THE OPEN ANNOTATION COLLABORATION PHASE II: DEMONSTRATION & REFINEMENT

1. Executive Summary

Annotating is a practice core to scholarship. Marginalia, glosses, and other recognized forms of annotation left behind by authors and by previous readers of a text bring an added dimension to how we understand that text, sometimes serendipitously, often through intent. The experience of reading can be altered by the presence of annotations. A student reading a well-annotated text will come away with a deeper understanding and appreciation of that text than he or she had before. Annotation of other media or in other forms can be just as powerful. In their survey of scholarly annotation practice past and present, Maristella Agosti, et al. note that annotation is often in its own right a "vehicle for carrying and transmitting ideas and knowledge to other people" and has been used historically as a "professorial tool."[1] The potential formats and modes of annotation should expand in a digital context. In his 2000 presentation at King's College London, John Unsworth identified annotating as a "scholarly primitive," but noted with concern the lack of progress towards applications that effectively supported the sharing of annotations on the Web.[2] This lack persists today. A number of issues and perceived obstacles have slowed development and adoption of tools for scholarly annotation of digital resources. Some of these have to do with the difficulties of mapping traditional practices of annotation into a digital context, but many others reflect a failure to systematically exploit opportunities afforded by the semantic Web and linked data initiatives, an unevenness in tools and limited options for annotating across clients and content repositories. Missing is implementation of a shared data model and ontology on which to base interoperable annotation tools and services.

The OAC seeks to remove these obstacles by engendering the emergence of a Web and Resource-centric interoperable annotation environment and by demonstrating the potential of this environment to support robust scholarly annotation in a multiplicity of digital humanities and digital library domains and across boundaries of clients, tools, and repositories. OAC Phase I focused on the development of a foundational data model and ontology for interoperable scholarly annotation, leveraging the collective experience of OAC members with Web architecture, model-theoretic semantics, Linked Data Initiative principles, and existing and emerging Web standards and best practices. OAC data model development has proceeded to the point where it is now feasible to broaden testing and, in collaboration with scholars and repository managers, to vet and refine the OAC data model against real-world scholarly use cases involving significant repositories of digital content. Work carried out during OAC Phase I has made clear that adoption of a shared annotation data model suitable to support humanities scholarship must happen initially domain by domain, and must begin happening soon to avoid the de facto entrenchment of a tower of Babel, project-by-project approach to scholarly annotation on the Web. To further our understanding of annotation interoperability, to create the Web and Resource-centric interoperable annotation environment sought, and to encourage in support of these goals the adoption and evolution of the open and robust data model for scholarly annotation on the Web developed during OAC Phase I, we will need to focus in OAC Phase II on directly engaging humanities scholars and involving existing collections of digital content having well-defined communities of scholars interested in annotating such content.

---

[1] Agosti, M., Bonfiglio-Dosio, G., & Ferro, N. (2007). A historical and contemporary study on annotations to derive key features for systems design. *International Journal on Digital Libraries*, 8(1), 1–19. DOI: 10.1007/s00799-007-0010-0. See also the more detailed discussion regarding the context for and importance of digital resource scholarly annotation in our proposal for OAC Phase I.

[2] Unsworth, building on a theme from Aristotle, uses the term scholarly primitives "to refer to some basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation." Unsworth, John. 2000. "Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?" Presented at Humanities Computing: formal methods, experimental practice, King's College, London, May 13, 2000. Available: http://www3.isrl.uiuc.edu/~unsworth/Kings.5-00/primitives.html (viewed 3 August 2010).

Accordingly, the continuing members of the OAC, i.e., the universities of Illinois, Maryland, and Queensland and the Los Alamos National Laboratory, request a total of $673,942. to undertake OAC Phase II in order to execute concrete, collaborative small-scale demonstration projects and experimentation featuring challenging scholarly use cases, existing digital content in situ, and the direct involvement of scholars and curators of scholarly digital resources. Through this demonstration and experimentation we will vet and refine the nascent OAC data model and ontology, demonstrate its utility in enabling a resource centric environment for interoperable scholarly annotation on the Web, and proactively encourage early adoption of this environment across a range of different scholarly domains. We propose to implement a total of 8 complementary annotation demonstration experiments, each involving active projects external to OAC that already have existing digital content and well defined scholarly audiences. Use cases will span a range of media formats, will address a complementary range of modeling requirements (e.g., a range of different approaches to segment description), and will include communities with prototype annotation tools for scholarly annotation or at least well-recognized and understood annotation tool needs. We propose to configure each demonstration to be a joint undertaking involving an OAC member institution team working in close collaboration with the repository staff and scholars proposing the use case.

Drawing on this experimental work, we also in OAC Phase II will advance the maturity, completeness and expressiveness of the OAC data model itself, the guide to this data model, and the ancillary documentation and best practice recommendations intended to inform implementers and facilitate consistent use of the data model by a broad range of scholarly communities. The latter category of documents will include recommendations for serializing Annotation Resource Maps in RDF, RDFa, and other formats. An outcome of Phase II will be a version 1.0, suitable-for-production OAC data model and specifications of similar maturity and completeness to 1.0 specifications developed for earlier projects such as OAI-PMH and OAI-ORE. Though not an ultimate end in itself, an interoperable data model of and ontology for scholarly annotation on the Web is a necessary pre-requisite to enable our larger goals.

Finally, we will observe closely the annotation demonstration projects in progress and require that all 8 projects collaborating with us during OAC Phase II report back in depth on their experiences along the way in applying the OAC data model to meet their domain-specific needs. This documentation in combination with what we learn by observation from each demonstration about needs and practices with regard to scholarly annotation of digital resources will be synthesized as part of our own report out. By documenting and synthesizing the results, both positive and negative, from the 8 demonstration experiment projects, we will be able to advance the state of our understanding of scholarly annotation needs in the Web environment and provide an adaptable and well illustrated model for others interested in implementing the OAC data model and ontology. There is broad-based interest in this problem space, and it is our expectation (depending in part on the success of our 8 demonstration experiments) that many other researchers and funders within multiple other domains will benefit from and be interested in further developing the OAC Web and Resource-centric interoperable annotation environment that we will engender and demonstrate through the work proposed for OAC Phase II.

PROPOSAL

THE OPEN ANNOTATION COLLABORATION PHASE II: DEMONSTRATION & REFINEMENT

Principal Investigators:

Timothy W. Cole, Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign

Jane Hunter, eResearch Laboratory, School of Information Technology & Electrical Engineering, University of Queensland

Doug Reside, Assoc. Dir., Maryland Institute for the Humanities
University of Maryland

Herbert Van de Sompel, Research Library
Los Alamos National Laboratory

Co-Principal Investigators:

Anna Gerber, eResearch Laboratory, School of Information Technology & Electrical Engineering, University of Queensland

Carole Palmer, Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign

Robert Sanderson, Research Library
Los Alamos National Laboratory

Allen Renear, Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign

Aaron Wood
Alexander Street Press

Proposal Table of Contents

2. Background & Context

2.1 OAC Overarching Goals

The overarching goals of the Open Annotation Collaboration (OAC) are:

i. To facilitate the emergence of a Web and Resource-centric interoperable annotation environment that allows leveraging annotations across the boundaries of annotation clients, annotation servers, and content collections. To this end, a shared data model of annotation is required, along with a shared vocabulary and guidelines for serialization and implementation.

ii. To demonstrate the interoperable annotation environment enabled by the OAC data model and interoperability guidelines in settings characterized by a variety of annotation client/server environments, content collections, and scholarly use cases.

A shared and well vetted OAC Data Model and ontology for shareable annotations is key to enabling annotation tool interoperability, the persistence of scholarly annotations of Web resources over time, and the sharing of annotations beyond the boundaries of individual solutions and/or content collections. This data model in combination with articulated principles for annotation interoperability, ancillary guides describing the data model and vocabulary, serialization options and interoperable annotation best practices, and a diverse range of exemplar demonstrations showing OAC in action, will allow for the emergence of value-added cross-repository and eventually cross-domain annotation services. It will also facilitate the implementation of advanced end-user annotation services targeted at humanities scholars and capable of operating across a broad range of both scholarly and general collections. Further, vetted data model and ancillary documentation will enable customization of annotation services for specific scholarly communities, without reducing interoperability. The proposed work also will enable more robust machine-to-machine interactions and automated analysis, aggregation and the opportunity for reasoning over distributed annotations and annotated resources.

2.2 OAC Phase I Outcomes

OAC Phase I (June 2009 to date) has set the stage and provided context for the annotation use case demonstration experiments proposed for OAC Phase II. Major accomplishments of phase I to date and anticipated before the end of the Phase I project in December 2010:

- Publication of the OAC Guiding Principles for Annotation Interoperability.[3]
- Publication of a guide to alpha version of the OAC data model for describing annotations involving Web resources, including a summary of OAC-specific vocabulary & semantics.[4]
- (Building on existing OAI-ORE documentation) Publication of a white paper outlining some of the ways that existing protocols such as RSS and OAI-PMH, etc. might be exploited to disseminate Annotation Resource Maps.
- Publication of notes from the May 2010 OAC Technical Review meeting in Albuquerque, NM, including a listing of hypothetical use cases posed & presented during the meeting.
- Release of enhanced AXE libraries supporting annotation of targets in more formats.[5]

[3] http://www.openannotation.org/documents/OAC_GuidingPrinciples_20091106.pdf

[4] http://annotation.lanl.gov/

[5] http://mith.umd.edu/research/?project=19

- Incorporation of AXE libraries into Zotero, resulting in an enhanced Zotero annotation facility
- Presentations on initial results and outcomes at CNI Membership meetings, the 2010 Joint Conference on Digital Libraries, and Digital Humanities 2010.
- Results (in various formats) of other early experimentation with Pliny, Danno, and other tools, demonstrating potential annotation use cases and exploring ways in which existing annotation and note-taking tools can be aligned with the nascent OAC data model.[6]

Key for OAC Phase II planning among the findings from Phase I:

- To support scholarly annotation the OAC data model must provide ways to express complex context and various kinds of nuanced relationships between Web resources, segments of resources, versions of resources, representations of resources, and interrelated and sometimes dependent resources at varying levels of granularity, including abstract conceptual information resources encompassing multiple digital instances.
- Data model implementation issues vary significantly by domain, usage scenarios, and formats of annotation body and annotation target resources.
- Digital humanities scholars and content providers supporting DH scholars are keenly interested in annotation data models and tools that support (for their domain) complete, persistent, and accurate descriptions and representations of annotations and notes made across multiple content repositories, even if only for personal annotation and note taking.
- At this point in time within the DH scholarly community, a high priority is assigned to developing domain-focused tools that can instantiate and exploit scholarly annotations in a manner compatible with Web architecture, standards & best practices; relatively less priority is assigned to work on large-scale aggregation and data mining of annotations across domains. To the extent that annotation sharing is of immediate interest, it is small-scale sharing within a domain or a collaborative research group that is given priority.

