CNI Spring 2009 Task Force Meeting Minneapolis, MN

**Project Briefing:** 

# **Open Annotation Collaboration**

Timothy W. Cole, University of Illinois at Urbana-Champaign

John Burns, JSTOR Daniel Cohen, George Mason University Neil Fraistat, University of Maryland Jane Hunter, University of Queensland Robert Sanderson, University of Liverpool Herbert Van de Sompel, Los Alamos National Laboratory

http://www.OpenAnnotation.org/



# annotation as a 'scholarly primitive'

- A definition of 'scholarly primitives' (from Unsworth): Basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation.
- Annotation is a scholarly primitive that supports, linking, relating, elaborating, comparing, referencing, illustrating, teaching, and other activities fundamental to scholarship in humanities & sciences.



## sharing annotations

#### "Shared annotation is,

for all scholarly intents and purposes, impossible on the Web."

-- John Unsworth, at *Humanities Computing* King's College, London, May 2000.

http://www3.isrl.uiuc.edu/~unsworth/Kings.5-00/primitives.html



# OAC goals & objectives

#### Overarching project goals:

- Facilitate emergence of a Web and Resource-centric interoperable annotation environment that allows leveraging annotations across boundaries of annotation clients, annotation servers, and content.
- Demonstrate benefits of interoperable annotation environment in settings characterized by a variety of annotation client/server environments, content collections, and scholarly use cases.
- Seed widespread adoption by deploying applications conformant with the interoperable annotation environment across ubiquitous and specialized services, tools, and content used by scholars



## definitions

- A basic definition in plain English: Annotations associate extra information with digital resources without modifying the original.
- Our initial working Web & Resource-centric definition: An Annotation is an autonomous Resource (Web Architecture) that groups one or more Source Resources with one or more Target Resources by means of an Annotation Relationship



# preliminary annotation model



6 April 2009 - CNI Spring Task Force

## starting assumptions

- Resources of any media type can be Sources or Targets; Annotations are conceptual Resources that have no media type.
- The author of the Annotation, of the Source and of the Target(s) may be different.
- Source may exist before annotation relationship is created or may be created at the same time as the relationship.
- Annotations, as Resources with URIs, can be annotated with further Annotations.



# key properties of annotations

- Interoperable annotation model will support at least the expression of the following annotation properties:
  - Annotation Creator(s) -- including ability to express creator's identity and associated identity verification authority
  - Date/Time Annotation was created
  - Annotation usage permissions
  - Segment metadata for Sources & Targets
  - Locations of cached Sources & Targets (optional, but useful given ephemeral nature of Web Resources in practice)

# possible annotation profile for ORE



6 April 2009 - CNI Spring Task Force

- Shared model of annotation will be based on:
  - Web Architecture
  - Linked Data -- best practices for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web
  - W3C Standards
- Targets, Sources and Annotations must all have URIs
- To achieve appropriate granularity, URIs of Targets, Sources and Annotations may be qualified with additional metadata
- Additional infrastructure will be required:
  - Distributed servers to store annotations and optionally cache targets
  - Annotation aggregators or overlay services
  - Clients supporting & exploiting Annotation Read interfaces as defined by spec

#### preliminary view of possible architecture



I

6 April 2009 - CNI Spring Task Force

Open Annotation Collaboration http://www.OpenAnnotation.org/

# OAC project phase 1

- Tasks:
  - Create data model & interoperability specification
  - Integrate AXE Libraries (MITH) into Zotero
  - Perform systematic scholarly annotation analysis
- Tentative timeline:
  - 14 months: 16 May 2009 through 15 July 2010

