EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2009, Issue 1

**January 13, 2009**

# Cyberinfrastructure for the Humanities, Arts, and Social Sciences

Orville Vernon Burton, University of Illinois at Urbana-Champaign

Simon Appleford, University of Illinois at Urbana-Champaign

4772 Walnut Street, Suite 206 ◆ Boulder, Colorado 80301 ◆ www.educause.edu/ecar/

## Overview

In a world where information can be reduced to bits and bytes and communicated instantaneously, humanities and social science technology has rapidly emerged as a necessary and fundamentally interdisciplinary method of archiving, analyzing, and interpreting human activity and history. Humanities technology can, for the first time, securely preserve and provide broad democratic access to the documents, images, languages, sound, and film that constitute the human record and facilitate its understanding. Social science technology can now allow us to analyze, model, and even predict human social behavior on a scale that was unimaginable just a few years ago.

According to the 1965 National Foundation on the Arts and the Humanities Act, the humanities encompass the academic disciplines that study people: their ideas, history, literature, artifacts, and values. Until recently, humanities scholarship and sources—which comprise the disciplines of history, languages, letters; the humanistic social sciences such as anthropology; and such interdisciplinary fields as folklore, ethnic studies, and women's studies—have been accessible for the most part only in museums, classrooms, and libraries. One reason for this is that the manuscripts, visual art, sound, video, and printed text that form the basis of humanities research do not fit neatly into a standard category of scientific method. Another reason is that these diverse fields encounter many research and technological barriers to accessibility. These barriers include limits on storage and bandwidth, as well as incompatible technologies used at different institutions to create their own digital humanities projects. Now, technologies such as pattern recognition and extensible markup language (XML) are helping to finally break down these barriers.

This research bulletin describes a framework for cyberinfrastructure for the humanities, arts, and social science (HASS) that is rooted in the history and challenges of those disciplines, as well as the working practices of faculty scholarship. It also describes the work of The Institute for Computing in Humanities, Arts, and Social Science (I-CHASS) at the University of Illinois at Urbana-Champaign, which serves the national research and education communities by making tools available for high-performance computing, communication and collaboration, data collection and analysis, geospatial inquiry, and visualization.

## Highlights of Cyberinfrastructure for the HASS Disciplines

Quantitative social science scholars have long used mainframes and personal computers for statistical analysis and other types of data processing, but qualitative social scientists have largely been subject to the technological limitations shared by humanists. This situation is rapidly changing. Huge data sets comprising input from large, multi-year surveys or the census, e-mail archives, large samples of Internet sites over time, and video recordings require advanced computing technologies and data management. New methods of analysis such as visualization and data-, text-, and image-mining are required to make sense of these complex data sets and to supplement

more traditional statistical analysis. Simulation and mathematical modeling, becoming more widely used, often draw on data sets for parameters or starting conditions and often require advanced computing applications, supercomputing, or grid computing support. Quantitative social science also has made important contributions to public policy deliberations and to education via tools that enable laypersons to interact with social science data and models.

Although computing and the information sciences have primarily concerned themselves with science, engineering, and business applications, trends in computing and communications illuminate new opportunities associated with electronic "content": visualization of all kinds of information, digitization of music and other audio material, electronic publishing with increasingly elaborate formats, and more complex paradigms for finding, selecting, displaying, and using greater quantities and varieties of information. In addition, humanities and social science technology addresses the fundamental paradox of language and technology: the need to preserve vernacular and traditional modes of expression while providing methods of universal translation and communication.

## The Importance of Digital Technologies

In a remarkably short period of time, digital technology has become essential to our intellectual processes, every bit as important as writing, and if humanists and social scientists do not embrace and study its potential they will not have access to a complete scholarly and pedagogical arsenal. This marvelously protean technology, with a potential to revolutionize teaching, outreach, and research across the HASS disciplines, must be made available in its fullest for discovering, synthesizing, and sharing knowledge. In recognition of both this potential and the challenges inherent in addressing these growing technological needs, there is a vital need for the development of a national cyberinfrastructure for humanities and social sciences. Indeed, this was the key recommendation of the American Council of Learned Societies (ACLS) in its 2006 report, *Our Cultural Commonwealth*, which urged universities, funding agencies, and the federal government to invest in such a cyberinfrastructure "as a matter of strategic priority."[1]

