

Persistent Annotations Deserve New URIs

Abdulla Alasaadi
Old Dominion University
Computer Science Department
Norfolk, VA 23529, USA
aalasaad@cs.odu.edu

Michael L. Nelson
Old Dominion University
Computer Science Department
Norfolk, VA 23529, USA
mln@cs.odu.edu

ABSTRACT

Some digital libraries support annotations, but sharing these annotations with other systems or across the web is difficult because of the need of special applications to read and decode these annotations. Due to the frequent change of web resources, the annotation's meaning can change if the underlying resources change. This project concentrates on minting a new URI for every annotation and creating a persistent and independent archived version of all resources. Users should be able to select a segment of an image or a video to be part of the annotation. The media fragment URIs described in the Open Annotation Collaboration data model can be used, but in practice they have limits, and they face the lack of support by the browsers. So in this project the segments of images, and videos can be used in the annotations without using media fragment URIs.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Data Sharing, Web-based services

General Terms

Design, Reliability

Keywords

Web Archiving, Persistence, Annotation, URI

1. INTRODUCTION

Many digital libraries allow users to annotate the documents using their systems, but these annotations often cannot be shared by other digital libraries. An annotation can be described as a relation between different resources with different media types like text, image, audio, or video. Expressing this relation between resources is described in the Open Annotation Collaboration data model [6]. The system described in this paper is based on the OAC model.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'11, June 13–17, 2011, Ottawa, Ontario, Canada.

Copyright 2011 ACM 978-1-4503-0744-4/11/06 ...\$10.00.

A web resource can be annotated, for example, by an image, a YouTube video, or a Facebook post. Since browsers do not come with an annotating tool, a plug-in is needed to build the annotation body. To make the annotation readable by the public, it should be readable by their browsers without requiring the installation of a special tool. The resources that refer to different websites might get changed, moved or deleted; in this case the annotation will be meaningless.

We present a system that gives a unique URI for every annotation, and this URI can be shared and dereferenced to retrieve the annotation. The generated URI will represent an open and persistent annotation that can be viewed even if the annotated resources have been changed or deleted. Object Reuse and Exchange (ORE) Resource Maps [5] are used to keep track of resources in the Annotation as well as their archived representation.

2. RELATED WORK

Web annotations are an active research field. This include projects such as the W3C project Annotea [4] that uses an RDF based annotation schema for describing metadata. Annotea stores the annotation metadata either locally or in an annotation server, but it needs a special client that reads the annotation and to communicate with the annotation server. Amaya is a client implementation that reads and presents the annotation metadata created by Annotea [4]. LEMO [2] is an annotation system that has more features than Annotea such as annotating embedded contents within web pages and supporting multimedia fragment identification. Annotea and LEMO need special client application to read the annotation metadata, while this project concentrates on making the annotation readable without any help of any add-ons. The problem that may come with web-centric annotation is that the web pages change all the time, things either get updated, deleted or added at any time.

The question of how can we keep these annotations persistent over time has been discussed by Sanderson and Van de Sompel [7]. Their solution to keep the annotation persistent is Memento [9], an inter-archive approach for accessing archived representations of web resources that utilizes HTTP content negotiation in the dimension of "datetime". Since Memento by itself cannot guarantee the existence of an archived representation at the desired datetime, at the time of the creation of the annotation we explicitly push a representation to the WebCite archive.

3. MINTING NEW URIS

The system gives a new URI for every annotation, this URI has an archived version which is called a memento. A short URI will be generated also to refer to the annotation. Each annotation has different resources associated with it, for example a target image annotated by another image (Figure 1), each resource and memento has a different URI. The associated resources and the relations between them are summarized in Figure 3(b). If the annotator is interested in a particular part of a resource, for example part of image or video. According to OAC alpha-3 specification [6], the fragment can have its own URI, and it has an extension for the Media fragment URI [8]. For example, in Figure 1, a user wants to annotate the central part of the map of Africa, with another image that shows the reason behind the decreasing number of elephants in the selected area. The user needs to select a specific fragment of the image to be annotated. The fragment URI for a rectangle part in an image is:

`http://www.example.com/img.png#xywh=100,200,360,260`

But the media fragment specification [8] does not support the non-rectangular shapes, and it does not support all the media types available today. The lack of support by current browsers for media fragment URIs made it unpractical to be used in our system which we want it to be open and independent. That is why we use SVG technology that allows users to select particular parts of the images of any shape, plus it supports the video and audio tags which support fragmentation by default.

```
<video id="v1" xlink:href="video1.flv"
  type="video/x-flv" begin="2.2s" dur="20s"
  x="50" y="50" width="400" height="300">
```

4. INDEPENDENT AND PERSISTENT ANNOTATION

This project concentrates mainly in building an annotating system that mints new URIs for the annotations and provides mementos for all the resources in the annotation.

