

MARCH 2010 WORKSHOP REPORT
THE OPEN ANNOTATION COLLABORATION PHASE II:
DEMONSTRATION & REFINEMENT
Research grant provided by the Andrew W. Mellon Foundation

Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
501 E. Daniel Street (MC-493)
Champaign, IL 61820
Digital Library Research Laboratory
Grainger Engineering Library Information Center
1301 W. Springfield Ave. (MC-274)
Urbana, IL 61801
University of Illinois at Urbana-Champaign
April 2011
Project Website: <http://www.openannotation.org/>

Summary

On March 24 & 25, 2011, the Open Annotation Collaboration (OAC) project held a one and half day workshop at the Illini Center in downtown Chicago. The purpose of this workshop was to bring together scholars, librarians, and systems designers involved in ongoing digital content projects using or planning to implement annotation tools and services. Participants were asked to talk about their projects dealing with annotation and to provide feedback to the OAC on data modeling work of the Collaboration to date.

32 representatives from 28 scholarly projects, initiatives, and institutions, including, the Annotation Ontology project, ARTstor, CNI, Elsevier, Project MUSE, Proquest, and TEI (see Appendix 1 for a complete list of participants) were in attendance. Following an in-depth introduction to the OAC data model and ontology¹ for describing scholarly annotations of Web-accessible information resources, representatives from 12 of the participating projects presented their annotation use cases, tools, and projects. The presentations served to highlight a variety of scholarly annotation classes and target media types.

The Workshop found four key areas (Structured/Machine-readable Body, Constraint Precedence/Workflow, Constraint Types, and Annotation Types) that need to be addressed before the Beta release of the OAC Data Model. Discussions of these use cases and the OAC data model itself provided a number of action items for OAC to address and prepared attendees for OAC's forthcoming Request for Proposals (RFP). These action items include establishing various best practice guidelines, building services to facilitate the sharing of annotations (through an annotation validation and repository service), and providing guidance on how to use the data model, especially with regards to the four significant areas above.

Presentations, Additional Details

- **Introduction to the OAC Data Model and Ontology:** OAC investigators, Herbert Van de Sompel and Robert Sanderson presented a technical overview and an introduction to machine-readable annotations to the workshop attendees. These presentations presented the primary argument for the Collaboration's work:
 1. Annotations are a core ingredient to scholarship and
 2. Existing annotation solutions are neither inter-operable nor shareable

The OAC data model (Figure 1) was presented to the audience. Its features, such as constrained targets and inline content, were explored.

The discussion session following the introductory presentation session was dominated by discourse surrounding the data model's features. Of particular interest to Workshop attendees was the issue of applying constraints to the target of the annotation. Attendees wanted to determine whether or not the data model would be able to support multiple constraints on a

¹ <http://www.openannotation.org/spec/alpha3/>

target. There was a consensus that the data model should be amended to support the multiple constraints on a single target scenario.

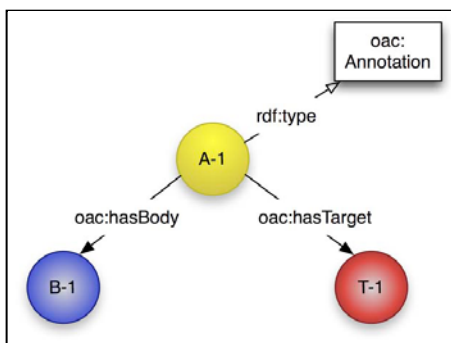


Figure 1: Open Annotation Data Model

- **Presentation Session 1 – Annotation of Marked-Up Text (including TEI):** Annotation projects from the University of Queensland and Brown University made presentations during the morning session of day 1.
 - OAC investigators Jane Hunter and Anna Gerber from the University of Queensland presented their project, “Annotation Supporting Collaborative Development of Scholarly Editions.” The overall goal of this project is to produce an annotation service for Aus-e-Lit, a project producing web-tools for the Australian AustLib web-portal. The annotation service is intended to aid in the creation of web-based scholarly editions of digitized literary works.
 - Andrew Ashton of Brown University next presented Brown’s project “Annotating Texts in the Brown Digital Repository.” The goal of the project is to produce an annotation service that can be used with Brown’s Fedora-based institutional repository (Brown Digital Repository). Targeting annotations to specific pieces of text within a document (text segments) is one of the most desirable bits of functionality that Brown hopes to implement in their annotation tool. They had tried using XPointer for this in the past but found that it was insufficient for TEI documents. They intend to investigate the use of constraints with the OAC model as they move forward.