The implications of this recommendation are startling, especially as the humanities, arts, and, to a somewhat lesser extent, the social sciences are notoriously underfunded, and they are coming under increasing financial pressures. The resources required to implement even a basic level of support for digital humanities and social science scholarship—demanding extensive hardware and software purchases, as well as the acquisition of significant technical expertise as a basic requirement—are clearly beyond the means of most campus units; indeed, even if it were feasible from a budgetary standpoint, this approach would only result in unnecessary duplication of resources. Creating an effective cyberinfrastructure is one method to avoid these problems. A key recommendation of the ACLS report was the establishment of a series of national centers at the university level dedicated to defining, implementing, and leading digital HASS research across discipline- and unit-based boundaries, while simultaneously participating in a healthy dialogue that contributes to and exploits cyberinfrastructure. These centers can act as hubs to provide stability within the community and allow long-term relationships between them and scholars across institutions.

## I-CHASS

One such center is I-CHASS at the University of Illinois at Urbana-Champaign. I-CHASS is envisioned as a nexus of scholarship, creativity, collaboration, outreach, and technical expertise—one hub among others in the growth of a vibrant community that spans both national and international collaborations and encompasses the HASS disciplines.

Advanced data acquisition, data storage and management, user-friendly data mining and visualization technologies, large-scale modeling and simulation, massive text and visual searches with complex relational analysis—these techniques, not possible a few years ago, are now galvanizing the humanities. But as humanities researchers harness these new capabilities, they must overcome major hurdles. Although new computational technologies give them access to a number of information sources and allow them to automatically generate data sets from those sources, researchers often struggle with the plethora of available information. Fortunately, the technologies that created the problem of having too much data can also help researchers gain insight and understanding from data resources. These new technologies are already available, and I-CHASS is a collaborative organization that introduces these technologies to humanities and social science users at all levels of computing expertise. I-CHASS provides scholars the resources, both human and computational, to enhance discovery and exploration, and it offers humanists access to hardware, computer applications, graphical user interfaces, and portals. Just as important, it offers educational opportunities and training to use these resources. For the most developed projects, I-CHASS helps scholars create their own cyberenvironments, an integration of distributed computing and data resources into end-to-end processes.

