

The Specimen Case and the Garden: Preserving Complex Digital Objects, Sustaining Digital Projects

Schlosser, Melanie

schlosser.40@osu.edu

University Libraries, The Ohio State University

Ulman, H. Lewis

ulman.1@osu.edu

Department of English, The Ohio State University

In a recent article entitled "Innkeeper at the Roach Motel", Dorothea Salo worries that focusing exclusively on preservation when designing institutional repositories leads to a situation in which documents are placed into repositories but never come out (Salo 2009). In this conceit, a "live" project gets placed in a repository and "dies" from lack of use. However, when attempting to preserve distributed, dynamic electronic textual editions, a somewhat different metaphor is needed. Like items in a specimen case, "live" digital projects must be "killed" before they are added to a conventional institutional repository such as DSpace. In such applications, they must be removed from the dynamic ecology of their production environments (the garden of our title) and frozen in a snapshot form that is substantially different in appearance and functionality. In our presentation, we will outline our own efforts to construct a preservation and sustainability plan for a multi-format, distributed, dynamic electronic textual edition that involves both the creation of a preservation "specimen" and the careful tending of the edition in the "garden". We will also share the general tools and workflows developed by the project that can help others tackle the same challenges.

Due to their innovative nature and the environments in which they are created, Digital Humanities projects are often dynamic and distributed. In other words, they often exist in multiple parts maintained on distributed hardware (which itself is supported by multiple

organizations), and are often compiled for viewing on the fly in response to readers' actions. The typical preservation strategy of frequent offsite backups is inadequate for these projects. The user or manager must be able to find the backups (including all of the constituent parts of a complex project, wherever they reside) and recognize what they are; the backups must be in usable condition; and their contents need to be understandable to the people who want to use them. Moreover, if backup files are to be used in any way similar to their original use, the files must be (1) compatible with current hardware and software, (2) translated into formats that are compatible with current hardware and software, or (3) used on reconstructed or emulated hardware and software that match the environment in which the project was originally developed.

Digital Humanities projects created by faculty often have the added vulnerability of relying on the creator's university computing accounts. Absent special accommodations, these accounts usually expire on a set schedule once the individual has moved on, taking with them information that often exists nowhere else.

1. Cultures of Preservation

In short, digital materials require a culture of description, preservation, and access every bit as robust as the practices and institutions that allow us to preserve manuscript and print materials. The devil of preservation — whether of print, digital, or other material artifacts — lies in the details of production, use, description, storage, conservation, and access. This holds true whether we are talking about acidic paper disintegrating on library shelves, digital files in obsolete formats, or media spread across computer systems whose links to one another have been broken. Preservation is further complicated by the distinction between preserving physical artifacts (books, manuscripts, floppy disks, flash drives) and preserving the information contained on those media in a useful format.

"The Specimen Case and the Garden: Preserving Complex Digital Objects, Sustaining Digital Projects" focuses on the preservation challenges posed by complex digital humanities projects, which present unique challenges to libraries

and repositories charged with accessioning, describing, and preserving the scholarly record. Our work, funded by the U. S. National Endowment for the Humanities, takes a two-pronged approach to the problem, developing technologies for preserving digital *objects* — and the relationships among them — that constitute complex projects, and establishing institutional structures for sustaining digital humanities *projects* after their creators are no longer actively involved in their development and maintenance. Over the course of more than a year, we have interviewed faculty involved in digital humanities projects, library professionals, and information technology professionals; assessed the need for new practices adapted to digital preservation at our institution; and documented the resources and workflows currently available for, or adaptable to, long-term preservation of digital objects. We have also begun to develop tools, institutional structures, and workflows for describing and archiving complex digital objects, as well as sustaining distributed digital production environments.

2. Preservation and Sustainability Tools and Workflows

Our presentation will outline the problems associated with preserving and sustaining complex digital projects, review the data we collected during our interviews, literature review, and environment scan, and share the tools that we have developed, including the following:

- a lifecycle map of complex digital projects that represents development and preservation milestones as interactions among scholars, library professionals, and IT professionals;
- a visual content manifest for complex digital projects that represents assets, the hardware on which those assets rely, and entities that enable the collaborative work of developing and preserving digital humanities projects;
- a Metadata Encoding and Transmission Standard (METS) profile for creating archival packages of complex digital projects;
- a visual representation of the roles of the scholars, library professionals, and IT professionals on our campus in the long-term preservation of digital humanities projects;
- a proposal for a Digital Humanities Network that could sustain selected distributed digital projects, without requiring that they sacrifice functionality for centralization.

Some of these tools and workflows will be easier to adapt to different projects, institutions, and cultural settings than others: for example, any library system should be able to adapt the METS profile to their needs, while our proposed Digital Humanities Network will serve mostly as a heuristic. Indeed, we hope to initiate a fruitful conversation about how to build cultures of preservation for complex digital projects among scholars, librarians, and IT professionals in a variety of institutional settings.

References

Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008). 'Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation'. <http://www.citeulike.org/user/AlisonBabeu/article/3801593>.

Cundiff, M. V. (2004). 'An Introduction to the Metadata Encoding and Transmission Standard (METS)'. *Library Hi Tech*. **22.1**: 52-64.

Dobрева, M. (2009). *Digitisation of Special Collections: Mapping, Assessment, Prioritisation, Final Report*. University of Strathclyde: Centre for Digital Library Research (CDLR).

Knight, G. (2009). *SHERPA Digital Preservation 2: Developing Services for Archiving and Preservation in a Distributed Environment, Final Report*. London: JISC; Centre for e-Research, King's College London.

The Library of Congress (2009). <http://www.loc.gov/standards/mets/mets-home.html>

Maron, N. L., Smith, K. K., and Loy, M. (2009). *Sustaining Digital Resources: An on-the-Ground View of Projects Today*. London: JISC: Ithaca Case Studies in Sustainability.

Rieger, O. Y., and Kehoe, B. (DATE). 'Enduring Access to Digitized Books:

Organizational and Technical Framework'. *The Fifth International Conference on Preservation of Digital Objects. Joined Up and Working: Tools and Methods for Digital Preservation*. Farqhar, Adams (ed.). London: The British Library.

Robertson, R. J., Mahey, M., and Allinson, J. (2008). *An Ecological Approach to Repository and Service Interactions*. Strathclyde, Scotland: JISC Centre for Educational Technology & Interoperability Standards.

Salo, D. (2008). 'Innkeeper at the Roach Motel'. *Library Trends*. 57: 2.

Skinner, K., and M. Halbert (eds) (2008). *Strategies for Sustaining Digital Libraries*. Atlanta, GA: Emory University.

Thompson, D. (2008). 'Archiving Web Resources'. *DCC Digital Curation Manual*. Ross, S. and Day, M. (ed.) .