#### Application for funding pending with the Andrew W. Mellon Foundation -- Research in Information Technology program

# OAC project team

- Herbert Van de Sompel (co-Pl) -- Research Library, LANL
- Timothy Cole (co-PI), Thomas Habing, Carole Palmer, Allen Renear -- University Library & GSLIS-CIRSS, U of Illinois at UC
- Neil Fraistat (co-PI), Douglas Reside -- MITH, U of Maryland
- Jane Hunter (co-PI), Anna Gerber -- eResearch Lab, School of ITEE, U of Queensland
- Daniel Cohen, Sean Takats -- CHNM, George Mason University
- Robert Sanderson -- Dept of Computer Science, U of Liverpool
- John Burns, Clare Llewellyn -- JSTOR



# specific objectives OAC phase 1

- Analysis of existing annotation models, systems, and architectures and scholarly practices related to annotation to inform creation of data model supporting interoperability that is widely adaptable & rooted in scholarly practice.
- Alpha release of annotation interoperability specification.
- Integration of Ajax XML Encoder (AXE) code libraries into Zotero; in addition to advancing both AXE and Zotero, this work will inform modeling & specification development.
- First-draft reference implementation, demonstrating proof-ofconcept & laying foundation follow-on phases of work



#### phase 1 work plan - thread 1 annotation interoperability specification

- Specification will detail:
  - Annotation data model
  - Annotation Information Architecture, defining requirements for & relationships between software components
  - Read Annotation Interfaces
- We will solicit community involvement (OAI model)
- Model & specification development will be informed by current practices, tools, and system architectures.
- A minimal reference implementation will be created as proof-of-concept

### phase 1 work plan - thread 2 implementation & integration experience

- Integrate AXE libraries into Zotero
  - AXE developed at MITH with NEH Digital Humanities grant
  - Will allow Zotero to annotate more than just Web pages
- AXE features:
  - Intuitive Web-based interface
  - Facilitates metadata creation directly by users
  - Makes linking digital artifacts more efficient and accurate
- Collaboration between MITH & CHNM; will be informed by analysis of experiences to date with Zotero's existing annotation client



#### File Edit View History Bookmarks Tools Help

#### Russian Folk Religious Imagination

Executive Summary

List of Participants Contact Details

Sample Materials



This video clip illustrates annotation of video segments with textual commentary. Click on video to view the annotation (since annotation is longer than text box in which it appears, click in text box and use the down arrow to see end of text). Links to complete commentaries are below the video window.

To experiment with the AXE user interface that allows for textual annotation of this RFRI video, <u>click here</u>.



Project Home



cemetery (for a discussion of this tradition in this region, see Y.

#### phase 1 work plan - thread 3 application analysis & model evaluation

#### • This thread will:

- Perform environmental scan & analysis of current annotation applications and repository service contexts
- Identify key repository dependencies and interactions with annotation clients & services
- Examine in detail scholarly-oriented tools (e.g.: Co-Annotea, Pliny) & content sets (e.g.: JSTOR, MONK, AustLit, Flickr)
- Examine in detail issues to do with annotating annotations and annotating compound resources & multiple targets
- Participate with Thread 1 in defining & developing model
- Evaluate model in terms of current & anticipated requirements of scholars carrying out annotation



# state-of-the-art

- Interest in digital resource annotation dates from earliest experiments with hypertext & hypermedia
  - See various: C. Marshall, M. Agosti, J. Kahan, M. Koivunen, ...
- Recent & ongoing projects (incomplete list):
  - W3C Annotea (RDF-based)
  - SANE (Scholarly ANnotation Exchange)
  - OATS (Open Annotation and Tagging System)
  - Other approaches embedded in clients & DL systems



# state-of-the-art (continued)

- Issues with work to date:
  - Often not fully compliant with Web Architecture
  - Conflates annotation Source (content) with assertion of annotation relationship -- e.g., not separately addressable
  - Focuses on particular media type and/or only limits media type of annotation Source
  - Does not support interoperability across annotations done using different clients or targeting content from elsewhere