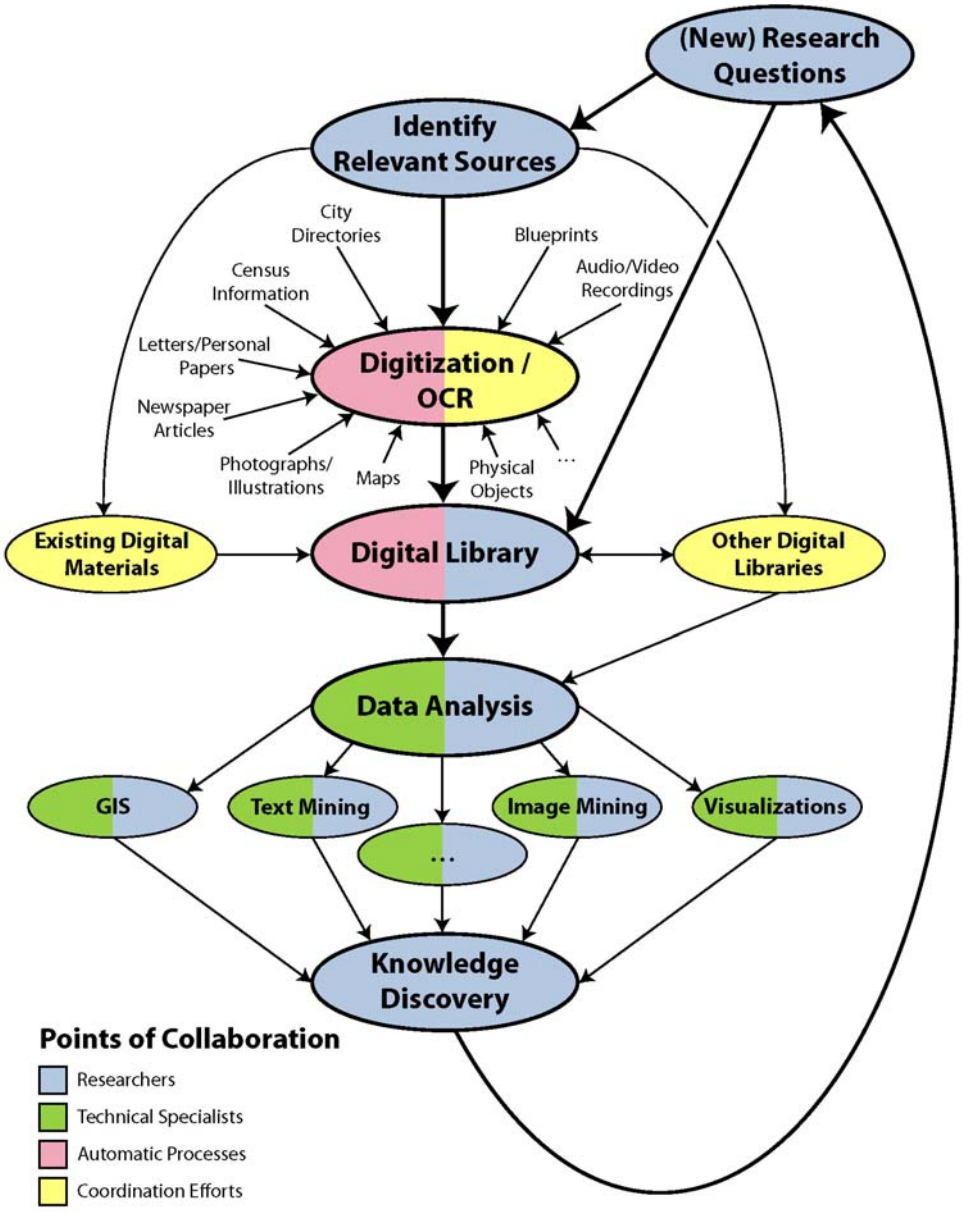
## Cyberenvironments

Figure 1 illustrates the structure of a typical humanities and social science cyberenvironment that combines disparate tools to facilitate new discoveries. It features:

- *Acquisition of print materials.* Automatic digitization of print materials and transmission of image files to the user's desktop.
- *Bulk upload and processing of digital materials.* Bulk upload of digital materials and integration with optical character recognition (OCR) technology. A user clicks a button on his or her desktop to upload scanned images to the cyberenvironment, which performs OCR, and stores original scanned imagery and OCR text in a searchable repository that is linked to supporting metadata.
- *Collaboration and dissemination.* Digital materials available to project collaborators via a secure Internal portal with features such as annotations and linking. The system allows individual documents or collections of documents to be accessible to the public via a customizable interface, with narratives or other related content.
- *Data and computational tools.* An array of relevant computational tools built into system. As an end-to-end portal, the cyberenvironment allows users to work with a data set and select computational tools for analysis and receive resulting output as a processed file.

- *Data conversion.* Files translated internally to a selection of common formats without any user intervention, including automatic geo-referencing of locations mentioned in text to allow spatial data analysis.
- *Grid-enabled resources.* Leveraging of grid resources, as needed, for both storage and computation needs.

**Figure 1. Key Components of a Cyberenvironment**



This type of complex interplay of processes is not possible using today's fragmented assortment of tools and data sources. Users must spend a great deal of time locating and compiling local stores of data and e-mailing files back and forth. Tools often require very specific file formats, and so data must constantly be converted between many formats, and users must be proficient with many different storage and computational tools to manage this process. The creation of cyberenvironments that feature end-to-end solutions addresses this entire information cycle by creating a transparent workflow that allows nontechnical users to create shared data warehouses and apply a variety of computational tools to their data without worrying about underlying technical issues. A further advantage of cyberenvironment development is that, where applicable, data can be distributed as online "cloud" computing applications. The advent of cloud computing allows researchers without knowledge of, or expertise with, the underlying technology infrastructure to take advantage of scalable, distributed computing resources.

## High-Performance Computing and Digitization

The next step in the digital humanities revolution, with the power to broaden humanities scholarship and as recently endorsed by the National Endowment for the Humanities, is high-performance computing (HPC). Just as traditional computing applications have expanded the type of questions that can be explored by humanists, access to the computational power afforded by the advent of advanced computer applications and HPC increases possibilities for researchers and educators in the humanities. The 2007 announcement of National Science Foundation funding for the National Center for Supercomputing Applications (NCSA) to build a petascale facility (dubbed "Blue Waters") at the University of Illinois positions I-CHASS to engage in digital humanities research and education at the highest end of computing. Humanists working with I-CHASS will have access to the Blue Waters supercomputing capacity and expertise that is not readily available to researchers elsewhere.