Discussion generated by these two presentations focused on the issues of text segmentation, use of constraints on annotation targets, and the semantics of sub-classes of annotation. Annotation provenance and the issue of trust [of sources] were also brought up. No specific conclusions were made as a result of the discussion session, although the use of constraints with the OAC model was explored as a means to address text segmentation issues.

- **Presentation Session 2 – Annotating Scientific Literature:** Presentations from University of Colorado, Elsevier, and Harvard Medical School were made during the early afternoon session on day 1.
 - Karin Verspoor of the University of Colorado Denver presented Colorado’s “Annotating the Biomedical Literature through Text Mining” project. The project uses text mining in conjunction with existing ontologies and knowledge bases to identify key concepts within the bodies of text. The overall goal is to build an annotation tool that can leverage their existing text mining techniques and tools in order to automatically create annotations that capture both the associations between text and concepts and the associations between text and arbitrary sets of assertions.
 - Ron Daniel from Elsevier presented “Production Publishing Considerations for Annotating the Scientific Literature.” His presentation annotation need and use from the publisher’s perspective. There is a need to build services that can be used in conjunction with text mining to add concept annotations to documents. These annotations can then be exploited by system users to improve document retrieval.
 - Finally, Paolo Ciccarese from Harvard Medical School presented “Annotation Ontology and SWAN Annotation Tool.” The Annotation Ontology project is an effort very similar to the Open Annotation effort but with a hard science focus. The Ontology’s data model was presented. It was very specific to link topics to the annotations, a link that is made to the annotation body in the Open Annotation model.

The discussion session focused on the difficulties of mapping the Annotation Ontology data model Open Annotation data model. A question regarding copyright was also raised at this point. Would the inclusion of snippets of text within constraints of an annotation’s target but beyond the text specifically targeted by the annotation (i.e. including text from before and after the targeted text) constitute a breach of fair use? This question remained unresolved.

- **Presentation Session 3 – Annotation of Time-Based Media:** Maryland Institute for Technology in the Humanities (MITH) and Alexander Street Press made a joint presentation during the late afternoon session, which was followed by the final presentation of the day was made by Indiana University.
 - OAC investigators Jim Smith of MITH and Aaron Wood of Alexander Street Press presented “Subscription Streaming Video Content.” This joint project explores how to apply annotations to video content. Specifically MITH is building an annotation tool that can be embedded in Alexander Street Press’s website. Alexander Street Press hopes to build a service that will allow annotations to be imported to and exported from their webservice.

- Will Cowan of Indiana University presented Indiana's "Annotation for Ethnographic Video and Audio" project. The project created an "annotator's workbench" for Indiana's EVIA Digital Archive. The workbench was developed as part of a suite of tools for the developers and users of the EVIA archive. As part of the tool's development, Indiana explored ways to describe and represent video segments.

The final discussion session of Day 1 focused on the video segmentation issue. It was noted how difficult it would be to implement constraints for video segments, especially in cases where only portions of the video frame were desired as the target of an annotation. There was also some discussion of both segmentation issues in general and in how the RDF for the annotations would be made (i.e. automatically generated or user generated RDF graphs).