2.3 The Collaboration

The OAC is an ongoing collaboration of humanists, librarians, and information technologists from 4 institutions with long and successful track records of innovation and recognized expertise in the domains of digital collection creation, use, and interoperability.

Current OAC participants at the Maryland Institute for the Humanities (University of Maryland), a founding member of *centerNet*,[7] and the Center for Informatics Research in Science and Scholarship (Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign) bring expertise and an extensive, diverse body of research and practical experience with regard to understanding the needs and interests of humanities scholars working with digitized information resources. The librarians at the University of Illinois Library at Urbana-Champaign and the information technologists at the Research Library at Los Alamos National Laboratory and the eResearch Laboratory in the University of Queensland's School of Information Technology & Electrical Engineering who are members of the Collaboration have been leaders on prior digital library interoperability initiatives including in particular the Open Archives Initiative (i.e., the OAI-Protocol for Metadata Harvesting and OAI-Object Reuse & Exchange protocols), highly visible and consequential work funded in the main by the Andrew W. Mellon Foundation. Collectively and individually the members of the Collaboration have been funded for prior research in the realms of digital humanities and digital libraries not only by the

[6] Various, linked from: http://www.openannotation.org/phaseI.html

[7] An international network of Digital Humanities centers (http://digitalhumanities.org/centernet/)

Mellon Foundation, but also by the National Science Foundation, the Institute for Museum and Library Services, the National Endowment for the Humanities, and others. During OAC Phase I, the recognized stature of OAC participants helped to lend immediate credibility to the initial data modeling work and attract the attention of a broad segment of the digital humanities and digital library communities. This interest has been sustained through the work accomplished during OAC Phase I. Papers integrating some of the early outcomes of OAC Phase I were selected for and presented at both Digital Humanities (DH) 2010 and the Joint Conference on Digital Libraries (JCDL) 2010, and were well received at both venues.

An objective of OAC Phase II is to broaden the scope and reach of the Collaboration by involving more leading members of the digital humanities and digital library communities in hands on work with the OAC data model for shareable annotations. As detailed subsequently in this proposal, we propose to begin by collaborating on 4 annotation demonstration experiments with humanities scholars directly involved in the use of structured texts (including scholarly editions) and digital emblematica, a commercial publisher of scholarly streaming media resources, and curators of digitized medieval manuscript content. Through a planned spring 2011 workshop and 4 subsequent projects funded as part of this grant via an open Request for Proposals, we intend to broaden involvement in the work of the Collaboration even further over the course of OAC Phase II. In this way it is our intent and expectation to demonstrate the utility of a Web and Resource-centric interoperable annotation environment and simultaneously grow a community of practice focused on enhancing the capacity for and functionality of digital resource scholarly annotations. Collectively OAC is well positioned by virtue of Phase I results and participants' records of accomplishment to reach these goals.

3. Rationale: The Need for Experimentation, Demonstration & Refinement

The work of Phase I was informed by limited, proof-of-concept experimentation undertaken by OAC participants, including: the successful integration of the MITH-AXE annotation software libraries into Zotero; an experiment migrating notes taken in Pliny into Danno, a Web-based annotation service; experimental evaluations demonstrating how the OAC data model can be used in conjunction with the Memento Framework to retrieve all annotations that involve a given archived version of a web resource or to reconstruct today a previously created annotation such that the correct archived version is displayed for all resources involved in the annotation. Phase I work also was informed by a body of both theoretical and speculative research examining the history of paper-based annotation and looking at how the traditional practices of scholarly annotation might translate and change in a context dominated by digital resources.[8] We also benefitted from prior annotation research specific to a range of domains and contexts,[9] as well as prior and contemporary looks at the potential for annotation interoperability and sharing more generally.[10] However, while useful in creating the OAC data model and ontology for shareable annotation, ultimately the literature and limited scope, proof-of-concept experimental work done solely within the Collaboration is insufficient to vet and promulgate a practical, implementable data model and ontology for annotation interoperability and annotation sharing.

Interoperability data models, protocols and best practices are necessarily meant to be useful and useable across a broad range of clients and contexts. Unavoidably there is a tendency to see and talk about works such as the OAC data model in abstract terms. The translation of such work into the concrete context of a particular scholar's work is not always immediate or intuitive. Some efforts to establish shared data models and digital library interoperability specifications have suffered because they have been undertaken in isolation or in the context of a single project. Both OAI-PMH and OAI-ORE demonstrated that successful efforts to achieve useful interoperability and engender environments supportive of broad resource sharing and interactivity must be informed by a breadth of practice and experimentation.

Demonstration, use, and experimentation are the means by which nascent interoperability models and standards are most effectively and efficiently vetted, improved, and promulgated. The OAC data model is part of a generalizable framework and vocabulary by which scholarly annotations of Web resources and components of Web resources can be described, shared, and instantiated as new Web resources in their own right. And while the development of the OAC data model for shareable annotations during OAC Phase I did benefit from some practical experimentation and development work carried out at OAC member institutions, and while the introduction of the alpha version of the OAC data model for shareable annotations has attracted attention and been generally well received by both the digital humanities community and the digital library community, several pragmatic questions about possible model implementation

[8] E.g.: Marshall, Catherine C. "Annotation: from paper books to the digital library" DL '97: Proceedings of the 2nd ACM international conference on Digital libraries, pp 131-140, 1997. Also Maristella Agosti paper cited above.

[9] E.g.: J.Hunter, “Collaborative Semantic Tagging and Annotation Systems ”, chapter for *Annual Review of Information Science and Technology*, American Society for Information Science & Technology, Volume 43, 2009

[10] E.g.: W3C Annotea project: http://www.w3.org/2001/Annotea/. More recent: The LEMO annotation framework: weaving multimedia annotations with the web. Bernhard Haslhofer, Wolfgang Jochum, King Ross, Sadilek Christian, Schellner Karin. *International Journal on Digital Libraries* 10(1): 15-32. DOI: 10.1007/s00799-009-0050-8

issues, concerns and trade-offs that surfaced during OAC Phase 1 remain to be investigated and further evaluated. Empirical results will be required to fully resolve these issues. The following incomplete list illustrates the kinds of issues we intend to address through the demonstration experiments proposed for OAC phase 2.

- On the Web URIs are essential, but correspondence between what a URI identifies and the intellectual content a scholar wishes to annotate may be poor (e.g., a URI exists for digitized medieval manuscript page image, but not for the illuminated initial on the page). What are use-case specific trade-offs between minting a new URI to identify illuminated initial and providing a segment description for illuminated initial? Do these trade-offs generalize?

- Some methods (e.g., SVG overlays) of segment description are obvious and accommodated well by the OAC data model and W3C Media Fragments Working Group recommendations. What other, domain-specific segment description options exist, and are they accommodated as well? Can existing scholarly citation schemes, e.g., Stephanus numbers, be leveraged?

- The OAC data model assumes a need to distinguish between the Annotation and the Annotation Body (content). Are there compelling use cases that demonstrate this need?

- Is there a need for an annotation "type" vocabulary that defines sub-classes of the *oac:Annotation* class, the semantics of such sub-classes and their special attributes?

- The OAC data model accommodates annotations involving multiple targets as well as annotations targeting aggregations of resources (e.g., OAI-ORE aggregations). What factors should implementers consider when deciding which approach to use?

- The OAC data model provides mechanisms for representation-specific annotations -- i.e., annotations that target not a whole Web resource, but only a specific disseminated representation of a Web resource. Are there real-world use cases that require this facility? If so, are the means provided in the model sufficiently rigorous and unambiguous in practice?

- Experiments from Phase I showed that the OAC data model can be used with the Memento Framework to insure annotation timelessness; how pervasive is the need for this facility?

- OAC Phase I work suggests issues of anchor vs. citation are important. A literary scholar annotates a novel passage while viewing a specific digital instance of the novel. While the annotation target can be anchored in that specific digital instance, often the intent is to annotate the passage in multiple digital instances of the novel, possibly spanning editions of the work. What are best practices for using the OAC data model to express this intent?

- An annotation may both target resources and reference other resources. When should the second relation, i.e., to the referenced work, be treated as a reference versus as a 2nd annotation target? What are trade-offs? How should implementers recognize the distinction?

- Biodiversity annotation use cases suggest that there are times when a user may want to annotate a resource in context, e.g., a scholar may want to say something about an image, but only about the image as it is embedded in a specific Web page. How important a use case is this in DH domains, and if important, is it well enough handled by our data model.

Broader interaction with digital humanities scholars and established and emerging repositories of digital scholarly resources are required to resolve these and similar issues. Only once proven through practical experience can we anticipate significant uptake of the OAC model.

4. Project Description

4.1 Project Objectives

Building on the successes of OAC Phase I, we propose in OAC Phase II to:

i. Demonstrate practical applicability, utility and benefits of implementing the OAC data model to support scholarly annotation for 8 demonstration experiments spanning multiple domains and involving content in a range of formats; each demonstration will result in a report detailing process, benefits realized, and implementation successes, difficulties and failures.

ii. Resolve open issues and residual uncertainties to do with the current alpha version of the OAC data model and ontology; this will allow us to refine and improve the OAC data model and ontology, moving from an alpha 3 release (expected before the end of 2010) to a version 1.0 production release by June 2012;

iii. Observe demonstration experiments in progress and synthesize results reported, yielding:
   - o augmented documentation helpful to assist would-be implementers of the OAC data model and ontology; and
   - o an overview report summarizing common themes and core requirements of scholarly annotation on the Web and highlighting ways in which scholarly annotation of digital resources differs across domains and models of content.

iv. Proactively encourage the immediate adoption and adaptation of the OAC data model for shareable annotation as the foundation for an environment supportive of scholarly annotation tools and services that span boundaries of annotation clients, annotation servers, and content collections.

4.2 Identifying Collaborating Projects

Essential to the success of OAC Phase II will be the recruitment and selection of collaborating projects involving scholars, librarians and content providers interested in using the OAC data model to address domain-relevant, challenging, user-driven scenarios exploiting high quality scholarly digital resources. To identify appropriate projects requires an articulation of the key attributes that characterize those projects well positioned for collaboration. Collaborating projects for OAC Phase II will be diverse as regards annotation use cases and kinds and formats of annotatable content, but all collaborating projects will share certain key attributes:

- Each will be associated with an existing collection or collections of digital scholarly information resources and will have in place collection services supporting persistent URI-mediated, Web-access to content (whether open or under IP or password-based controls);

- Each will have an annotation tool or service need that is both germane to current discipline interests and has inherent, challenging data modeling requirements not fully addressed by existing, implemented models of scholarly annotation;

- Each will be targeted at and have access to a recognized and discipline-specific audience(s) of scholarly users, i.e., as sources for use cases and requirements for annotation tools, and in some cases as potential experimenters;

- Each will present a use case(s), important in the domain of the project, that will allow us to demonstrate benefits of a Web and Resource-centric approach to annotation data modeling;

- Each will be positioned to demonstrate an annotation use case scenario(s) that potentially could span multiple repositories and/or service platforms;

- Each will be led by investigators with established track records in DH and/or DL research.