## distinguishing features of this project

- Begins with commitment to interoperability
- Provides option to ensure persistence of Targets & Sources (i.e., as well as Annotations)
- Allows for single Annotation to have multiple Targets
- Provides ability to annotate annotations, relationships and graphs of relationships
- New approaches for research into problem of merging annotations across different representations of a resource



# challenges

- Determining and exploiting incentives for scholarly annotation
- Addressing tension between general and flexible annotation framework versus discipline-specific research challenges
- Supporting synchronization of multiple annotation body vocabularies
- Support for user management: Authentication, Authorization/Access, Trust



# preliminary use cases

Critical that interoperability data model accommodate diverse, real-world scholarly use cases. For example:

- Citation linking into performance video (e.g., dance)
- Use streaming audio to annotate streaming video
- Share annotations for processing while protecting sourcing of annotation Sources & Targets
- Extract & transform social tags from Flickr Commons into shareable annotations; demonstrate some measure of sort & merge across repositories with overlapping content



# preliminary use cases (continued)

- Annotate mixed media objects in ways that define new compound objects, which in turn can be annotated
- Demonstrate support for *comparing* with Annotations that exploit multi-part Sources and/or Targets
- netchaining: show how chains across primary & secondary sources can be created, shared, & extended
- Share annotations that employ GIS techniques to document conservation state/history of extremely high-resolution digital images
- Additional???



#### use-case: multi-part annotation target

Inspired by Pasanek & Sculley (2008).

A scholar studying the evolution of metaphor uses Pliny to annotate and make a comment about the metaphor in John Locke's *An Essay Concerning Human Understanding* comparing the human mind to blank white paper.



## multi-part annotation target (part 2)



Another scholar annotates a similar metaphor from *Leviathan* by Thomas Hobbes

## multi-part annotation targets (concluded)

But one of the important points here is that the Hobbes metaphor predates Locke's use of the similar metaphor. To express this notion, the preferred solution is to create an annotation that targets both the Hobbes and the Locke passages together – i.e., an annotation with a multi-part Annotation Target.

In Pliny this concept is supported by a construct called an *annotation container.* Our shared interoperable data model must provide a means to express and share these kinds of annotations.



I

## use-case: netchaining

# Inspired by Suzana Sukovic (2008).

The initial link: A scholar asserts that a JSTOR journal article effectively "annotates" a passage from a digital copy of Stowe's *The Two Altars* she finds in the MONK Project data store.





# netchaining (part 2)



Subsequently another scholar, using different annotation tools, creates an annotation having a multi-part annotation Target that encompasses this same JSTOR article as well as a series of Lectures on African Colonization and Kindred Subjects delivered to the Ohio State House of **Representatives in 1849 (text of** which has now been digitized by OCA). The annotation Source contrasts Stowe's view of race and slavery to those held by fellow-Ohioan and contemporary David Christy (i.e., he who delivered the lectures).

# netchaining (concluded)

A third scholar using yet different tools and repositories illustrates an evolution in Stowe's thinking about race & slavery by annotating passages in *The Two Altars* in concert with mentions of pre-war slavery conditions in a cookbook written by Stowe and her sister in the 1870's that has been digitized as part of Harvard's *Women Working* collection

Being able to share these annotations across tools & repositories, makes it possible to expose chains of links between digital resources.



**Open Annotation Collaboration** 

http://www.OpenAnnotation.org/



# motivating questions

- Can we describe a broadly useful model of annotation not tied to repository design or type of content being annotated?
- Using this model, can we enable new opportunities for digitally-based scholarship built around annotation interoperability?
- What are the defining scholarly use cases and can we embed our model in existing applications to demonstrate benefits for these use cases?
- Are there additional benefits to be had by treating annotations as first-class Web Resources?



# more information

- Website: <a href="http://www.OpenAnnotation.org/">http://www.OpenAnnotation.org/</a>
- Contact: Tim Cole (<u>t-cole3@illinois.edu</u>)

Need community input on use cases, tool exemplars, ...

What's your favorite annotation use case?

