, 1996. http://www.ietf.org/rfc/rfc2046.txt.

[7] F. Halasz and M. Schwartz. The Dexter Hypertext Reference Model. *Communications of the ACM*, 37(2):30–39, 1994.

[8] W. Halb, Y. Raimond, and M. Hausenblas. Building Linked Data For Both Humans and Machines. In *International Workshop on Linked Data on the Web (LDOW'08)*, Beijing, China, 2008.

[9] L. Hardman. *Modelling and Authoring Hypermedia Documents*. PhD thesis, CWI, Amsterdam, 1998.

[10] L. Hardman, D. C. A. Bulterman, and G. van Rossum. Links in hypermedia: the requirement for context. In *HYPERTEXT'93: 4^{th} ACM Conference on Hypertext and Hypermedia*, pages 183–191, Seattle, Washington, USA, 1993.

[11] L. Hardman, J. van Ossenbruggen, K. S. Mullender, L. Rutledge, and D. C. A. Bulterman. Do you have the time? Composition and linking in time-based hypermedia. In *HYPERTEXT'99: 10^{th} ACM Conference on Hypertext and Hypermedia*, pages 189–196, New York, NY, USA, 1999.

[12] M. Hausenblas, W. Bailer, T. Bürger, and R. Troncy. Deploying Multimedia Metadata on the Semantic Web (Poster). In *2^{nd} International Conference on Semantics And digital Media Technologies (SAMT'07)*, Genova, Italy, 2007.

[13] M. Hausenblas and W. Halb. Interlinking Multimedia Data. In *Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics'08)*, 2008. http://triplify.org/Challenge/Nominations?v=51f.

[14] M. Hausenblas, W. Halb, and Y. Raimond. Scripting User Contributed Interlinking. In *4^{th} Workshop on Scripting for the Semantic Web (SFSW'08)s*, Tenerife, Spain, 2008.

[15] M. Hausenblas eds. Multimedia Vocabularies on the Semantic Web, W3C Incubator Group Report. http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/, 24 July 2007.

[16] I. Jacobs and N. Walsh. Architecture of the World Wide Web, Volume One, W3C Recommendation. http://www.w3.org/TR/webarch/, 15 December 2004.

[17] M. Krötzsch, D. Vrandecic, M. Völkel, H. Haller, and R. Studer. Semantic Wikipedia. *Journal of Web Semantics*, 5:251–261, 2007.

[18] Y. Lafon and B. Boss. Describing and retrieving photos using RDF and HTTP, W3C Note. http://www.w3.org/TR/photo-rdf/, 19 April 2002.

[19] Z. Li, H. Davis, and W. Hall. Hypermedia Links and Information Retrieval. 14th Information Retrieval Colloqium, Lancaster University, http://eprints.ecs.soton.ac.uk/772/, 1992.

[20] MPEG-21. Part 17: Fragment Identification of MPEG Resources. Standard No. ISO/IEC 21000-17, 2006.

[21] L. Nixon and E. Simperl. Makna and MultiMakna: towards semantic and multimedia capability in wikis for the emerging web. In *Semantics'06*, Vienna, Austria, 2006.

[22] S. Pfeiffer, C. Parker, and A. Pang. Specifying time intervals in URI queries and fragments of time-based Web resources. http://www.annodex.net/TR/draft-pfeiffer-temporal-fragments-03.html, 2005.

[23] S. Pfeiffer, C. Parker, and A. Pang. The Continuous Media Markup Language (CMML), Version 2.1. http://www.annodex.net/TR/draft-pfeiffer-cmml-03.html, 2006.

[24] N. Popitsch, B. Schandl, A. Amiri, S. Leitich, and W. Jochum. Ylvi - Multimediaizing the Semantic Wiki. In *1^{st} International Workshop on Semantic Wikis (SemWiki'06)*, 2006.

[25] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The Music Ontology. In *8^{th} International Conference on Music Information Retrieval (ISMIR'07)*, pages 417–422, Vienna, Austria, 2007.

[26] Y. Raimond and M. Sandler. A Web of Musical Information. In *9^{th} International Conference on Music Information Retrieval (ISMIR'08)*, Philadelphia, USA, 2008.

[27] Y. Raimond, C. Sutton, and M. Sandler. Automatic Interlinking of Music Datasets on the Semantic Web. In *International Workshop on Linked Data on the Web (LDOW'08)*, Beijing, China, 2008.

[28] S. Schaffert, J. Baumeister, F. Bry, and M. Kiesel. Semantic Wikis. *IEEE Software*, 25(4):8–11, 2008.

[29] R. Schroeter and J. Hunter. Annotating Relationships between Multiple Mixed-media Digital Objects by Extending Annotea. In *4^{th} European Semantic Web Conference (ESWC'07)*, pages 533–548, Innsbruck, Austria, 2007.

[30] R. Troncy, L. Hardman, J. van Ossenbruggen, and M. Hausenblas. Identifying Spatial and Temporal Media Fragments on the Web. W3C Video on the Web Workshop, 2007.

[31] V. Tzouvaras eds. Multimedia Annotation Interoperability Framework, W3C Incubator Group Editor's Draft. http://www.w3.org/2005/Incubator/mmsem/XGR-interoperability/, 14 August 2007.

[32] L. von Ahn. Games with a Purpose. *IEEE Computer*, 39(6):92–94, 2006.