We will prefer projects in disciplines or domains that have a nascent annotation-related client or service model for annotation tool interfaces under development or well described (it is not the objective of the OAC to design domain-specific end-user interfaces for annotation).

In identifying our group of collaborating partners, we will take into account the complementary nature of the projects, collections, and annotation use cases. We will look to balance domains explored, content formats to be exploited (both for annotation body and annotation target), the complexity of domain's digital annotation needs and how well these needs have been surfaced and refined, the range and diversity of data modeling challenges to be explored, and the potential for sustainability and/or continued development once demonstration / experimental phase collaboration with the OAC has been completed.

Through contacts established during OAC Phase I -- some we initiated, some initiated by discipline-specific projects interested in annotation -- we are well acquainted with much of the ongoing work on scholarly annotation of digital resources. We have an established rapport with several discipline-specific projects. We have made headway in identifying specific issues and concerns raised by these projects and have some existing momentum in our examination of these issues and concerns. As described above work carried out during OAC Phase I also has allowed us to identify a number of issues relating to scholarly annotation and the modeling of scholarly annotation that absolutely need to be researched further if we are to achieve overall OAC goals. This argues that the OAC should recruit projects we know will allow us to address these most critical issues; the only certainty that the issues we identified during OAC Phase I as the most critical will be addressed during OAC Phase II is to take this approach for selecting at least some of the annotation demonstration experiments to be carried out.

On the other hand we clearly are not familiar with all current projects involving digital annotation of scholarly resources on the Web. Moreover, it is reasonable to assume that we will encounter and want to surface additional issues and concerns to do with this problem space that warrant further research and investigation. In addition to cementing initial strategic contacts established during OAC Phase I, we want to make new contacts. These factors argue for an open call to the community soliciting fresh proposals for annotation demonstration experiments.

Accordingly, we propose a two-tiered approach to identifying the annotation demonstration experiments to be undertaken during OAC Phase II.

Tier I: collaborators for four demonstration experiments will be recruited (from projects already known to us from Phase I) based on suitability to advance our understanding of critical research questions surfaced during OAC Phase I. These first four annotation demonstration experiments

(additional details on each given below and in section 8 of this proposal) will begin 3 January 2011 (preserving current momentum) and will be completed during the fall of 2011. For efficiency, economy, and to maximize confidence of early, positive results, two of these first 4 demonstration experiments will involve projects in which OAC institutions already participate.

Tier II: four demonstration experiments will be selected from responses received to a Request for Proposals (RFP) that will be issued in March 2011. These annotation demonstration experiments will then be undertaken beginning in summer 2011, with completion no later than May 2012. To encourage quality proposals, to provide with more context and details about OAC Phase I, and to provide another avenue for promulgating the OAC data model and ontology, release of the RFP will be coordinated with an in-depth workshop on the implementation of the OAC data model; however, **workshop attendance is not a pre-requisite for submitting a response to the RFP**. Additional details of the RFP, the workshop, and the process that will be used to select workshop attendees and from among responses to the RFP are given below.

4.3 Jan. through Dec. 2011: OAC Demonstration Experiments 1 - 4

OAC Phase II annotation demonstration experiments 1 through 4 will allow us to explore a range of the scholarly annotation and data model research questions identified as among the most critical from OAC Phase I (see Section 3 above). Collectively the demonstration experiments selected span 3 complementary formats of target content: structured text, still image, and streaming media (video), as well as offering some opportunity to experiment with structured text and other non-trivial formats of annotation body. This insures we will demonstrate annotations involving a range of annotation body and target formats. Each of these demonstration experiments are associated with projects having the essential attributes enumerated above. Each builds on ongoing relationships and/or work begun during OAC Phase I. Each is described in further detail in section 8 of this proposal. Briefly presented here are the main factors that led to the selection of each of these projects:

Annotation Supporting Collaborative Development of Scholarly Editions

- *Critical issues addressed / demonstrated*: A key requirement in developing a scholarly edition is the ability to create, share and reply to scholarly commentary attached to variations (textual differences) between versions of a particular literary work. This demonstration experiment will seek to address this need and will explore: annotations involving multiple targets and/or ORE aggregations as targets; the use of segment descriptions in the context of structured literary texts; distinctions between citation and annotation target (anchor), including multi-target annotations addressing fragments spanning multiple manifestations of FRBR expressions and works; and the potential for robust retrieval by target of annotations described using the OAC data model.

- *Collaborating project & collection suitability*: As the primary resource for scholars of Australian literature, AustLit is well established with a large scale and an international audience of scholarly users. AustLit provides access to a significant body of structured texts using a metadata model based on the IFLA FRBR report.[11] Working with advisers such as Paul Eggert (University of New South Wales, AustLit Advisory Board), and as the lead/host organization for both AustLit and Aus-e-Lit, the University of Queensland is well-positioned to undertake this annotation demonstration experiment. The proposed work builds on

[11] http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records

already cultivated relationships between OAC, AustLit and Aus-e-Lit, maintaining momentum from Phase 1 and giving high confidence of a useful outcome.

Annotation of Digitized Medieval Manuscripts

- *Critical issues addressed / demonstrated*: This demonstration experiment will feature annotation of both structured text and still-images and will explore: a range of annotation granularities and discipline-relevant segment description options, leveraging in particular the TILE (Text and Image-Linking Environment tools project, University of Maryland) and T-PEN (Transcription-Paleographic and Editing Notation Tool project, St. Louis University) initiatives as exemplars; annotations having structured text bodies; multi-target annotations having targets of different formats (i.e., text and image) and/or spanning multiple repositories (e.g., Stanford's "return-and-reuse" scenarios); and annotation of annotations including some that may change over time (e.g., transcriptions).

- *Collaborating project & collection suitability*: Contingent on both projects being funded, this annotation demonstration experiment will be carried out jointly by the OAC and a proposed new Stanford University Libraries project, *Defining a Modular and Interoperating Environment for Collections of Digitized Medieval Manuscripts, Tools, and Users*. As host of the *Parker Library on the Web* and given its other work in this arena, the Stanford University Libraries have access to exemplary content, are well established in this domain, and are familiar with end-user interests and needs. Contingent on a favorable response to their proposal, work on their new project is scheduled to begin in October 2010. During Phase I, OAC has looked to the digitized medieval manuscript domain for use case examples; OAC members are involved in leading the TILE initiative and participated in the May 2010, Stanford-hosted *DMSTech* meeting,[12] a broadly inclusive meeting to examine the issues and technology needs of medieval manuscript repositories. A joint OAC-Stanford experiment would build on these existing relationships and on established momentum in both domains.

Annotation of Subscription Streaming Video Content

- *Critical issues addressed / demonstrated*: This demonstration experiment will feature annotation of streaming video and will explore the nuances, limitations and persistence of streaming segment descriptions as complement or alternative to existing service to mint URIs for full-frame video clips. Also explored: multi-target annotations having targets of different formats and with different access restrictions (e.g., public domain text and subscription performance video); relationship between synchronization of content in different formats and anchor vs. citation distinction issues; and techniques for importing / exporting annotations in custom client serializations from and to OAC-compliant serializations.

- *Collaborator & collection suitability*: Alexander Street Press, founded in 2000 to bring together the skills of traditional publishing, librarianship, and software development in order to create quality electronic collections supporting scholarship in the humanities and social sciences, provides one of the few collections of well curated streaming content supporting digital humanities scholarly research and pedagogy, and provides sophisticated services over this content including content retrieval by persistent URI, user-mediated clip services that mint new URIs on demand, and basic annotation functionality. Alexander Street Press critical edition video collections are especially notable for scholarly quality and extent and for

[12] http://lib.stanford.edu/DMSTech

the complementary content (e.g., play scripts) also published by the Press. Work with annotation of saved to desktop video was part of the AXE-Zotero integration work done in OAC Phase I, so this demonstration will build on and extend a major thread from Phase I.

Annotation of Digital Emblematica

- *Critical issues addressed / demonstrated*: This demonstration experiment will feature annotation of both structured text and still-images and will explore: annotations with bodies drawn from a hierarchical vocabulary (Iconclass), thereby exploiting OAC data model separation of Annotation and Annotation Body; annotations involving targets at multiple levels of granularity (e.g., motto, pictura, emblem, and book all potentially considered at item, manifestation, expression, and work levels); distinctions between citation and anchor; annotations targeting dissimilar resources (e.g., a pictura segment in a published emblem and in an applied context such as architecture or porcelain); and annotations of targets in specific context (e.g., of an emblem pictura as used in one context, but not in another).

- *Collaborating project & collection suitability*: The *Emblematica Online* project, itself a collaboration between the University of Illinois and the Digital Library of the Herzog August Bibliothek (HAB) in Wolfenbüttel, Germany and currently supported with joint funding from the NEH in the U.S. and the DFG in Germany, focuses on interoperability across digital emblematica repositories and features a model of metadata[13] that recognizes multiple levels of resource granularity and the distinctions (and attendant implications for identification) between digital surrogates for an object and intellectual resources as citable elements of scholarship. The HAB and Illinois collections of digitized emblem books (rivaled only by the Sterling Maxwell Collection in Glasgow) are ample for this demonstration and already richly annotated with original and normalized motto transcriptions and Iconclass descriptors. The proposed work will build on the applicable knowledge and content that has been and will continued to be developed by the *Emblematica Online* project and will exploit already cultivated relationships between *OAC* and *Emblematica Online* researchers at Illinois, maintaining momentum from Phase I and giving high confidence of a useful outcome.

4.4 March 2011: The Using OAC Workshop -- Promulgation & Recruitment

In addition to the initial 4 annotation demonstration experiments outlined above it is our intent during the first 6 months of OAC Phase II to select a second set of 4 annotation demonstration experiments. As described below, these projects will be selected through an open, competitive Request for Proposals (RFP).

In order to encourage well-informed responses to the RFP, we propose to coordinate release of the RFP with a ***Using OAC Workshop*** in March 2011. While in no way a pre-requisite for responding to the RFP, we suspect many would-be RFP respondents will find the workshop helpful in preparing their RFP responses. If nothing else they could be more confident after attending the workshop in preparing RFP responses. The workshop will be geared to provide attendees with an in depth introduction on how to implement the OAC data model and ontology. In addition to recruiting potential collaborators and helping to ensure informed responses to the RFP, this workshop also will be designed to promulgate the OAC data model and ontology to a larger audience of would-be implementers and project managers. Also, through exposure of use cases contributed by each of the participants, the workshop will provide an opportunity for

[13] http://www.ces.arts.gla.ac.uk/html/spine.htm

cross-discipline discussion of common issues and requirements spanning the domain of digital resource scholarly annotation, giving the OAC additional data points for understanding the state of research and thinking in this realm.

