Digital libraries in the humanities

Building a cyberinfrastructure for the humanities

Linda Cantara

The author

Linda Cantara is Metadata Librarian, Case Western Reserve University, Cleveland, Ohio, USA.

Keywords

Digital libraries, Society, Education

Abstract

This article discusses the American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences, and its potential impact on the development of enhanced digital libraries in the humanities. It focuses on the problems inherent in creating the technological environment needed by humanities scholars and the tools required to build, repurpose, and interact with digital libraries in the humanities.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1065-075X.htm

Teaching and research in the humanities is traditionally text- and image-based, focused on the artifacts that preserve cultural heritage and the documented records of human thought, action, and creativity. Typically collected, stored, and preserved in libraries, archives, and museums, primary resources – such as manuscripts, letters, and diaries; images of works of art; musical scores and recordings – are the substance of research in the humanities. Digital libraries now bring many of these documents and artifacts directly to researchers' desktops. For example, the Library of Congress's American Memory Historical Collections and the Oxford Digital Library (ODL) provide free access to high resolution images of manuscripts, letters, photographs, and sound recordings. Other sites, such as the Research Libraries Group (RLG) Cultural Materials and the Andrew W. Mellon Foundation's ArtStor, provide licensed access to comparable digitized primary materials and images of works of art. Open Archives Initiative metadata harvesters such as the University of Michigan's OAIster harvester facilitate federated searching of thousands of "hidden digital resources", while RLG's licensed-access Archival Resources provides finding aids to hundreds of special collections. In addition, digital libraries of XML-encoded text collections – notably the Oxford Text Archive (OTA), the University of Virginia's Electronic Text Center, and the Miguel de Cervantes Digital Library – provide in a variety of formats authentic electronic versions of printed books and editions[1].

Nevertheless, while the number and quality of primary sources available via digital library collections continues to grow, only a small fraction of the world's analog collections have been digitized. What is also missing are the tools and technologies humanities scholars need to visualize, analyze, interpret, collate, and edit texts and images in order to publish secondary monographs, scholarly editions, and teaching materials – both printed and electronic – based on these primary resources. Last spring, in response to *Revolutionizing Science and Engineering through Cyberinfrastructure: The Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* which outlined the technological requirements for new "knowledge environments and organizations" in the sciences and engineering (Atkins *et al.*, 2003), the American Council on Learned Societies (ACLS) established a national commission to investigate the cyberinfrastructure needs of scholars in the humanities and social sciences. Working with a definition of cyberinfrastructure as "the new research environments in which capabilities of the highest level of computing tools are available to

167

researchers in an interoperable network", the ACLS Commission notes: "Effective cyberinfrastructure for the humanities and social sciences will allow scholars to focus their intellectual and scholarly energies on the issues that engage them, and to be effective users of new media and new technologies, rather than having to invent them" (ACLS, 2004).

At a conference on "Transforming Disciplines: Computer Science and the Humanities", held at the National Academy of Sciences in January 2003, John Unsworth, chair of the ACLS Cyberinfrastructure commission, remarked that in spite of ten years of tool-building for humanities computing, very little progress has been made[2]. The problem, he states, is that: "We need (we still need) to demonstrate the usefulness of all the stuff we have digitized over the last decade and more – and usefulness not just in the form of increased access, but specifically, in what we can do with the stuff once we get it: what new questions we could ask, what old ones we could answer" (Unsworth, 2003). Unsworth notes that the tools that have been developed in the past decade have been specialized and self-contained, whereas what is needed is a standards-based toolkit that is open, modular, and extensible. Humanists, he states, need tools that facilitate more than just searching and browsing digital collections, tools that enable data-mining, annotation, comparison, and sampling, and an interoperable architecture that facilitates networked collaboration and sharing of data.

Thus far, the best-case scenarios are digital library collaborations that recognize that the humanities present computing problems that can engage the interests of computer scientists, and correspondingly, computer scientists can provide technical solutions to these problems, leading to the development of new tools and technologies for humanities scholars. One such endeavor is the ARCHway Project at the University of Kentucky, Lexington. Funded by a National Science Foundation Information Technology Research (ITR) grant, the ARCHway (Architecture for Research and in Computing for the Humanities through collaborative research, teaching, and learning) Project is the work of Kevin Kiernan (English), Jerzy Jaromczyk (Computer Science), Alexander Dekhtyar (Computer Science) and their students to apply computer science theory and expertise in the development of a technological infrastructure, an Edition Production Technology (EPT): "to identify and solve problems of mutual importance in building image-based electronic editions of significant cultural materials". The testbed comprises digital images of damaged Old English manuscripts from the Cotton Collection of the British Library. Built on Eclipse, an open source software development environment[3], the EPT will include a tool to integrate images with texts; an image annotation tool; a tool for layering digital images derived from multiple sources or captured with different techniques; a paleographical tool for organizing, analyzing, describing, and displaying scribal letterforms; a 3D imaging and image-flattening tool; a tool for creating XML-encoded glossaries; and an integrated XML editor, validator, and transformer. When completed, the EPT "Editor's Workbench" will be available for the use of scholars through an open source license (Kiernan et al., 2004).