- **Presentation Session 4 – Annotation of manuscripts & other coordinated text & images:** The second day of the workshop opened with presentations from Stanford University, Princeton University, and the University of Waterloo (Canada). The focus of this session was the annotation of digitized manuscripts from both a text and image point of view.
 - Ben Albritton from Stanford and Open Annotation investigator Robert Sanderson presented their joint annotation project, "Shared Canvas: Interoperability for Digitized Medieval MSS Repositories" (<http://www.shared-canvas.org/>). The primary focus of the project is to produce tools that allow scholars to annotate digitized manuscripts across multiple repositories. The interoperability of the annotations produced is the key feature of the annotation tool.
 - Sylvia Stoyanova, from Princeton, next presented "Giacomo Leopardi's *Zibaldone*: a hypertext template for scholarly annotation". The *Zibaldone* presentation highlighted several lingering annotation issues. Specifically how to treat existing annotation structures within digitized content (which broadly applies to the text segmentation problems discussed during Day 1), how to handle internal references within series of annotations, and typing (i.e. classification of) annotations.
 - Christine McWebb of the University of Waterloo presented mock-ups of the "MARGOT Annotation Tool". The overall goal of the project was very similar to SharedCanvas, but focused more on some of the practical issues of annotation, such as, how to determine what section of an image is being annotated, how to export annotations, how to import external materials into annotations, etc.

This session produced a lot of discussion revolving around what kind of data was on the backend of the html manuscripts, TEI mark-up text or images of physical pages. Clarification on the identity of an annotation's target (e.g. whether it was a segment of text or an entire document) was needed. Lingering issues regarding the provenance of annotations were also raised. The question of whether or not an annotation tool should or would need to include an

authentication layer so that annotations could be linked with their authors was raised but remained unresolved.

- **Presentation Session 5 – Annotation of maps & geographic texts:** The final presentation session of the Workshop addressed the annotation of maps and geographic texts. Cornell University and Drew University annotation projects were both presented to the Workshop attendees at this time.
 - Bernhard Haslhofer presented Cornell’s map annotation project, “Historic Map Annotations with YUMA”. The project began as a set of open-source annotation tools for Europeana. It is evolving into an annotation framework for multi-media objects on the web. While the project was originally using the LEMO data model, they hope to switch to the Open Annotation data model in the near future.
 - Shannon Bradshaw of Drew University made the final presentation of the workshop, “An OAC-Compliant Toolbox”. Using funds supplied by NEH and Mellon, Drew is developing a holistic toolbox that is designed to address all of the scholarly primitives, which include annotation, citation, and discovery, among others. The goal is not just to make annotations that are interoperable with one another but to make annotations that also better support other parts of the scholarly research cycle.

The final discussion session focused on the technical issues of annotating maps, such as how to annotate events specific to a map, how maps change over time, and when maps are used not as maps but rather as illustrations within another work. Identity of an annotation’s target continued, with the question of whether or not the data model supported annotations as the target of other annotations (the spec does address this issue).

Key Workshop Outcomes & Collaboration Action Items

As is clear from the summary above, the OAC Workshop generated a great deal of useful feedback and helpful guidance for the Collaboration going forward. Summarized here is the distillation and synthesis of this feedback and guidance. Further steps and future OAC activities, such as development of the Beta spec for the data model and the RFP to recruit additional projects to collaborate with are also summarized. Additionally, community building activities resulting from the workshop are also noted.

Further work is needed to provide guidance on producing structured/machine-readable annotation bodies. Issues of constraint precedence and workflow within the data model need to be addressed. More guidance is needed from the Collaboration on how to address inheritance, provenance, constraint typing, annotation typing, and target segmentation when applying the data model and ontology. From the discussion sessions it is evident that four priority issues & services for Open Annotation to address have been identified:

- text segmentation

- inheritance
- provenance
- sharing/interoperability

To facilitate the Open Annotation community in addressing these and other issues identified during the presentation sessions, the Collaboration will find ways to make it easier to share tools and methods for sharing annotations. Open Annotation will also establish best practice guidelines for creating structured bodies, dealing with multiple, aggregate, or discontinuous targets, typing annotations, typing constraints, and addressing architectural issues. The Collaboration will look at services to facilitate sharing and building a repository of shareable annotations conformant to the Open Annotation data model.

Open Annotation is moving forward with the development of a Beta spec for the data model in order to refine areas of the data model that were identified as needing more work at the workshop. Those areas for refinement are:

1. Structured/Machine-readable Body
2. Constraint Precedence/Workflow
3. Constraint Types
4. Annotation Types

As work proceeds on the Beta spec, the Collaboration is also moving forward with its RFP (Request For Proposal). This RFP will allow Open Annotation to partner with four additional projects or institutions that have existing annotation tools or interesting scholarly annotation cases. These funded collaborations will provide opportunities to demonstrate the implementation of Open Annotation's data model and ontology.