To be effective and to achieve desired outcomes, the size of such a technically oriented and in depth workshop must be managed. (We anticipate that there may be subsequent opportunities later in Phase 2 for larger, shorter, more PR-focused OAC "roll-out" sessions in North America, Europe and/or Asia, possibly in conjunction with conferences like JCDL, DH, OR; the March 2011 workshop is intended more to train early adopters willing to help with refinement of our alpha and early beta specification releases.) We plan to limit workshop participation to a maximum of between 15 and 20 participants (not including OAC participants involved in the leading the workshop). In order to avoid over-subscription for the workshop and to help insure pre-workshop preparation by the participants and leaders alike, we propose to include in the request for workshop participation a requirement that interested attendees submit a 1 to 2 page brief describing an annotation use case of particular interest to them. Proposed draft text of request for workshop participation is included below as section 6 of this proposal. We propose to use a small amount of the resources remaining from OAC Phase I to issue the call for workshop participation in October 2010, and then follow-up with each individual responding to the call as necessary to clarify and flesh out each use case submitted. The call for participation would go out through Centernet, DLF, and other appropriate venues.

In January, at the start of OAC Phase II, the submitted use case briefs will be reviewed by OAC Phase II PIs and co-PIs. As necessary the number of participants will be winnowed based on the use case briefs submitted. In any winnowing, we will seek to select complementary (albeit diverse), domain-relevant, challenging, user-driven scenarios featuring high quality scholarly digital resources. All else being equal, we will favor briefs that look to have the potential to be developed into viable responses to the RFP, which means we will favor briefs submitted by projects and institutions having the pre-requisites to author a viable RFP response. Workshop leaders will then leverage the selected use cases as fodder for workshop discussions and illustrations of how the OAC data model and ontology could be implemented in a variety of practical contexts.

Finally, to incentivize participation, we will fund as part of this project meeting space, participant air travel (up to $900 per participant) and participant hotel accommodations for at least 15 attendees (i.e., in addition to funding travel for OAC participants leading the workshop). To minimize meeting room costs and simplify logistics, the workshop will be held on the Campus of the University of Maryland or in facilities in Chicago managed by the University of Illinois or the Committee on Institutional Cooperation (CIC).

4.5 June 2011 through May. 2012: OAC Demonstration Experiments 5 - 8

As mentioned, an additional 4 use cases and collaborators will be selected through an open, competitive, RFP-based process (draft RFP language is included below as section 7 of this proposal). Sufficient resources have been included in the budget of this proposal to fund each project selected via the RFP for their part of the collaboration at between $30,000 and $45,000 for a period of performance of 9 to 11 months, beginning in June 2011. Budgets proposed for Illinois and Maryland assume each will collaborate with 2 of the selected projects.

RFP responses will be reviewed by a panel consisting in equal numbers of funded OAC principal participants and experts from outside of OAC (e.g., drawn from the OAC Phase 1

advisory committee and from the 2 technical review panels convened during OAC Phase I). Criteria for ranking proposals will include those outlined above as regards project suitability for collaboration, congruence of OAC and proposer objectives, confirmation of pre-requisites for a focused scholarly annotation experiment of appropriate scale and scope, track record of proposer, and assessment of likely potential that experiment / demonstration will further the goals of OAC. Panel will be instructed to select a set of 4 annotation demonstration experiments that are complementary to demonstration experiments 1 - 4 and one to another in terms of data modeling challenges, range of formats explored, maturity of digital content and maturity of annotation needs of discipline or community involved. We anticipate that at least 2 of the projects will be humanities discipline based, but 1 or 2 of the projects may focus more on the sciences, or on the general problem of Web annotation as deemed appropriate by the review panel. A goal of the selection process will be a set of demonstrations that explore requirements different one from another and complementary to those annotation needs and requirements explored in the initial 4 demonstration experiments, e.g., ideally 4 distinctive segment description problems will be explored by the set of demonstration projects selected. Capacity and special capabilities (and potentially limitations) of Maryland and Illinois collaborators will also be taken into account in the selection process.

4.6 Spring 2012: Finishing Data Model Refinement & Review

Refinement of the OAC Data model and ontology will be ongoing throughout Phase II. Los Alamos will take the lead on this task, assisted by Queensland and supported by Illinois and Maryland. We also will seek input and feedback from each annotation demonstration experiment. Our goal before the end of OAC Phase II is to provide a guide to the OAC data model, documentation of the OAC ontology, one or more guides to recommended serializations, and any ancillary documentation as required to insure that the wording and technical explanation is clear and complete. This should then enable the implementation of independent OAC-based clients and services with access only to the documents, rather than the more hands-on approach that we anticipate will be required during the Phase II annotation demonstration experiments.

OAC Phase I established the benefit of periodic in-person technical reviews (for Phase I these were held in October 2009 and May 2010). We plan to exploit such meetings during OAC Phase II as well. The occasion of the workshop in March 2011 will provide an opportunity, i.e., either immediately before or immediately after the workshop, for face-to-face discussion of the data model and ontology. This will serve as part of our pre-beta release review. Then, prior to 1.0 production release in the late spring of 2012, we will convene another in-person technical review of the OAC data model, ontology, and ancillary documentation.

The actual membership of the technical review panels remains to be decided, but we will follow the same pattern as used during OAC Phase I, specifically a panel composed of principals involved in OAC, participants experimenting with the data model, and 2 or 3 additional outside reviewers knowledgeable about Web architecture, linked data, and/or annotation. Rosters for technical review panels will not be set and invitations not issued until closer to the events; however, the following draft rosters are meant to be illustrative of likely invitees:

Draft roster for March 2011 mini-technical review in conjunction with *Using OAC Workshop*:

- OAC: Herbert Van de Sompel and/or Rob Sanderson (LANL)

- OAC: Jane Hunter and/or Anna Gerber (Queensland)
- OAC: Tim Cole, Tom Habing, and/or Allen Renear (Illinois)
- OAC: Doug Reside and/or MITH developer (Maryland)
- Participant from experiments 1-4, e.g., Ben Albritton (Stanford)
- Participant from experiments 1-4, e.g., Aaron Wood (Alexander Street Press)
- Outside expert, e.g., Michael Nelson (Old Dominion University)
- Outside expert, e.g., Bernhard Haslhofer (University of Vienna / Cornell University)
- Outside expert, e.g., Zotero Developer and/or JSTOR Developer

Draft roster for Spring 2012 pre 1.0 release technical review meeting:

- OAC: Herbert Van de Sompel and/or Rob Sanderson (LANL)
- OAC: Jane Hunter and/or Anna Gerber (Queensland)
- OAC: Tim Cole, Tom Habing, and/or Allen Renear (Illinois)
- OAC: Doug Reside and/or MITH developer (Maryland)
- Participant from experiments 1-8
- Participant from experiments 1-8
- Participant from experiments 1-8
- Participant from experiments 1-8
- Outside expert, e.g., Michael Nelson (Old Dominion University)
- Outside expert, e.g., Bernhard Haslhofer (University of Vienna / Cornell University)
- Outside expert, e.g., Zotero Developer and/or JSTOR Developer
- Outside expert (budget allowing)

4.7 Advancing Understanding of Digital Resource Scholarly Annotation

While annotation demonstration experiments are ongoing, the Illinois team will conduct a project-wide analysis of development activities and outcomes that will document advances, challenges, lessons learned, and the applicability of data model implementations across use cases illustrated by each demonstration project. Each individual demonstration project will serve as a case study representing particular requirements scholarly users and associated demands on the data model. Through comparative cross-case analysis, we will assess range of user requirements addressed, how well they are accommodated by implementation of the data model, and areas needing further study, research and/or development or testing.

Case data will include documentation from each demonstration project, as well as any data we are able to collect through interviews with and observation of demonstration experiment collaborators and participants (as a way to capture project team perspectives, solutions, and challenges in development as they happen). Where possible, views of scholarly end users' affiliated with demonstration experiments will be captured to determine annotation needs and viability of the approaches taken by the demonstration projects to meet these needs. Each of the 8 projects will be developed as a basic case study, with 2-3 selected for deeper investigation to highlight the challenges associated with certain kinds of content and/or certain types of scholarship, e.g., to understand the varying demands and requirements of historical researchers vs. literary researchers when annotating text, images, or video.

A CIRSS data analyst will lead the case study research under the direction of the investigators at Illinois. Baseline documentation for each demonstration experiment will provide a balanced descriptive foundation of 8 cases, which will be extended through systematic data collection using interview and observation protocols developed to capture team activities through the course of the experiments, and building up longitudinal data across the duration of development. The cases selected for deeper analysis will include further interaction with key informants on those experiments, to document team views and activities for the particular annotation problems addressed in that demonstration. A final report will synthesize the results across cases, identifying the common issues and unique aspects of scholarly annotation needs and the robustness and limitations of the data model in meeting the aims of the various demonstrations and the OAC project overall.

5. Staffing, Outcomes and Sustainability

5.1 OAC Phase II Staffing

The table below details staff by name, both OAC team members and those associated with collaborating projects, who will be involved in each of the first four annotation demonstration experiments. For each OAC team member an entry is included showing their broader role in OAC Phase II; note that for these individuals, the period of performance and equivalent level of effort shown is total, i.e., total for the annotation demonstration experiment plus for their broader role in the core OAC Phase II project. Level-of-effort of in-kind contributions of collaborating project personnel are defined by the collaborating projects. Listed at the end of the table are the additional OAC core personnel not directly involved in any of the first four demonstration experiments.