Creating the annotation requires an extra tool to do it online, this tool will annotate the resource then send the annotation metadata to the annotation server. The user will get a shortened URI that can be used later to retrieve the annotation from any browser. Users will be able to view the annotation by clicking on the shared short URI given by the annotation server.

Persistent annotation is a very important issue because the annotation has resources that refer to different websites on the internet, and these resources might get reallocated, deleted or updated for any reason.

5. ANNOTATION EXAMPLE

In Figure 1 the user annotated the central part of the map of Africa (ex:AFRICA) with another image (ex:Tusk) that shows the reason behind the increasing number of elephants killed in the specified area. SVG enabled the body (ex:Tusk) to annotate the selected part only from (ex:AFRICA). The system archives all the resources from the annotation and uses the archived mementos (M-x) to retrieve the annotation when the original resources are deleted.

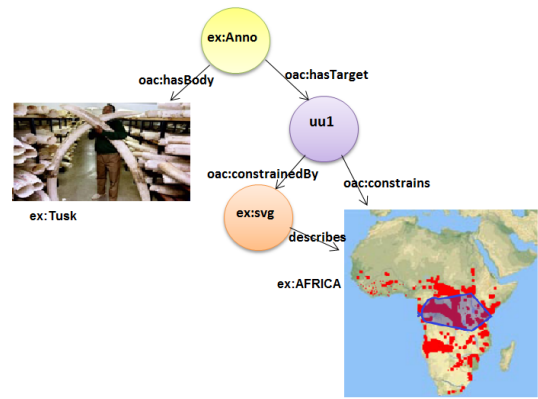


Figure 1: Annotation example.

Table 1: URIs used in the example.

| URI Table | |
|-----------|---|
| ex:Anno | http://annota.com/RDF/343.rdf |
| uu1 | urn:uuid:2AB92A55-6325-43D4-421D-D9DFB7ADA59E |
| ex:AFRICA | http://bioweb.uwlab.edu/bio203/s2007/shahrach/map.JPG |
| ex:Tusk | http://bit.ly/g55gvD |
| ex:svg | http://annota.com/SVGs/image344.svg |
| ex:ReM | http://cs.odu.edu/aalasaad/RDF/344.atom |
| M-1 | http://www.webcitation.org/5xp8yGp5U |
| M-2 | http://www.webcitation.org/5xp9A6qfn |
| M-3 | http://www.webcitation.org/5vp95UiJP |
| M-4 | http://www.webcitation.org/5xp94SX1P |

RDF syntax for Figure 1:

```
ex:Anno a oac:Annotation ,
  oac:hasBody ex:Tusk ,
  oac:hasTarget uu1 .

uu1 a oac:ConstrainedTarget ,
  oac:constrains ex:AFRICA ,
  oac:constrainedBy ex:svg .

ex:svg a oac:SvgConstraint ,
  dc:format "image/svg+xml" .

ex:Tusk a oac:Body .
```

5.1 URI Shortener

Since one of the main goals of this project is to give the annotations new URIs and enable users to share their annotations with the public by placing the annotation URI in their blogs, Facebook, or twitter accounts. Some social-websites have character limits for their messages. URI shorteners make the annotation's long URI shorter, and look nicer when users share it in their blogs. Bit.ly¹ is one of the most popular, stable and fast shortening service available and it has been used in this project.

¹<http://www.bit.ly>

5.2 RDF annotation and SVG representation

Each annotation is an RDF file that describes the relationships between the resources. Annotations usually applied to certain part of the target and sometimes annotated by specific part of another resource. For example, in Figure 2 the image of elephant tusks annotates the highlighted area and explains the reason behind the decreasing number of elephants in this particular area.

The idea of reading the annotation from any browser and sharing or embedding it in a blog has motivated us to make the annotation viewed as one object although it contains different resources with different media types. This led us to create a Scalable Vector Graph (SVG) representation for every annotation. SVG is an XML-based file format. SVG used in this project to constraint the annotated part of the targets. SVG also used to represent the RDF annotations because it is interactive, searchable, supports audio/video tags [3], supports Javascript, and hyperlinks. Google started indexing SVG content on August 31st, 2010. SVG is supported by all popular browsers FireFox, Chrome, Opera, Safari, and IE9.