Proposals for work that addresses some of the practicalities of implementing Open Annotation's data model and ontology that were noted in the workshop session summaries above will be recognized as of priority. These practicalities include:

- Viable methods for identifying arbitrary segments of text as annotation targets or bodies
- Use cases involving annotations that propagate across formats and/or FRBR entities
- Demonstrations of annotation portability, e.g., device and resolution independence
- Implementations that incorporate provenance, fixity, and/or target context into annotations
- Strategies for mapping from existing native formats for annotation to the Open Annotation data model
- Use cases demonstrating the relevance of Open Annotation to Social Reading, etc.
- Logic for dealing with alternate constraints and supporting graceful fall back

- Experiments examining the practical utility of annotation and/or constraint typing
- Use cases involving complex, structured bodies
- Experiments focusing on interoperability across disciplines

The final outcome of the workshop was the identification of three “Birds of a Feather” groupings which met at the end of the Workshop. These groups were focused on three specific areas of annotation:

- Annotation for Education
- Annotation of Video
- Annotation of Editions of Texts/Manuscripts

These community efforts will foster further discussion of open annotation practices, guidelines, and issues.

Appendix A: Workshop Attendee List

28 projects / initiatives / institutions were represented at the workshop (32 individuals), in addition to members of the core OAC team. Shown below are participants listed by project or institution; for a list ordered by name see: [\[2\]](#)

- Alfalab
 - ❖ Alexander Witteveen (Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Sciences)
- Annotation Ontology
 - ❖ Paolo Ciccarese (Harvard Medical School)
- ARTstor
 - ❖ William Ying
- AustLit
 - ❖ Anna Gerber (The University of Queensland)
 - ❖ Jane Hunter (OAC co-PI) (The University of Queensland)
- BioNLP
 - ❖ Karin Verspoor (University of Colorado Denver)
- Canadian Writing Research Collaboratory
 - ❖ Susan Brown
 - ❖ James Chartrand
- CATCHPlus Project
 - ❖ Hennie Brugman (Meertens Institute)
- Center for Informatics Research in Science and Scholarship, University of Illinois
 - ❖ Allen Renear
- CLARIN
 - ❖ Menzo Windhouwer (The Language Archive, Max Plank Institute for Psycholinguistics)
- Coalition for Networked Information
 - ❖ Cliff Lynch
- The Collaborative Annotation Tool
 - ❖ Philip Desenne (Academic Technology Group, Harvard University)
- Elsevier
 - ❖ Ron Daniel
- EVIADA
 - ❖ William Cowan (Institute for Digital Arts & Humanities, Indiana University)
- Giacomo Leopardi's Zibaldone
 - ❖ Silvia Stoyanova (Princeton University)
- The Long Civil Rights Movement Project
 - ❖ Jenn Riley (University of North Carolina at Chapel Hill)
- MARGOT Annotation Tool
 - ❖ Christine McWebb (University of Waterloo)
 - ❖ Ian Davis (University of Waterloo)
- MediaThread
 - ❖ Jonah Bossewitch (Columbia Center for New Media Teaching and Learning)
 - ❖ Schuyler Duveen (Columbia Center for New Media Teaching and Learning)
- Northwestern University Library
 - ❖ William Parod