The Open Annotation Collaboration Phase II: Demonstration & Refinement

Staffing Summary

Name & Affiliation	Period of Performance	Level of Effort[1, 2]	Role in OAC Phase II
Use Case Demo #1: Annotation Supporting Collaborative Development of Scholarly Editions			
Prof. Jane Hunter eResearch Lab (Queensland)	Jan 2011 - June 2012	10% (in-kind)	PI
Anna Gerber eResearch Lab (Queensland)	Jan 2011 - Dec. 2011	100%	co-PI
Roger Osborne Sch. Of English, Media Studies & Art History (Queensland)	Jan 2011 - Dec. 2011	in-kind from partner project researcher (~10%)	
Prof. Paul Eggert Sch. of Humanities & Social Sci. (Univ. of New South Wales)	Jan 2011 - Dec. 2011	in-kind from partner project researcher (~10%)	
Use Case Demo #2: Annotation of Digitized Medieval Manuscripts			
Rob Sanderson Research Library (Los Alamos)	Jan 2011 - June 2012	20%	co-PI
Benjamin Albritton Digital Library Systems & Services (Stanford)	Jan 2011 - Dec. 2011	in-kind from partner project researcher	
John Haeger Digital Library Systems & Services (Stanford)	Jan 2011 - Dec. 2011	in-kind from partner project researcher	
To-be-named Developer(s) Digital Library Systems & Services (Stanford)	Jan 2011 - Dec. 2011	in-kind from partner project researcher	
Use Case Demo #3: Annotation of Subscription Streaming Video Content			
Doug Reside Maryland Inst. for the Humanities (Maryland)	Jan 2011 - June 2012	10%	PI
Developer-to-be-named Maryland Inst. for the Humanities (Maryland)	Jan 2011 - June 2012	100%	Experiment support, testing & development
Designer-to-be-named (Maryland) Maryland Inst. for the Humanities (Maryland)	Jan 2011 - June 2012	4%	Experiment support, testing & development
Aaron Wood Alexander Street Press	Jan 2011 - Dec. 2011	10%	
Developer-to-be-named Alexander Street Press	Jan 2011 - Dec. 2011	27.9%	
Notes:			
[1] For OAC core team members, level-of-effort shown is total level of effort on OAC (i.e., level-of-effort for annotation demonstration experiment plus level-of-effort of the individual for other OAC tasks)			
[2] The effort reported as "in-kind" are resources that will be available should this proposal be funded; however, these resources will not be formally tracked nor reported and should not be construed as cost sharing on the part of the University of Illinois			

The Open Annotation Collaboration Phase II: Demonstration & Refinement			
Staffing Summary (continued)			
Name & Affiliation	**Period of Performance**	**Level of Effort[1, 2]**	**Role in OAC Phase II**
Use Case Demo #4: Annotation of Digital Emblematica			
Prof. Timothy Cole Library, Grad. Sch. of Library & Info. Sci. (Illinois)	Jan 2011 - June 2012	5% (in-kind)	PI
Tom Habing Library (Illinois)	Jan 2011 - June 2012	10%	Experiment support, testing & development
Developer-to-be-named Library (Illinois)	Jan 2011 - June 2012	50%	Experiment support, testing & development
Prof. Emeritus Tom Kilton Library, Grad. Sch. Of Library & Info. Sci. (Illinois)	Jan 2011 - Dec. 2011	1 academic yr @ 7%	
To-be-named Academic Hourly (grad students) Germanic Languages & Literatures (Illinois)	Jan 2011 - Dec. 2011	1 academic yr @ 30%	
Prof. Mara Wade Germanic Languages & Literatures (Illinois)	Jan 2011 - Dec. 2011	in-kind from partner project researcher	
Prof. Myung-Ja Han Library (Illinois)	Jan 2011 - Dec. 2011	in-kind from partner project researcher	
OCA Core Personnel (not listed above)			
Herbert Van de Sompel Research Library (Los Alamos)	Jan 2011 - June 2012	10%	PI
Prof. Carole Palmer Grad. Sch. of Library & Info. Sci. (Illinois)	Jan 2011 - June 2012	2%	co-PI
Prof. Allen Renear Grad. Sch. of Library & Info. Sci. (Illinois)	Jan 2011 - June 2012	3%	co-PI
Kevin Trainor Grad. Sch. of Library & Info. Sci. (Illinois)	Jan 2011 - June 2012	10%	Project coordination
Jacob Jett (Academic Hourly) Grad. Sch. of Library & Info. Sci. (Illinois)	Jan 2011 - June 2012	30%	Data synthesis
Notes:			
[1] For OAC core team members, level-of-effort shown is total level of effort on OAC (i.e., level-of-effort for annotation demonstration experiment plus level-of-effort of the individual for other OAC tasks)			
[2] The effort reported as "in-kind" are resources that will be available should this proposal be funded; however, these resources will not be formally tracked nor reported and should not be construed as cost sharing on the part of the University of Illinois			

5.2 Advisory Committees

As with OAC Phase I, an Advisory Board (approximately 12 individuals) will be named at the outset of Phase II to serve as a resource for the OAC team during Phase II. This Board will consist of outside experts well versed regarding scholarly annotation practices, digital humanities, Web architecture, and/or linked data and the semantic Web. Board members will not be recruited until funding for Phase II is in place, but please see section 10 of this proposal for a list of potential nominees for the Phase II Advisory Board. We anticipate inviting several members from the Phase I Board to continue for Phase II. We do not anticipate an in-person meeting of the Advisory Board but will convene as a whole at least once by phone or Skype and will work with Board members one-on-one and in small groups on an ongoing and ad hoc manner over the course of Phase II (as we have done during Phase I).

Also as was done during OAC Phase I, we plan to convene 2 in-person technical review panels during the course of OAC Phase II (see Section 4.6, above). These panels will consist of a mix of selected OAC team members and outside experts. For OAC Phase II outside experts will be both at large and drawn from projects collaboration in ongoing annotation demonstration experiments. Section 4.6 gives preliminary rosters for planned technical reviews (subject to change).

5.3 Deliverables & Dissemination of Outcomes

The primary deliverables planned for OAC Phase II are listed below:

- *Using OAC Workshop* (March 2011)
- Beta release of OAC Data Model & Ontology (Spring 2011)
- Workshop Summary Report (June 2011)
- Reports from annotation demonstration experiments 1 - 4 (by December 2011)
- *Pre 1.0 release Technical Review Panel Meeting* (Spring 2012)
- Reports from annotation demonstration experiments 5 - 8 (by May 2012)
- Report synthesizing results from demonstration experiments 1 - 8 (June 2012)
- Prototype tools, services and software libraries as developed during demonstration experiments

To conveniently co-locate them, the workshop summary report, reports from annotation demonstration experiments 1 - 8, and the report synthesizing results from these experiments will all be released as volumes in the new *CIRSS Technical Report* series published by the Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Volumes in this series are automatically deposited in and made freely accessible and publicly discoverable through IDEALS (Illinois Digital Environment for Access to Learning and Scholarship), the institutional repository of the University of Illinois at Urbana-Champaign. In addition we anticipate that some of these reports will be simultaneously published as part of report series at OAC partner institutions and/or demonstration experiment collaboration project institutions. Such simultaneous publication will be noted in IDEALS metadata. All project publications will be linked from the OAC project Website.

Also, as has been the practice during OAC Phase I, summaries of and significant findings from the OAC project as a whole and from each of the various annotation demonstration experiments undertaken will be submitted to peer-reviewed conferences and journals for presentation and potential publication. Potential conference venues for dissemination would include (for example, all listed held at least annually): Coalition for Networked Information (CNI) membership meeting, the Joint Conference on Digital Libraries (JCDL), the Digital Humanities (DH) conference, the Annual Meeting of the American Society for Information Science & Technology (ASIST), the Open Repositories (OR) conference, and the European Conference on Research and Advanced Technology for Digital Libraries (ECDL). Dissemination of the OAC data model and ontology, and guides and best practice usage recommendations regarding same, will be via the Web. These documents will be linked from the primary OAC project Website and/or from Websites maintained by partners and collaborating projects, as appropriate. Notice of major releases will be posted to the oac-discuss Google group and to other listservs (e.g., DLF listserv, Web4Lib listserv) as appropriate.

5.4. Intellectual Property

As more fully described in our agreement with the Foundation governing intellectual property matters (separate document), all software deliverables will be made available to the non-profit educational, scholarly and charitable communities on a royalty-free basis under an open source license allowing free redistribution, derived works, etc.; all pre-existing software that will be

embedded in or used to derive deliverables is already made available under appropriate open source license. Reports and Web-posted deliverables will be made freely and openly available to the non-profit educational, scholarly and charitable communities on a royalty-free basis, under a Creative Commons Attribution license permitting non-commercial use and modification.

5.5 Reporting

Separate from and in addition to the deliverables enumerated above, the Illinois OAC Phase II project team will provide to the Foundation at least annually during the course of the OAC Phase II project and also at the conclusion of the project (anticipated June 2012) a narrative report detailing project progress, expenditures and accomplishments during the reporting period. These reports, submitted electronically in PDF (with hard copy available if requested) within 90 days of project start anniversaries or project end date, will describe activities during the reporting period and summarize progress to date on OAC Phase II goals and objectives. Reports will comment on all grant expenditures and performance against planned budgets during the reporting period and will include financial reports generated from the Banner Enterprise system used at the University of Illinois at Urbana-Champaign and containing a summary of financial expenditures that encompasses the information outlined in the Foundation's Scholarly Communications and Information Technology Program reporting guidelines and instructions (available: http://msc.mellon.org/guidelines/reporting-instructions).

5.6 Possible Next Steps after OAC Phase II

A primary objective of OAC Phase II is to demonstrate in real-world scholarly use cases and contexts the viability and usefulness of a shared, Web and resource-centric data model for describing and exploiting scholarly annotations of digital resources on the Web. Through proof-of-concept experiments with and demonstrations of the OAC model in the context of a range of use cases spanning multiple domains, we expect to stimulate adoption and adaption of our data model, most immediately in the domains of and with the projects with which we collaborate on the annotation demonstration experiments planned for Phase II. Assuming we are successful in meeting this expectation, we then envision 3 kinds of follow-up efforts that may be warranted:

1. **Workshops and additional documentation**. Maintenance of the OAC data model and related developer resources must be assumed long-term by the digital humanities and digital libraries communities, but in the immediate aftermath of OAC Phase II we anticipate there may well be a modest, short-term role for the Collaboration to still play in promulgating standards and best practices for scholarly annotation of digital resources, in serving as a resource for new implementers, and in disseminating the findings of Phase II work.

2. **Demonstration and prototyping of OAC tools and services in new domains**. While significant effort and additional funding will still be needed to extend our initial modeling and prototyping work even in the domains explored during Phase II, we anticipate that such further work will be (and should be) integrated into new and ongoing domain-specific projects and initiatives and so would not on its own warrant continued involvement of the Collaboration. On the other hand, we can anticipate our successes also may stimulate interest and/or uptake in domains beyond those directly addressed by Phase II annotation demonstration experiments. It is not clear now, however, whether there might still be a need to further grow interest piqued with additional experiments in new domains building on the model used for and the results from Phase II. We will begin to gain a better sense of what is

needed and how proactive the Collaboration might want / need to be in this regard from the responses next spring to the RFP issued to identify annotation experiments 5 - 8.

3. **Identification & implementation of cross-domain annotation services**. Most intriguing and potentially most impactful is the possibility that through our work in Phase II we will be able to detect, recognize, and flesh-out common themes and features of scholarly annotation of digital resources that span use cases and domains. Already there are suggestions from Phase I work that some annotation data modeling issues span domains, e.g., issues of how best to perform ontological annotation show up in the literary analysis domain, in digital emblematica studies, in bio-medical journal article annotation. Similarly there appear to be needs in searching for and filtering annotations that span domains and use cases. Of keen interest is the possibility that given the foundation of a shared, interoperable model of scholarly annotation, annotation services and tools built in one scholarly domain may be easily adaptable for use in another. This potential has been in mind since the outset of the OAC; so, a natural next step for the Collaboration, if preliminary indicators are borne out by the work done during Phase II, would be to experiment, design, and implement (in prototype) a few of the most broadly useful interoperable scholarly annotation services and tools. Phase II is intended in part to help us identify, prioritize, and preliminarily assess the potential value of such domain-spanning annotation applications.