A similarly ambitious initiative is HASTAC (pronounced "Haystack"), the Humanities, Arts, Science, and Technology Advanced Collaboratory, "a consortium of humanists, artists, scientists, and engineers ... dedicated to working together to develop innovative computing and information systems that support interdisciplinary research and teaching in the humanities and arts". Founded by Cathy Davidson, vice provost for Interdisciplinary Studies and co-founder of the John Hope Franklin Humanities Institute at Duke University, and David Theo Goldberg, director of the California Humanities Research Institute at the University of California, Irvine, HASTAC's primary goals are to create new knowledge in the humanities based on interdisciplinary scholarship; to develop new information technologies that will expand the study of humanities domains; to develop communication software that will facilitate multi-site, multi-user communication and data sharing; to educate a generation of scholars in the humanities and social sciences whose work, in collaboration with scientists and engineers, "is impacted by and has significant impact on information and computing systems"; and to transform institutions by establishing guidelines for the evaluation of new forms of collaborative, interdisciplinary scholarship. The deliverables of HASTAC-sponsored projects will undoubtedly contribute innovative and dynamic content to digital libraries in the humanities as well as the computing environment used to create them (Davidson et al., 2004).

The outcomes of the ARCHway and HASTAC projects (and others like them) will add to the considerable achievements of long-standing humanities computing research centers and projects, such as the University of Virginia's Institute for Advanced Technology in the Humanities (IATH), the University of Maryland's Institute for Technology in the Humanities (MITH), and Tufts University's Perseus Digital Library[4]. Meanwhile, the ACLS Commission on

Cyberinfrastructure for the Humanities and Social Sciences is convening public information-gathering sessions across the USA to gain input from experts in the fields of digital scholarship, teaching, librarianship, publishing, and administration in preparation for issuing a report in early 2005. Although the Commission's primary focus is the technological environment and tools required to reuse digital content in the research and teaching of humanists and social scientists, it is aware that new tools and technologies for users of the digital resources created by scholars are also required. The envisioned cyberinfrastructure, assuming it is successfully implemented, will result in significantly enhanced digital libraries in the humanities, digital libraries that not only provide online access to primary sources also but, tools to interact with and repurpose content as well as dynamic electronic scholarly editions, teaching materials, and resources for networked collaboration and sharing of data. Hopefully, in ten years' time it will not be necessary to again revisit the issues under review by the ACLS Commission.

Notes

1 Library of Congress, American Memory, http://memory.loc. gov; Oxford University Libraries Services (OULS), Oxford Digital Library, http://www.odl.ox.ac.uk/collections/index. html; Research Libraries Group (RLG) Cultural Materials, http://culturalmaterials.rlg.org/; Andrew W. Mellon Foundation, ARTstor, http://www.artstor.org/index.html; Open Archives Initiative, http://www.openarchives.org/; University of Michigan Digital Library Production Service (DLPS), OAIster, http://oaister.umdl.umich.edu/o/oaister/; Research Libraries Group (RLG) Archival Resources, http://www.rlg.org/en/page.php?Page_ID=120.

2 John Unsworth, former director of the University of Virginia's Institute for Advanced Technology in the Humanities (IATH), is dean of the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
3 Eclipse Project, http://www.eclipse.org/.
4 Institute for Advanced Technology in the Humanities (IATH), http://www.iath.virginia.edu/; Maryland Institute for Technology in the Humanities (MITH), http://www.mith.umd.edu/; Perseus Digital Library, http://www.perseus.tufts.edu/

References

American Council of Learned Societies (2004), "Commission on Cyberinfrastructure for the Humanities and Social Sciences", available at: www.acls.org/cyberinfrastructure/cyber.htm

Atkins, D.E., Droegemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, D.G., Messina, P., Ostriker, J.P. and Wright, M.H. (2003), *Revolutionizing Science and Engineering through Cyberinfrastructure: The Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation, available at: www.cise.nsf.gov/sci/reports/CH1.pdf

Davidson, C.N. *et al.* (2004), "The HASTAC vision: humanities, arts, science, and technology advanced collaboratory", available at: www.aas.duke.edu/comp/ascc/agendas/HASTAC_Vision_March_17.pdf

Kiernan, K., Jaromczyk, J. and Dekhtyar, A. (2004), "The ARCHway Project: architecture for research in computing for humanities through collaborative research, teaching and learning", available at: http://beowulf.engl.uky.edu/~kiernan/ARCHway/entrance.htm

Unsworth, J. (2003), *Tool-time, or 'Haven't We Been Here Already?', Transforming Disciplines: Computer Science and the Humanities*, National Academy of Sciences, available at: www.carnegie.rice.edu/ppt-unsworth.cfm