Open Annotation Update

OAC Experiment Results
Ongoing Work of the W3C OA Community Group

Timothy W. Cole (t-cole3@illinois.edu)
Paolo Ciccarese (paolo.ciccarese@gmail.com)
Order of presentation

• Introduction: the basic open annotation data model
• Illustrative use cases: digital humanities & biomedicine
• The collaboration between AO & OAC
• The W3C Open Annotation Community Group
• Plans for Open Annotation 1.0 Release
• The OAC Experiments (2011-12)
• Spring 2013 Open Annotation Rollout Meetings

• Q & A
Open Annotation

“... towards a common, RDF-based specification for annotating digital resources. ”

• Specification defines a data model & essential vocabulary

• Aligns with WWW Architecture & Linked Open Data best practices

• Focuses on scholarly use cases & interoperability – the sharing of annotations across clients, servers, repositories and applications

• By our definition, annotations can be any media type and are treated as first class Web resources, i.e., subject to persistent reference and available themselves to be annotated
The basic Open Annotation data model

Res T – A resource (any media type) identified by a URI

http://somewhere.edu/MyPage.html
The basic Open Annotation data model

Res B – A resource (any media type) identified by a URI (existing or new)

- **Res B**
  - URN:uuid:1234567890

- **Res T**
  - http://somewhere.edu/MyPage.html
The basic Open Annotation data model

Res B
URN:uuid:1234567890

Says Something About

Res T
http://somewhere.edu/MyPage.html
The basic Open Annotation data model

http://somewhere.edu/MyAnnotation

Anno1 – A resource of class oa:Annotation identified by a URI

oa:hasBody

Res B

URN:uuid:1234567890

oa:hasTarget

Res T

http://somewhere.edu/MyPage.html
The basic Open Annotation data model

Anno1 gives identity to the relationship between Res B and Res T and instantiates the annotation as a First-Class Web Resource

http://somewhere.edu/MyAnnotation

Anno1 – A resource of class oa:Annotation identified by a URI

oa:hasBody

Res B

URN:uuid:1234567890

oa:hasTarget

Res T

http://somewhere.edu/MyPage.html
Key features (partial list)

Open Annotation specification provides support for

• expressing provenance, annotation authorship, ...
• expressing motivation for annotating
• annotating version / segment / representation of target resource
• multiple targets and multiple bodies
• annotation of targets in context
• semantic & data annotations

Includes semantics and basic, extensible vocabularies for describing target/body state and expressing selectors, annotation motivations, ...
Illustrative annotation use case 1

Collaborative Curation

Consider a digitized text that has been scanned, then OCR’d or keyboarded off-shore, then transformed into basic TEI XML.

You now want to allow annotations by colleagues and graduate students to identify needed corrections, add mark-up to expose additional structures, etc.
Illustrative annotation use case 1

Collaborative Curation

Consider a digitized text that has been scanned, then OCR’d or keyboarded off-shore, then transformed into basic TEI XML.

You now want to allow annotations by colleagues and graduate students to identify needed corrections, add mark-up to expose additional structures, etc.
You create an annotation identifying two character encoding errors.

**Annotator:** Tim Cole  
**Motivation:** EditRequired  
**Text:** Character encoding problem  
**Date Annotated:** 10 Dec 2012
Possible annotation graph for this illustration
(keeping in mind that 1.0 spec not yet released)
Annotation provenance, motivation, ...

- Record agents involved in annotation (both machine & human)
- Record when annotated / generated
- Mappings / points of contact with W3C Provenance Ontology

- Express annotation-level attributes such as motivation, rdf:type, style, ....
Inline annotation body

The content (body) of Annotation can be a stand-alone resource or can be embedded inline within the Annotation RDF description:

One way to do this is to use the W3C Representing Content in RDF Working Draft as illustrated here.

Also support other classes of annotation bodies.
Granularity and multiplicity of targets

Specification allows for specific targets/bodies and multiple targets/bodies

**oa:SpecificResource**
Allows targeting of versions and regions of a Web resource as it appears in a particular context

- **oa:hasSource**
- **oa:hasSelector**
- **oa:hasState**
- **oa:asIncludedIn**

Supports multiple classes of selectors

**Multiple Resources**
Semantics for differentiating meaning of multiple bodies and/or multiple targets – choice, list, set
Granularity and multiplicity of targets

Specification allows for specific targets/bodies and multiple targets/bodies

**oa:SpecificResource**
Allowing targeting of versions and regions of a Web resource as it appears in a particular context
- `oa:hasSource`
- `oa:hasSelector`
- `oa:hasState`
- `oa:asIncludedIn`

Supports multiple classes of selectors

**Multiple Resources**
Semantics for differentiating meaning of multiple bodies and/or multiple targets – choice, list, set
Paolo Ciccarese:

Illustration: biomedical annotations

The collaboration between AO & OAC

The W3C Open Annotation Community Group

Plans for Open Annotation 1.0 Release
OAC Experiments

The University of Queensland

• Collaborative Development of Scholarly Editions
  – This experiment explores use of annotations in the creation of electronic scholarly editions.
OAC Experiments

Stanford University & the Los Alamos National Laboratory

- Digitized Medieval Manuscripts (Shared Canvas)
  - This implementation of the OA data model leverages annotations to build a view of distributed resources through a single canvas.
OAC Experiments

University of Maryland

• **Annotation of Streaming Video**
  
  Focusing on issues of annotation target segmentation and mixed media, this experiment has built a series of browser-based tools for the annotation of videos.
University of Illinois at Urbana-Champaign

• Annotation of Digital Emblematica

• In collaboration with the Herzog August Bibliothek, Wolfenbüttel, researchers at Illinois are exploring the range of annotation types needed to support research needs of Emblem Scholars.
Cornell University

- **MapHub**
  - Cornell University has developed a Web portal for geo-referencing and annotating digitized, high-resolution historic maps.