At this point the need for and relative value of these potential next steps is uncertain. Also uncertain is the degree to which we can expect these next steps to happen organically, picked up and moved forward by the larger DH and DL communities without substantial further intervention of agency like the OAC. Given interest observed during Phase I, it seems likely that given the model for and examples of successful domain-based OAC implementations we will create during OAC Phase II, we may be able to leave domain-specific adoption after Phase II to individual domain-scoped projects. However, with regard to proving the potential for, utility of and benefits of annotation interoperability across domains and of cross-domain annotation services and tools, we see more likelihood of a continuing need for the OAC, at least in the near term after the conclusion of Phase II.

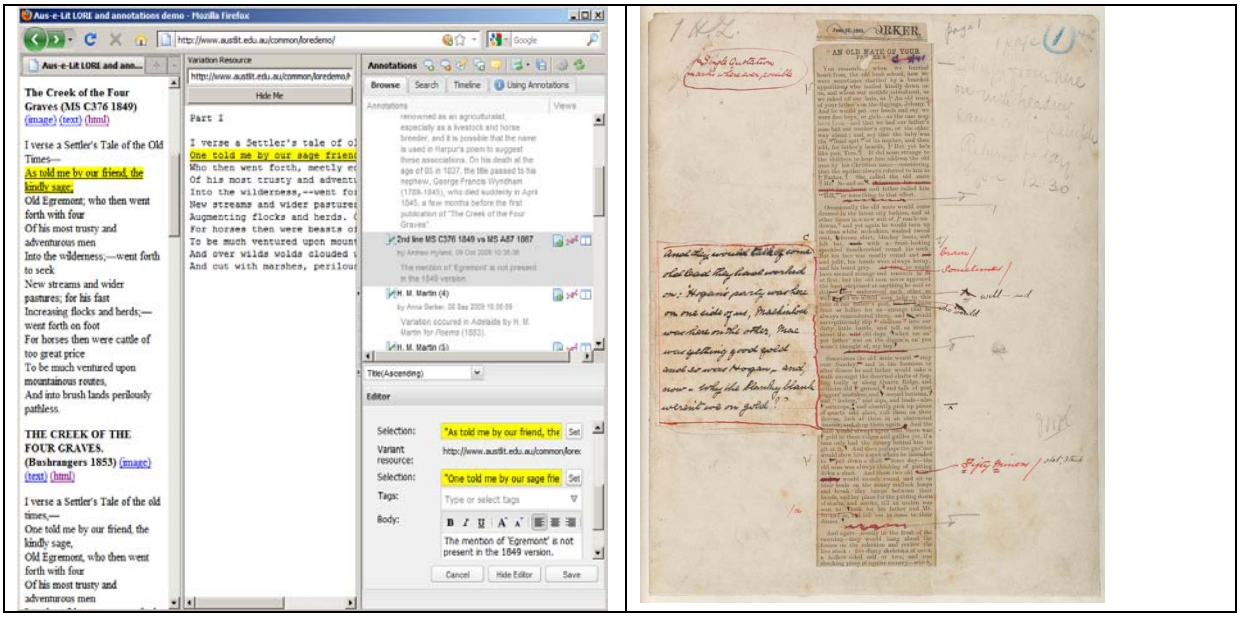
9. Annotation Demonstration Experiments 1 - 4

9.1 Annotation Supporting Collaborative Development of Scholarly Editions

Summary & Rationale: Scholarly editions in print form have long been important to the study of literature. There is a critical need for useful Web-based tools and services to support the collaborative development of scholarly editions by distributed researchers. A key requirement in developing a scholarly edition is the ability to create, share and reply to scholarly commentary attached to variations between versions of a particular work. Although there currently exist a number of tools for comparing and displaying differences between multiple versions of digitized texts, e.g., MEDITE[14], Juxta[15] and MVDDeskViewer[16], there is no established, common model or tool for documenting, sharing or replying to annotations on specific differences between digitized versions of a common work.

The aim of this collaboration with AustLit is to evaluate and demonstrate the applicability of OAC in the context of annotating variations between versions of a literary work. This experiment will explore annotation of annotations (including nested annotations); annotations involving multiple targets and/or ORE aggregations as targets; annotations involving structured text as annotation body (content); distinctions between reference and targeting; and issues arising with tools that enable annotations spanning multiple manifestations of FRBR expressions and works. It will also demonstrate the potential for robust retrieval by target of annotations described using the OAC data model.

Example: "While the Billy Boils" by Henry Lawson, is a set of 50 short stories first published in 1896. Prior to the publication of the first edition of the book, many of the stories had been published in newspapers such as the Bulletin. Subsequent editions of the book also were



published (1905, 1913 etc). In each publication, the stories had undergone numerous editions

[14] http://www-poleia.lip6.fr/~ganascia/Medite_Project

[15] http://www.juxtasoftware.org/

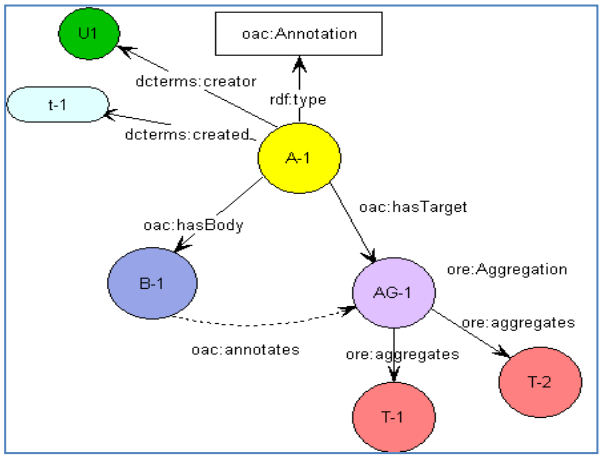
[16] http://multiversiondocs.blogspot.com/

and changes, which are documented in scholarly editions, e.g., “An Old Mate of Your Father’s” – first story in the collection.

Methodology:

Step 1. Integrate the nmerge[17] program with the Aus-e-Lit Textual Variants Interface - to extract, collate and display differences between versions of same work from different publications (e.g., “An Old Mate of Your Father’s” – version 1 from the newspaper serial in the Bulletin, version 2 (1896 book published by Angus and Robertson), version 3 (1905 book published by E.W. Cole).

Step 2 – Evaluate the OAC data model’s ability to capture, share and re-use, and reply to commentaries on selected differences between texts – via a Web interface e.g., “this change was done by Lawson’s wife Bertha in 1897 whilst living at Karumbah Downs and the evidence is provided by the handwritten notes on the original manuscript”. **In particular, we aim to investigate the use of nested aggregations for this use case and any performance issues that may arise from this approach.**



Step 3. Develop an annotation tool (based on the OAC data model) that is an extension of the Aus-e-Lit Textual Variants Interface – that:

- enables scholars to highlight/select textual variants across versions and to attach structured commentary and publish that commentary is a form that is shareable and re-usable
- enable other scholars to respond/reply to the commentary
- enable search, browse and retrieval of annotations on specific textual variants

We also anticipate that investigation of this use case will be informative as to the utility/applicability of OAC segment descriptions and time context constructs (oac:when) for this annotation class.

Institutions & Collaborators: AustLit[18] is a collaboration of the National Library of Australia, the Australian Research Council, and 12 Australian universities led by the University of Queensland. This annotation demonstration experiment also will leverage outcomes of the Aus-e-Lit project,[19] an ongoing research project examining and developing tools and services, including for annotation, to meet the needs of researchers involved in the study of Australian

[17] http://code.google.com/p/digitalvariants/
[18] http://www.austlit.edu.au/
[19] http://www.itee.uq.edu.au/~eresearch/projects/aus-e-lit/

literature and Australian print culture. In addition to AustLit and the University of Queensland, other collaborators in Aus-e-Lit include the Association for the Study of Australian Literature, the Australian National Data Service, and the Australian Research Collaboration Service. The main technical infrastructure for this annotation demonstration experiment will be hosted by the University of Queensland, which also hosts the primary technical infrastructure for Aus-e-Lit.

The work of this annotation demonstration experiment will be directed jointly by Professor Jane Hunter (University of Queensland, OAC, Aus-e-Lit) and Professor Paul Eggert (University of New South Wales, AustLit Advisory Board).

Collections: There are approximately 25,000 full text works available through AustLit. This includes works digitized by AustLit as well as complementary creative works of Australian authors held by the National Library of Australia, state and university library repository throughout Australia, and freely available services such as the Internet Archive. Access to AustLit digitized full-text and some third-party digitized full-text is by subscription; AustLit subscription content is currently accessible to patrons of all Australian State Libraries, the National Library of Australia, almost all university libraries in Australia and a number of Australian public libraries and schools. AustLit users at subscribing institutions may use AustLit content for the purposes of individual research, study, criticism and review. Search encompassing additional resources such as Google Books is also available through AustLit federated search services. The AustLit data model implements parts of the FRBR model and recognizes distinctions between works, expressions, and manifestations, making AustLit particularly well suited as a partner for a demonstration experiment addressing the use of annotation to support distributed development of literary scholarly editions.

9.2 Annotation of Digitized Medieval Manuscripts

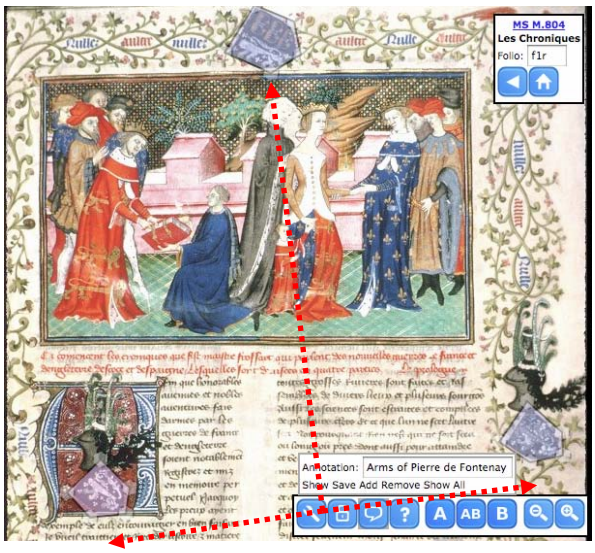
Summary & Rationale: Use cases involving annotation of digitized medieval manuscripts are important, timely and challenging. As targets of annotation, digitized medieval manuscripts are of interest to scholars both as textual documents and for their artistic elements and use of graphic embellishments. Scholars annotate digitized manuscripts at multiple levels of granularity, e.g.: a line of text on a page; an arbitrary, and arbitrarily shaped, segment of a page image; a sequence of manuscript pages (facing/verso or when multiple orderings of pages are possible). Tools must handle a range of structures (e.g., manuscripts may present as bound volumes, scrolls, unbound pages, maps, fragments of these). Annotations often compare and contrast across manuscript repository boundaries; this is currently hampered by a siloization of digitized manuscript archives and the corresponding lack of a shared data model for annotation and collection interoperability.