- The Nyangwe Diary of David Livingstone
 - ❖ Heather Ball (ASA Institute of Business & Computer Technology)
 - ❖ Adrian Wisnicki (Birkbeck University of London)
- Old Dominion University
 - ❖ Michael Nelson
- Open-Source Toolbox for Annotation
 - ❖ Shannon Bradshaw (Drew University)
- The Pico Project
 - ❖ Andrew Ashton (Center for Digital Scholarship, Brown University)
 - ❖ Michael Park (Center for Digital Scholarship, Brown University)
- Project MUSE
 - ❖ Brian Harrington (John Hopkins University Press)
- ProQuest
 - ❖ John Burns
- Shared Canvas
 - ❖ Robert Sanderson (Research Library, Los Alamos National Laboratory)
 - ❖ Benjamin Albritton (Stanford University Libraries)
- Subscription Streaming Video Content
 - ❖ James Smith (Maryland Institute for Technology in the Humanities)
 - ❖ Aaron Wood (Alexander Street Press)
- Text Encoding Initiative
 - ❖ Peter Gorman (University of Wisconsin Digital Collections Center)
- YUMA
 - ❖ Bernhard Haslhofer (University of Vienna)

Also attending were Principal Investigators:

- ❖ Tim Cole (Center for Informatics Research in Science and Scholarship, University of Illinois)
- ❖ Anna Gerber (The University of Queensland)
- ❖ Jane Hunter (The University of Queensland)
- ❖ Robert Sanderson (Research Library, Los Alamos National Laboratory)
- ❖ James Smith (Maryland Institute for Technology in the Humanities)
- ❖ Herbert Van de Sompel (Research Library, Los Alamos National Laboratory)

and CIRSS's staff members:

- ❖ Jacob Jett
- ❖ Kevin Trainor

Appendix B: Workshop Agenda

Thursday - 24 March 2011

- 8:30 AM [Welcome & Open Annotation Project Overview](#) (Tim Cole)
- 9:00 AM Introduction to the OAC Shareable Annotation Data Model & Ontology
(Herbert Van de Sompel & Rob Sanderson)
- [Technical Overview](#)
 - [Machine Readable Annotations](#)
- 10:30 AM Break
- 10:45 AM USE CASES -- Annotation of marked-up text, including TEI
Chair & discussion facilitator: Allen Renear
[Annotation Supporting Collaborative Development of Scholarly Editions](#)
(Jane Hunter & Anna Gerber)
[Annotating Texts in the Brown Digital Repository](#) (Andy Ashton)
Discussion - 40 minutes
- 12:15 PM Lunch (provided) -- continued discussion in small groups
- 1:15 PM USE CASES -- Annotating Scientific Literature
Chair & discussion facilitator: Herbert Van de Sompel
[Annotating the Biomedical Literature through Text Mining](#) (Karin Verspoor)
[Production Publishing Considerations for Annotating the Scientific Literature](#) (Ron Daniel)
[Annotation Ontology and SWAN Annotation Tool](#) (Paolo Ciccarese)
Discussion - 25 minutes
- 2:45 PM Break
- 3:00 PM USE CASES -- Annotation of time-based media
Chair & discussion facilitator: Jane Hunter
[Subscription Streaming Video Content](#) (Jim Smith & Aaron Wood)
[Annotation for Ethnographic Video and Audio](#) (Will Cowan & Alan Burdette)
Discussion - 40 minutes
- 4:30 PM Perspectives on the Project and Day 1
- 5:00 PM Adjourn Day 1

Friday - 25 March 2011

- 8:30 AM USE CASES -- Annotation of manuscripts & other coordinated text & images
Chair & discussion facilitator: Christine McWebb
Shared Canvas: Interoperability for Digitized Medieval MSS Repositories [Part 1](#) & [Part 2](#)
(Ben Albritton & Ron Sanderson)
[Giacomo Leopardi's Zibaldone: a hypertext template for scholarly annotation](#)
[\[Zibaldone Sample\]](#) (Silvia Stoyanova)
[MARGOT Annotation Tool](#) (Christine McWebb)
Discussion - 25 minutes
- 10:00 AM BOF Break
- 10:45 AM USE CASES -- Annotation of maps & geographic texts
Chair & discussion facilitator: Rob Sanderson
[Historic Map Annotations with YUMA](#) (Bernhard Haslhofer)
[An OAC-Compliant Toolbox](#) (Shannon Bradshaw)
Discussion - 40 minutes
- 12:00 PM [Next Steps](#) (Tim Cole, et al)
- 12:15 PM Action Item Summary (Tim Cole, et al)
- 12:30 PM Adjourn Workshop
- 1 - 3 PM Room Available