The Perseids Platform for Collaborative Editing and Annotation: A Technical Summary

The Perseids platform supports collaborative editing, annotation and publication of born-digital editions of source documents in the classics. Perseids is not one single application but an integrated environment built from a loose coupling of heterogeneous tools and services from a variety of sources. The development of the Perseids platform was inspired and motivated by the work of several pre-existing projects: the *Tufts Miscellany Collection* at Tufts University; the *Homer Multitext Project* at the Center for Hellenic Studies at Harvard University; and the *Papyri.info* project (Almas and Beaulieu, 2013).

A key goal behind the initial development of the platform was to enable original undergraduate research in the field of classics. This work was supported by grants from Tufts University, the National Endowment for the Humanities [grant HD-51548-12] and the Institute from Museum and Library Services. New funding from the Mellon Foundation will allow us to expand the platform over the next two years to support classroom-based collaboration on digital editions, beginning-to-end scholarly workflows, and the development of dynamic syllabi using the resources managed by the platform. Adherence to established data models and standards, publication of all code and data under open source licenses, and reuse of existing code whenever possible are core tenets of our development approach.

Perseids uses Uniform Resource Identifiers (URIs) to reference all documents and data objects subject to editing on the platform. These URIs have CTS and CITE URNs embedded in them to enable seamless integration with resources from the Perseus Digital Library.

The SoSOL application sits at the core of the Perseids platform. SoSOL is a Ruby on Rails application, originally developed by the Papyri.info project, that serves as front-end for a

Git repository of documents, metadata and annotations. It includes a workflow engine that enables documents and data of different types to pass through flexible review and approval processes. The SoSOL application includes user interfaces for editing XML documents, metadata and annotations. While it does not include a full-featured XML editor, it supports alternative text based input of XML markup, and can enforce XML schema validation rules on the documents being edited. SoSOL currently supports working with documents that adhere to specific variations of TEI, including EpiDoc, and annotations that adhere to the Open Annotation specification.

Future plans for the Perseids platform involve extending the core SoSOL application with RESTful Application Programming Interfaces (APIs) so that users can work with annotations and documents in domain-specific tools and user interfaces of their choice. Editors have been identified for working with Optical Character Recognition (OCR) output, morpho-syntactic annotations, translation alignments and text reuse as the first to use these APIs but the goal is for any web based tool which supports RESTful interactions to be able to be used with the platform.

In addition to user-facing tools, the APIs will support integration of analytical services to populate annotations and documents with automatically-generated data that can then be manually corrected and curated by users of the platform. For example, users will be able to request an automated morphological analysis of a text and then refine the results to create a set of treebank annotations to accompany their text. Examples of other types of analytical services to be integrated include ones for identification of named entities (i.e. places and people) and text reuse.

Other enhancements focus on integration with the academic environment. The ability to login with institutional credentials via the Shibboleth/SAML 2 protocols allows a user's contributions to be linked to her identity at her educational institution. A dissemination module

will support publication of work product as part of an electronic portfolio. This module, together with the support for the CTS and CITE protocols, will also enable the development and distribution of dynamic syllabi - collections of reading lists, resources and assignments targeted to specific goals and/or proficiency levels.

A "deploy early and often" approach to development of the platform, making features on an incremental basis, is being taken. Students at the speaker's university will be using Perseids in the fall of 2013 to contribute essays as commentaries on texts and artifacts by linking from a dynamic display of the texts and objects being commented on, and to contribute transcriptions and translations of Greek funerary inscriptions. Through a pilot program at another university during the 2013-2014 academic year Latin students will be using the platform to correct automatically generated morphosyntactic analyses of texts of Caesar. Additional users and usage scenarios will be identified as we proceed through a series of scholarly visits and workshops.

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/llc/fqt046.