5.3 The Video/Audio Tags

The Video/Audio tags are part of the SVG specification, but current browsers are facing problems in viewing the videos embedded within the SVG files. To solve this problem, when the system detects a video tag in the annotation, it renders the SVG file into a flash file in the server side, and sends it back to the browser as a regular flash movie which is supported by all browsers.

5.4 Web Archive WebCite

To keep the annotation persistent, the annotation, and all resources that are related to it, must be archived in one of the existing internet archives. Many popular archives have been investigated to choose the most suitable web archiving service that can help us. Wayback Machine², Archive-it³, Diigo⁴, and Webcite⁵ have been investigated and WebCite has been chosen because it has an API interface.

5.5 The Annotation Plugin

To enable users annotate the resources online, the project used the SVG-Edit plugin (Figure 2) and modified it to make it meet the project's goals. SVG-Edit [1] enables users to create SVG graphics. As part of this project, modifications were added to the SVG-Edit plugin to make it pick the target resource from an open web page, create the annotation and send it to the Annotation server.

6. IMPLEMENTATION

The two main services of this system are 1) the service that creates and shares the annotation, and 2) the service that retrieves the annotation.

The first web service will receive the annotation data from the browser, and then it sends the SVG data and the URIs of the related resources to the database. After that, the service will push all the related resources to the Archiving service to retrieve the URIs of the archived objects and save them in

²<http://www.archive.org/>

³<http://www.archive-it.org>

⁴<http://www.diigo.com>

⁵<http://www.WebCitation.org>

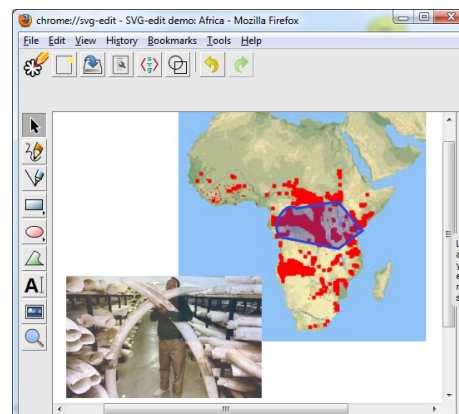


Figure 2: SVG Edit Plugin. A fragment of an image can be selected and annotated using SVG.

the database. A Resource Map [5] will be created for every annotation; this Resource Map contains aggregations of all the resources and mementos related to a given annotation. Then, the service will contact the shortening URI service and send the returned shortened URI to the browser.

Now the annotation has its own URI (ex:Anno) that returned by the system, and by dereferencing this URI the second web service will be called. The second web service will send back the annotation to the browser. First it connects the annotation server to find the annotation and then replace the URIs of target resources with the URIs of the archived copies of these resources. The SVG representation of the RDF annotation will be sent back to the browser. Links to the RDF annotation and other resources will be sent back. The web service returns a link HTTP header with every annotation, this link header explain the relation between the different annotations, resources and mementos.

6.1 Resource Map

The system creates a different memento for every resource, this leads to different types of resources or URIs associated with every annotation Figure 3(c).

The URIs and the relation between these URIs are represented with Resource Maps and Link headers.

A Resource Map is generated for every annotation, it lists all the associated resources to this annotation and the relationship between them. The following example shows how the aggregated resources are listed in the Resource Map:

```
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://bioweb.uwlax.edu/bio203/
s2007/shah_rach/map.JPG"
title="The original Target" type="image/jpg"/>
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://cs.odu.edu/~aalasaad/RDF/
344.rdf" title="The RDF annotation"
type="application/rdf+xml"/>
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://2.bp.blogspot.com/
_2ASHdLr0Gi0/TNmzuwr3kKI/AAAAAAAAAzI/Y91p52Huq3s/
s1600/elephant-tusks-460_1016757c-739042.jpg"
title="The Body" type="image/jpg"/>
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://www.webcitation.org/
```

