PROJECT bamboo
Building applications and shared infrastructure for humanities research

News • Spring 2012

Collaboratively Curate and Explore Texts

by Martin Mueller, Professor of English and Classics, Northwestern University

The scholarly explorer of our textual heritage in digital form is perforce chief cook and bottlewasher. S/he would rather cook than wash bottles, but s/he does not want to eat or drink from dirty plates or glasses. Unfortunately, there is nobody around to do the dishes. That pretty much sums up the situation of the humanities scholar working with most digital texts in their current form. The data are typically "dirty." Not perhaps so dirty as to make them useless, but dirty enough to make work with them uncomfortable. It is a major goal of Bamboo to find efficient and tolerable (if not pleasurable) ways of engaging scholars in the necessary tasks of cleaning up valuable textual data. Exploration and curation need to go hand in hand for quite a while before most of the texts can be trusted in the way in which we trust printed books from a reputable house. But there is some comfort in the fact that comparisons are lame: a textual error may be like a dirty bottle, but it needs to be "washed" only once.

Complete, Correct, and Connect

Bamboo is facilitating the collaborative curation of digital texts by building services and connecting tools. Basic curation tasks can be grouped under the imperatives of complete, correct, and connect. To promote these activities we have built relationships with the developers of curation tools such as TypeWright, Abbott, and MorphAdorner and will be working to integrate these and other tools into the ecosystem in phase two. We hope that the curated text will not only benefit other researchers but also be sent back to the original repository to improve the quality of the digital texts available to the public.

We are also incorporating tools for text analysis into the Bamboo ecosystem, thereby making hand-in-hand curation and exploration possible. The orchestration of tools into workflows will facilitate scholarly activities ranging from correction and annotation to analysis and visualization. The research environment will support the use of tools for collation, topic modeling, and visualization – including Juxta and software being developed by the Visualizing English Print project – and allow scholars to work collaboratively.

Project Bamboo is generously funded by the Andrew W. Mellon Foundation.

Scholarly Workflows

Imagine that a scholar wants to collect digital surrogates of Gothic novels for comparison and visualization. Once the digital objects have been gathered, the first step would be to use TypeWright to correct the OCR and create high-quality digital editions for analysis. Behind the scenes, the standardization of TEI could be accomplished by running the texts through Abbot, followed by linguistic annotation using MorphAdorner. Once these editions are prepared in such a way, tools from Visualizing English Print could provide visualization of the topics and linguistic features of the genre. Enabling these customizable workflows will be a big step forward in the use of digital tools for humanities research.

Bamboo DiRT Partners with DHCommons

Project Bamboo will soon launch Bamboo DiRT (dirt.projectbamboo.org), a directory of tools and collections for digital research and pedagogy that builds upon Lisa Spiro's well-known DiRT (digital research tools) wiki. The new Bamboo DiRT makes it easy to browse and search for tools by category, platform, license, or where the tool has been discussed (e.g. ProfHacker, DH Answers, etc.) Developers are encouraged to use Bamboo DiRT to make their work more visible to the community.

Bamboo DiRT
Digital Research Tools | Plant seeds. Grow ideas.

dh Commons

Bamboo DiRT is partnering with DHCommons, a project/collaborator matching hub, through a data-sharing arrangement that will benefit both sites. Basic information about the tools a project is using will be available on DHCommons and Bamboo DiRT will display a list of DHCommons projects that are using a particular tool on that tool's profile. We anticipate that the code developed for integration with DHCommons will be reusable on other directory-oriented sites.

Get Involved

- Visit Bamboo DiRT, **dirt.projectbamboo.org**, to explore digital research tools, or create an account and contribute to the site.
- Bamboo DiRT is currently seeking an editorial board to assist with data curation and shaping future site improvements. If you're interested, please contact the Bamboo DiRT team by emailing **bamboo-dirt@lists.projectbamboo.org**.
- Email **feedback@lists.projectbamboo.org** if you’re interested in working with Bamboo as an Affiliate.