Through this collaboration with the proposed Stanford-led effort to define a modular and interoperating environment for digitized medieval manuscript collections, the OAC will simultaneously explore annotation of both structured text and still-image media, implying also a broad range of use cases and classes of annotation, e.g., from granular (evidence on a page image of a scribe's hand) to broadly scoped (on the cultural significance of a particular graphic construct used). This work will inform the OAC regarding annotation of annotations (e.g., annotation of transcribed text annotations); multiple approaches to and tools for segment description (see also below regarding TILE and T-PEN initiatives), including trade-offs between minting new URIs and providing segment descriptions; annotation bodies that are structured (e.g., TEI); annotations of aggregate targets and/or multiple targets, some of which may be discontinuous and/or distributed across repository boundaries (e.g., Stanford's "return-and-reuse" scenario, which can lead, for example, to a page image in one repository linked to a structured TEI transcript in another location); annotations that may change over time (e.g., need to update transcription annotations), providing an opportunity to examine both degree of annotation mutability supported by OAC data model and the potential importance in this domain of being able to exploit the Memento Framework to preserve the meaning of annotations that target mutable annotations.

From the perspective of a proposed new Stanford University Libraries project, *Defining a Modular and Interoperating Environment for Collections of Digitized Medieval Manuscripts, Tools, and Users*, annotation interoperability is a core exemplar. Scholars working with digital resources in this domain need to be able to research, compare and contrast manuscripts and manuscript pages, text fragments and graphic elements across collections of digitized manuscripts. Current tools, including nascent annotation and note-taking tools, written for one repository do not work with other repositories. Missing, among other things, is a shared data model that spans repository architectures. The pending Stanford project will seek to "catalyze and coordinate a process in which collections of digitized manuscripts reinvent themselves as modular repositories in an interoperating system, linked with tools designed to assist users seeking to analyse digitized source documents." [Stanford Proposal] The Stanford project has identified the OAC data model as a promising, "generalizable framework for exposing, consuming and annotating data" [ibid] across existing manuscript collections. Contingent on receipt of funding, the Stanford project plans to exploit the OAC data model, in combination with outcomes from T-PEN (Transcription-Paleographic and Editing Notation Tool project, St. Louis

University), TILE (Text and Image-Linking Environment tools project, University of Maryland), and other ongoing work to accomplish their goals for more interoperable digitized medieval manuscript collections. Given central role of annotation in manuscript research, direct collaboration with the OAC will facilitate achieving this objective.

Illustration & additional detail: Annotation of manuscripts will require the ability to target multiple resources and/or multiple segments of resources. In the illustration below, the substance (Body) of the annotation is about multiple appearances of a particular coat of arms on a given manuscript page. Flexibility built into the OAC data model supports multiple approaches for describing annotation targets in this instance. There are nuanced differences between each approach, but testing is required to determine whether these differences are significant in the domain of medieval manuscript annotation, and if so which approach is better in which circumstance. Interaction with curators and scholars who study these resources will help the OAC assess whether distinctions in modeling approach are significant, and if so, why, informing further refinement of the OAC data model.



Annotation associating & commenting on variants of a coat of arms found on a single manuscript page

As another example, transcription of text found in medieval manuscripts can be modeled as an annotation targeting part of a page image. Transcriptions have structure and are often modeled in this domain using TEI. Expressing transcriptions as annotations can be complicated if segments of text on the page image are discontinuous or if multiple page image instances exist for a given manuscript page. As noted, both TILE and T-PEN are looking at these issues. Potential mutability of transcriptions is another, potentially complicating issue. This collaboration will allow us to collectively vet and refine as necessary the OAC model and demonstrate the potential interactivity of the OAC model with the infrastructure and modular tools being developed by Stanford, TILE, et al. This work will simultaneously inform (and presumably influence) the direction of OAC, the Stanford infrastructure development, both TILE and T-PEN, and other anticipated projects involving annotations services for manuscript collections.

Methodology: The Stanford project plans to investigate a number of use cases related to annotation, including annotation use cases that demonstrate/interact with "return and reuse" scenarios (e.g., as mentioned above). Stanford also plans reference implementations (e.g., one

making use of Blacklight) to demonstrate the benefits of the proposed modular and interoperating environment for digitized medieval manuscript collections they are creating. OAC will work with Stanford investigators to identify use cases which might implement and demonstrate the benefits of the OAC data model in this domain. As work progresses, OAC participants will lend expertise and assist in mapping components of these use cases to the OAC data model; multiple mappings may be possible, in which case OAC participants will collaborate with Stanford investigators and potentially TILE and T-PEN researchers to examine trade-offs and other optimization considerations. Process will be iterative with feedback and lessons learned applied to refinement of the OAC data model (and presumably also to the broader development work being done by Stanford). OAC participants will help examine implications for repositories (e.g., object identification, etc.) and tool construction vis-à-vis larger context of Stanford's image stack, Web service, return-and-reuse approaches to manuscript repository interoperability. What are the requirements of use cases and planned reference implementations in this domain that can inform / drive refinement of the OAC data model? How do requirements to support annotations that span manuscripts and manuscript repositories contribute to / help drive design of technical infrastructure and modular architecture supporting manuscript services and the work being done by TILE and T-PEN?

Institutions & Collaborators: Stanford University Libraries, our primary collaborator for this demonstration experiment, has been a leader in the digitization of medieval manuscripts and the development of digitized manuscript repository design and services. Their newest pending project, *Defining a Modular and Interoperating Environment for Collections of Digitized Medieval Manuscripts, Tools, and Users* will feature interactions with the broader medieval manuscript research community including the TILE and T-PEN projects mentioned above, as well as a large number of digitized medieval manuscript repositories in North America and Europe,[20] all of which have the potential to generate additional, useful requirements-gathering use cases in this domain. Through our collaboration with Stanford, we anticipate the potential to interact with some of these repositories as well.

The work of this annotation demonstration experiment will be directed jointly by Rob Sanderson (Los Alamos National Laboratory, OCA) and Benjamin L. Albritton (Stanford University Libraries) supported by Doug Reside (University of Maryland, OCA, TILE).

Collections: This annotation demonstration experiment will exploit a significant and representative sample of existing digitized medieval manuscripts. In particular, the new Stanford project will use a reference repository implementation featuring a subset (or possibly all) of the content accessible through the production *Parker on the Web* service, possibly augmented with additional representative content from other sources. The content behind the *Parker on the Web* portal is a digital version of the Parker Library of Corpus Christi College, consisting of nearly 500 manuscripts (some 200,000 manuscript page images). The Parker Library includes a large number of pre-1066 manuscripts, some of the oldest manuscripts of English music, lavishly illustrated travelogues, famously illuminated Bibles, bestiaries and other treasures. Another likely target for this annotation demonstration experiment will be the digitized *Roman de la Rose* manuscript (hosted by Johns Hopkins University). Access to some digitized manuscript content requires a subscription.

[20] E.g.: http://romandelarose.org/, http://www.e-codices.unifr.ch/en, http://www.stgallplan.org/stgallmss/, http://parkerweb.stanford.edu, http://image.ox.ac.uk/, etc.

9.3 Annotation of Subscription Streaming Video Content

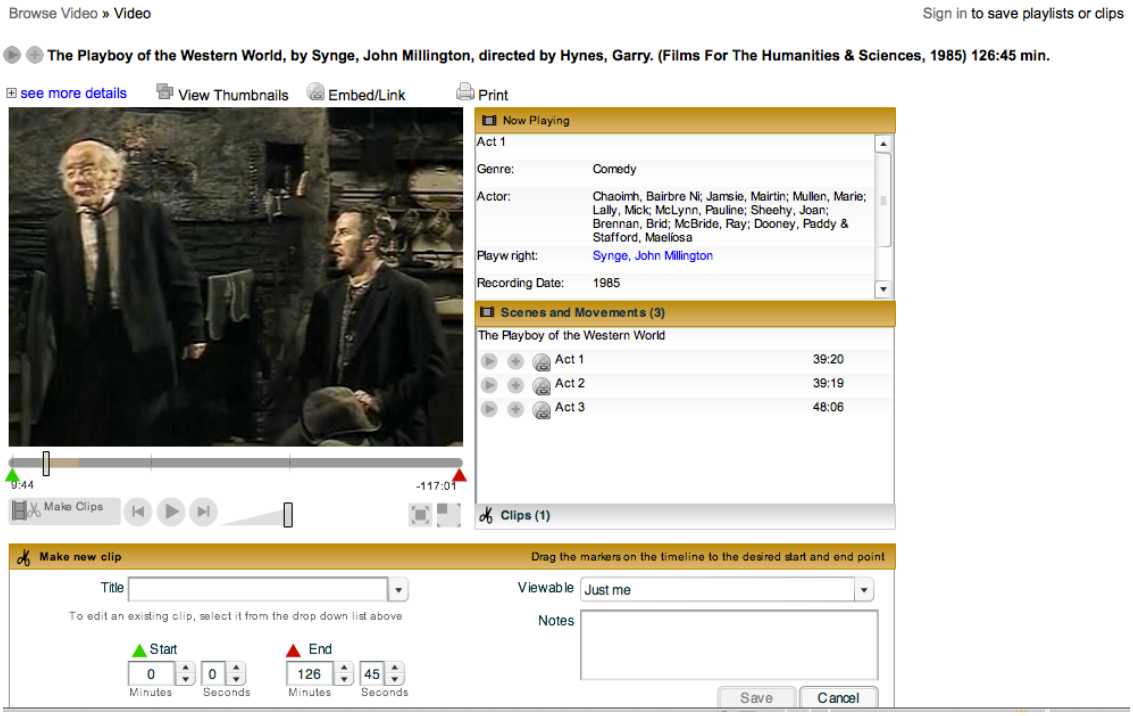
Summary & Rationale: Although there now exist several web based interfaces for tagging text and images, there are relatively few tools that allow a user to annotate a video. However, since the advent of YouTube and similar video sharing services, this format is among the most common posted on the web, and any initiative that seeks to annotate media fragments on the Internet must take the format into account. Much of the video content on the web is streamed rather than downloaded, which presents several important challenges for the OAC data model, especially in regard to segment description. How, for instance, do you reference moments or ranges in a video when frames may be dropped and timings distorted in the streaming process? How should movement of an object in a frame be described? Is it sufficient if the OAC segment description simply record a start and stop position, or should it be possible/useful to provide segment descriptions that mathematically describe the motion of both an object in the frame and the position of the camera?