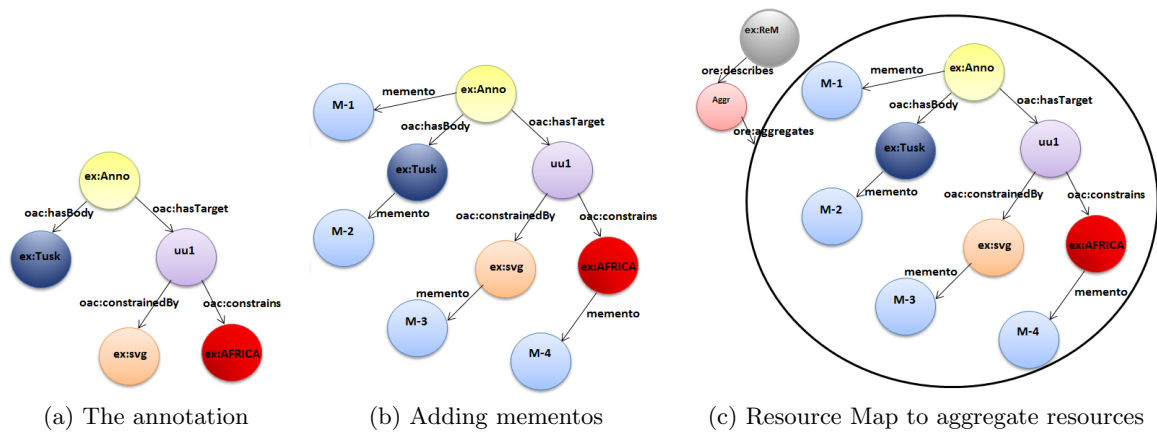


Figure 3: Resource map to track all resources and mementos of the annotation

```
5xqiA4E0f" title="Archived Target image"
type= "text/html"/>
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://www.webcitation.org/
5xqhw5tI9" title="Archived RDF Annotation"
type="text/html"/>
<link rel="http://www.openarchives.org/ore/terms/
aggregates" href="http://www.webcitation.org/
5xp9A6qfn" title="Archived Body image"
type= "text/html"/>
```

6.2 Response Header

When dereferencing the annotation URI (e.g., ex:Anno), an HTTP Link response header similar to the following is returned:

Link:

```
<ex:Anno>;rel="http://www.openannotation.org/ns/
hasBody"; <ex:Tusk>,
<ex:Anno>;rel="http://www.openannotation.org/ns/
hasTarget"; anchor="uu1",
<uu1>;rel="http://www.openannotation.org/ns/
constrains"; anchor="ex:AFRICA",
<uu1>;rel="http://www.openannotation.org/ns/
constrainedBy"; anchor="ex:svg",
<ex:ReM>;rel="resourcemap",
<M-1>;rel="memento"; datetime="Wed, 08 Dec 2010
09:00:00 GMT"; anchor="ex:Tusk",
<M-2>;rel="memento"; datetime="Wed, 08 Dec 2010
09:00:00 GMT"; anchor="ex:AFRICA"
```

7. CONCLUSION

The paper concentrates mainly on minting new URIs for the annotations, annotating media fragments without hacking fragment identifier specification and building an independent and persistent annotation system based on the OAC data model. Annotating the media fragments was made possible using the SVG and its media tags. Using the web archives to retrieve the annotated resources solved the issue of keeping the annotation persistent over time. Since the annotation is associated with many resources, the resource map is the perfect solution to keep track of all the related

resources and their copies and to represent the relations between them.

8. ACKNOWLEDGMENT

This research is supported in part by NSF IIS-1009392.

9. REFERENCES

- [1] SVG edit web site.
<http://code.google.com/p/svg-edit/>.
- [2] B. Haslhofer, W. Jochum, R. King, C. Sadilek, and K. Schellner. The LEMO annotation framework: weaving multimedia annotations with the web. volume 10, pages 15–32. Springer Berlin / Heidelberg, 2009. 10.1007/s00799-009-0050-8.
- [3] L. House and E. Pearlman. Introduction to SVG, 2003. <http://bit.ly/fscVcA>.
- [4] M. Kahan, J. Koivunen. Annotea: An open RDF infrastructure for shared web annotation. In *Computer Networks* 39, pages 589–608, 2002.
- [5] C. Lagoze, H. Van de Sompel, P. Johnston, M. L. Nelson, R. Sanderson, and S. Warner. Object Re-Use & Exchange: A Resource-Centric Approach. Technical Report arXiv:0804.2273, 2008.
- [6] R. Sanderson and H. Van de Sompel. Open annotation spec, alpha 3.
<http://www.openannotation.org/spec/alpha3/>.
- [7] R. Sanderson and H. Van de Sompel. Making web annotations persistent over time. In *Proceedings of the 10th annual joint conference on Digital libraries, JCDL '10*, pages 1–10, New York, NY, USA, 2010. ACM.
- [8] R. Troncy, E. Mannens, S. Pfeiffer, and D. Van Deursen. Media fragments URI.
<http://www.w3.org/TR/media-frags/>.
- [9] H. Van de Sompel, M. L. Nelson, R. Sanderson, L. Balakireva, S. Ainsworth, and H. Shankar. Memento: Time travel for the web. Technical Report arXiv:0911.1112v2, 2009.