

## Representing Humanities Research Data using Complementary Provenance Models

### **CONTACT DETAILS:**

NAME: BRIDGET ALMAS, SAYEED CHOUDHURY, DAVID DUBIN  
RDA IG: RESEARCH DATA PROVENANCE IG  
EMAIL: BRIDGET.ALMAS[AT]TUFTS.EDU, SAYEED[AT]JHU.EDU, DDUBIN[AT]ILLINOIS.EDU

### **Objectives**

The IMLS-funded “Libraries: Transformation of the Humanities” (LTH) project is focused on designing infrastructure to support a collaborative network of collections of resources for the humanities. The LTH project team comprises researchers from John Hopkins University, Tufts University, and the University of Illinois.

The vision of the LTH project is to create tools to facilitate contributions from a wide range of users in a distributed and increasingly collaborative research environment. Users may come from diverse disciplinary cultures, be located in different countries, or be affiliated with different kinds of institutions or organizations. In addition, the ability to move the results of work by students and citizen scholars out of individual projects and into an infrastructure that incorporates data from disparate fields will allow broader participation in scholarly activities. These diverse user communities call for modeling approaches that will support precise identification of resource and user entities, along with the assignment of roles with respect to the annotation and description of resources. Similarly, due to the expected interdisciplinary range of these activities, our modeling approaches should aim for a level of abstraction that is independent of discipline-specific terminologies and constructs.

### **On-going activities:**

The Systematic Assertion Model (SAM) is a formal account of research data and its content (Wickett et al., 2012). Developed at the University of Illinois, the SAM framework relates data identity, change over time, and varieties of scientific or scholarly equivalence directly to research transactions, including data translation, annotation, and curation.

The LTH use cases include a breadth of computational and interpersonal activities that call for a more inclusive understanding of provenance than SAM alone can provide. Rather than sacrifice precision by generalizing one model, we are integrating complementary views of the same research enterprise in a single RDF description that combines SAM entities and properties with those from other models such as W3C's PROV and Open Annotation (OA). Taken together, these vocabularies can provide a richer, more contextualized view of data encoding and use in humanities research.

### **Results:**

The LTH project has identified two specific use cases - a simple data curation scenario and a more complex scenario involving the annotation of text reuse in fragmentary texts - and developed prototype representations of the data and activities using a combination of the SAM, PROV and OA data models. Perseids is a platform being developed by the Perseus Digital Library that integrates and extends open source tools and data models from a variety of projects to provide support for collaborative editing and annotation of textual resources (Almas and Beaulieu, 2013). Through new support from the Mellon Foundation, Perseids developers over the next two years will be implementing tools in Perseids that produce data according to the prototype representations for these use cases identified here. Our poster presents the details of these use cases and their forthcoming implementation.

## References

Almas, B., Beaulieu M. (2013). Developing a New Integrated Editing Platform for Source Documents in Classics. In *Literary and Linguistic Computing*; doi: 10.1093/lc/fqt046.

Wickett K. M., Sacchi, Simone, S., Dubin, D., and Renear, A. H. (2012). Identifying content and levels of representation in scientific data. In *Proceedings of the 75th Annual Meeting of the American Society for Information Science and Technology*. Baltimore, MD, October 26-30, 2012.