OAC has begun to address some of these questions in its initial reference implementation of the Ajax XML Encoder (AXE) in Zotero, but there is more work to be done. In Zotero, users can make personal notes about particular regions of frames on videos they have downloaded and plan to cite in future work, but, cannot easily annotate streaming video. Further, despite its new shared Commons, Zotero is a tool primarily designed for personal notes and citation management rather than the sharing of annotation. OAC, on the other hand, is intended to support annotations of online content (i.e., not limited to downloaded content). Practical approaches for describing annotation targets (or bodies) that are segments of streaming content are therefore in scope. For OAC Phase II, we need another use case: ideally a repository of streamed scholarly videos. Some of the best such repositories useful for scholarly humanities research and pedagogy are those published by Alexander Street Press. Their "*Theatre in Video*," "*Opera in Video*," and "*History in Video*" (among others) offer licensed, high quality videos of significant pedagogical and research value.

Through this proposed collaboration with the Alexander Street Press, OAC will explore sophisticated annotation of streaming video, focusing on issues to do with annotation target segment description for streaming media; the persistence vs. variability of such segment descriptions; and mixed media multiple target annotations (e.g., as when comparing and/or contrasting a segment of a performance with the corresponding segment of digitized text play script or music score). In many cases, scholarly content on the web not be freely available at open URIs, but scholars may still wish to annotate such content and link it to other content that is freely available. The OAC data model must be supple enough to allow this. This collaboration will allow us to examine issues to do with annotations having multiple targets spanning both subscription and freely accessible content. Finally, the Press already provides a tool to annotate clips of videos in their collection, but it does not currently provide a means to further annotate segments of the frame. Further, the annotation format used is specific to the interface, and thus cannot accept annotations from other sources or export its own annotations to external tools. The existence of an independent, domain and collection-specific annotation tool will suggest functional requirements and more importantly allow us to examine ways to import and export OAC-described annotations.

Methodology: In this OAC collaboration with the Alexander Street Press we will seek to demonstrate the adaptability of the OAC data model and ontology to the specific needs of a

subscription-based video repository, and to develop a Web-based import/export tool capable of transferring annotations both into and out of the Press's collection-specific tool using an exchange format based on the OAC data model. MITH developers will extend the functionality of the video annotator built for Zotero in phase I to demonstrate the capacity to annotate curated, subscriber-restricted streaming video, using content provided by the Press as representative. Working in consultation with Aaron Wood (director of software product management for the Alexander Street Press), strategies for streaming video segment description will be examined and a preferred option implemented. Working in collaboration with the Press, MITH developers will then create an open source tool for importing annotations into and out of the collection-specific Alexander Street Press annotation tool. An open source FireFox plug-in that can capture annotations of streaming video resources through the Web browser and pass annotations back and forth between itself and the web-based tool import/export tool will then be created.



Screenshot of the video annotation tool already built into Theatre in Video

Institutions & Collaborators: Alexander Street Press was founded in 2000 to bring together the skills of traditional publishing, librarianship, and software development in order to create quality electronic collections supporting scholarship in the humanities and social sciences. It provides respected, scholar-quality, well curated digital collections in formats appropriate to the content. It has ongoing alliances with The Networked Infrastructure for Nineteenth-Century Electronic Scholarship (NINES) and its sister organization, 18thConnect: Eighteenth-Century Scholarship Online.

This annotation demonstration experiment will be directed by Doug Reside (University of Maryland, OAC) and Aaron Wood (Alexander Street Press).

Collections: Subsets from the Alexander Street Press Critical Video Edition Collection will be used for this annotation demonstration experiment. Video content has been semantically indexed, provided with permanent, persistent URLs and made searchable. Custom clips can be created by users and assigned a persistent, citable URL. (Subscription is required to access Critical Video Edition content.) Content is streamed and viewed in user's Web browser window through embedded video player. The Press, also publishes a database of full-text play scripts in North American Theatre Online and musical scores in Music Online, offering the opportunity to experiment with annotations simultaneously targeting performance video and representations of corresponding script and/or music score segments.

9.4 Annotation of Digital Emblematica

Summary & Rationale: As with digitized medieval manuscripts, digitized emblem books are of interest to scholars both as literary textual documents and for their artistic elements and use of graphics and iconography. More so than with the study of medieval manuscripts, however, the range of scholarly use cases involving digitized emblematica has become quite broad. Emblematic forms appear and are studied in architecture,[21] embroidery, tapestry, painted ceilings, porcelain,[22] political propaganda, and court and civic festivities. Emblem books represent a key genre of Renaissance texts, and through the study of applied emblematics, research into the use of iconography and symbolism in printed emblems now helps to inform our understanding of art, architecture, politics, vernacular languages, and popular culture during the European Renaissance. Additionally, emblem books appear occasionally in multiple editions and printings, sometimes identical, sometimes variant in small but significant ways, e.g., re-ordered, with variants on original pictura, with original pictura associated with variant mottos or other text. Plates for pictura occasionally were reused not only in other editions or translations of the emblem book for which they were created, but also in entirely different books or collections of emblems. For these reasons emblematica annotation use cases often involve multiple diverse annotation targets and multiple levels of the FRBR Group 1 entities hierarchy.

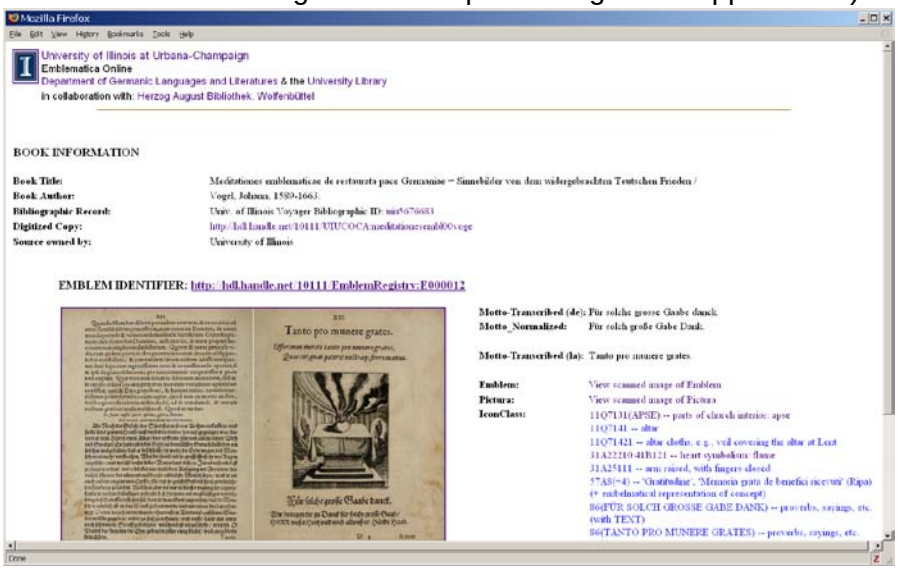
Through this collaboration with the *Emblematica Online* project, itself a collaboration between the University of Illinois and the Digital Library of the Herzog August Bibliothek (HAB) in Wolfenbüttel, Germany and funded jointly by the NEH in the U.S. and the DFG in Germany, OAC will explore a broad range of use cases and classes of annotation involving both structured text and still image annotation targets. This work will inform the OAC regarding annotation of annotations (e.g., annotation of transcribed and normalized mottos as annotations); annotations involving targets at multiple levels of granularity (e.g., motto, pictura, emblem, and book all potentially considered at item, manifestation, expression, and work levels); annotations having dissimilar targets and/or bodies (e.g., a Web accessible picture of an architectural emblem feature annotated by the motto of the print emblem instance from which the feature derived); annotations having annotation bodies drawn from a hierarchical controlled vocabulary with persistent identifiers (i.e., the multi-lingual Iconclass Thesaurus), thereby exploiting the separation of Annotation from Annotation Body inherent in the OAC data model. *Emblematica Online* has implemented a prototype emblem registry shared by UIUC and HAB and plans to expand this registry to include emblem books. CNRI-style handles are utilized. This collaboration will provide an opportunity to test the utility of these registry and identifier services to support interoperable annotations.

Illustration: This screen shot of a staff view of a digitized emblem shows the emblem-level identifier; metadata for the individual emblem including transcribed, normalized, and alternate language transcribed versions of the emblem motto; links to pictura only view (using an aDORe djatoka-based service); and Iconclass descriptors assigned to this pictura. Note that Iconclass descriptors are used both to identify the significance of entire scenes (e.g., gratitude, grateful

[21] E.g.: *The Emblem and Architecture: Studies in Applied Emblematics from the Sixteenth to the Eighteenth Centuries*. 1999. Ed. by P. Daly and H. Boker. Turnhout: Brepols.

[22] *Fired by passion: Vienna Baroque porcelain of Claudius Innocentius Du Pacquier*. 2009. Melinda and Paul Sullivan Foundation for the Decorative Arts, Meredith Chilton, editor-in-chief. Stuttgart: Arnoldsche Art Publishers: 388-95.

memory of benefits received) and to identify individual elements represented within an image (e.g., arms raised with fingers closed). Conventionally these descriptors are tied to the pictura as a whole and in the context of emblem. This provides an opportunity to examine whether these descriptors or select of these descriptors then represent annotations on the pictura, on the pictura but only as it appears in the context of the emblem (cf. bioinformatics scenario mentioned in section 3 above), or on a segment of the pictura (e.g., as might be useful to create a training set to assist machine recognition description assignment applications).



Methodology: Classes of annotations germane to emblem research will be defined in collaboration *Emblematica Online* researchers and users. Potential use cases involving these classes of annotations and providing opportunities to explore the issues outlined above will then be described. These use cases will be explored and developed or rejected by trial and error against a small selection of emblems from the UIUC collection of digitized emblems. A modified version of Pliny note-taking tool (with export to OAC data model capability) will be used to allow exploration of potential use cases. Variations on use cases will be examined to compare different options for describing classes of annotations.

Institution & Collaborators: The HAB at Wolfenbüttel is an established and important European research centre for the study of the medieval and early modern period and a well-matched partner to UIUC in this scholarly domain. Both institutions are especially strong in German-language emblem books and have developed communities of researchers engaged in emblem research, and both institutions have been actively developing and using their collections of digitized emblem books for more than a decade.

This annotation demonstration experiment will be directed by Professor Timothy Cole (University of Illinois, OAC) and Professor Mara Wade (University of Illinois, Past Head of the Department of Germanic Languages & Literatures, U.S. PI for *Emblematica Online*) supported by Professor Emeritus Thomas Kilton (University of Illinois, co-PI *Emblematica Online*).

Collection: Only a single emblem collection worldwide, the Sterling Maxwell Collection, University of Glasgow Library, rivals the emblem book collections at Illinois and the HAB that together offer the most comprehensive collection of emblematica in the world. The current project has digitized almost 400 additional volumes at UIUC, including approximately 50 German-language volumes, encompassing more than 7,200 individual emblems.