Annotation

Annotations can easily be added by creating overlays on top of map images.

A window will pop up, prompting you to add an annotation for that region. While you are typing your annotation, Maphub will suggest possibly relevant Wikipedia tags for your annotation.
Open Annotation plugin for Fedora

Experimenters at Brown developed a content-agnostic web service that allows developers to create, query, retrieve, and serialize annotations using the Fedora Commons repository software to store annotations and their content.
OAC Experiments

Meertens Institute (KNAW)

• CODA (CATCHPlus Open Document Annotation)
  • The Netherlands-based CATCH and CATCHPlus programs added an OA compliant annotation repository, document annotation tool, workspace services, and associated software tools using two bodies of existing annotations that have been scanned.

CODA – CATCHPlus Open Document Annotation
OAC Experiments

University of Colorado at Denver

• Automated Annotation of Biomedical Text
  • Researchers at Colorado explored applications of Open Annotation in automated natural language processing and biological data analysis framework.

CRAFT: the Colorado Richly Annotated Full Text corpus

Quick Facts

- 67 full text articles
- >560,000 Tokens
- >21,000 Sentences
- ~100,000 concept annotations to 7 different biomedical ontologies/terminologies
  ◦ Chemical Entities of Biological Interest
  ◦ Cell Ontology
  ◦ Entrez Gene
  ◦ Gene Ontology (biological process, cellular component, and molecular function)
  ◦ NCBI Taxonomy
  ◦ Protein Ontology
  ◦ Sequence Ontology
- Penn Treebank markup for each sentence
- Multiple output formats available
- Integrated with UIMA

The Colorado Richly Annotated Full Text Corpus (CRAFT) is a manually annotated corpus consisting of 67 full-text biomedical journal articles. Each article is a member of the PubMed Central Open Access Subset.
OAC Experiments

New York University Libraries

- Annotation Middleware for Scholarly Publications
  - Investigators at NYU built a working prototype of a stand-alone annotation web service that works with content management systems, such as Drupal, that have robust commenting systems already built-in.

NYU Annotations

Demos

Annotate a region I  Annotate a region II  Using Annotated.js

NYU Digital Library Technology Services Group
Other projects using or influenced by Open Annotation

NISO / Internet Archive Bookmarks Initiative

Hypothes.is

the pund.it (Semlib, DM2E)

Digital Mappaemundi

Annotator / AnnotateIt (Open Knowledge Foundation)

Pleiades (Part of Pelagios & Perseus)
2013 Open Annotation rollout meetings

All day sessions featuring presentations and audience Q&A

Learn about:

• Open Annotation Data Model & Ontology
• Tools & Services that can help you build your implementation
• Goings on & future plans of the W3C Open Annotation Community Group
• How others are using Open Annotation

No fee to attend but registration will be required due to limited seating.

<table>
<thead>
<tr>
<th>Host</th>
<th>Date (Tentative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Coast (US)</td>
<td>9 April 2013</td>
</tr>
<tr>
<td>East Coast (US)</td>
<td>7 May 2013</td>
</tr>
<tr>
<td>Europe (UK)</td>
<td>11 June 2013</td>
</tr>
</tbody>
</table>
Acknowledgements

While the speakers are solely responsible for any errors or omissions in this presentation, we gratefully acknowledge the contributions of many others to this work:

• AO investigators: Paolo Ciccarese, Tim Clark
• OAC investigators: Tim Cole, Neil Fraistat, Anna Gerber, Jane Hunter, Rob Sanderson, Jim Smith, Herbert Van de Sompel
• Members of the W3C Open Annotation Community Group
• Other contributors and advisers too numerous to list here
• OAC Sponsor: the Andrew W. Mellon Foundation
• The generous support of participating institutions
• UIUC staff, in particular Jacob Jett, Suzi Harmon, Sally Eakin, ...
Questions?

• Links:
  – W3C Open Annotation Community Group
    http://www.w3.org/community/openannotation/
    http://lists.w3.org/Archives/Public/public-openannotation/
    http://www.w3.org/community/openannotation/wiki/Main_Page
  
  – Open Annotation Collaboration
    http://www.openannotation.org/
    http://www.openannotation.org/Partners.html
  
  – Annotation Ontology Initiative
    http://code.google.com/p/annotation-ontology/

• Contacts:
  Tim Cole (t-cole3@illinois.edu)
  Paolo Ciccarese (paolo.ciccarese@gmail.com)