Research capabilities necessitate much more than digital libraries. If the major organizational principle is knowledge about how to use HPC, advanced applications, and networks, then supercomputing facilities become the logical partner. Google has collaborated with libraries, not supercomputing centers, to digitize books. Most definitely the digital humanities is interested in digital libraries, digital repositories, ontologies, metadata, and curation and has traditionally been the province of libraries and graduate schools of library sciences. For applications requiring storage, curation, preservation, and access to large data sets, the library or schools of library science makes sense as a partner for the scholar. I-CHASS, advanced computer applications, and HPC fill a different need.

## Collaborative Teams

Because scholars in humanities disciplines must be involved in the development of computing solutions for their community, I-CHASS forges teams of humanists and scientists for mutually beneficial exchanges of expertise. I-CHASS can substantially expand the number of researchers, educators, and specialists influenced by these collaborations. I-CHASS has proven that these different spheres are compatible because it has already been playing a critical role in helping humanities scholars

embrace new technologies and work in collaboration with computer scientists. Humanities researchers can come with their computational and storage challenges, receive advice and help, and have access to the most appropriate advanced computing applications and even HPC resources.

## What It Means to Higher Education

It is clear that the creation of a viable cyberinfrastructure for the humanities and social sciences requires the investment of significant resources at the university level if it is to succeed. I-CHASS is extremely fortunate in that its activities are supported by the University of Illinois at Urbana-Champaign, NCSA, and the Illinois Informatics Institute. The University of Illinois has a multitude of strengths in the humanities and social sciences as well as a significant library, computer science department, and a graduate school of library and information science. It is therefore logical that the university would seek to capitalize on these strengths. Moreover, Illinois' strategic vision for the 21st century endorses the ideals articulated in *Our Cultural Commonwealth* and lays the groundwork for much of I-CHASS's efforts to encourage the adoption of digital scholarship across the campus. The Illinois Informatics Institute was recently established to further that vision.

Although I-CHASS is fortunate to be co-located with a national supercomputing center, not all digital humanities centers will enjoy such an advantage. There are, nevertheless, other ways in which these centers can marshal the resources available on their own campuses to promote the use of digital tools in HASS research and teaching. Many important projects may require relatively modest computing resources, which are becoming cheaper by the year; storage, for example, is often more important than computing. It might be fruitful, for instance, for a center to act as a broker between humanities scholars and computer scientists to match research interests with on-campus computer resources. One strategy is to identify research questions that require a relatively modest commitment from the computing and information scientists to produce results—the proverbial low-hanging fruit. Such an early success could be used to promote the advantages of digital scholarship in numerous directions—to the HASS departments; to the computer scientists; and to the institution's administration. Another approach might be to designate several specific research areas as "digital hallmarks" and channel computational efforts in these areas until some successes have been achieved that can then be leveraged for broader institutional support. Since I-CHASS is committed to collaboration, advancing HASS cyberenvironments, and enabling scholars and other centers in their research, it regularly partners with other centers and individual scholars so that they can also benefit from its expertise and resources.

### Sharing the Vision

Communicating the cyberenvironment vision to the departments and engaging them in it is a challenge, not least because each department must be approached as a separate entity and each has its own priorities and interests in how or why to pursue digital scholarship. Furthermore, many HASS researchers remain uncertain as to exactly how computational techniques can enhance their scholarship, and they are, understandably,

reluctant to commit a significant amount of their time and energy to what appears to be an endeavor replete with risks and uncertain rewards. Without addressing these concerns, the efforts of digital humanities centers to establish a cyberinfrastructure are doomed to failure. Any effort to seed digital humanities centers consequently must focus considerable energy in building relationships, in fostering collaborations in specific disciplines, in disseminating as widely as possible the products of these collaborations, and in organizing informational meetings and themed workshops that focus on the application of specific technologies to research interests in HASS scholarship. It is also important that education on the use of digital technologies be available, including to students. All these activities are designed to alleviate fear and to demonstrate the profound power inherent in advanced computational methodologies for these disciplines. It is our experience that such outreach efforts are enthusiastically received, and they result in exciting new collaborations. This outreach effort, in research and in education, is the core mission of the Illinois Informatics Institute, which now hosts I-CHASS.

## Core Funding

Marshaling resources for building a cyberinfrastructure is, of course, a critical endeavor. Some core funding must, of necessity, be provided through a center's home institution. One model is for the home institution to provide start-up support, with the understanding that the commitment will be reduced as external funding is acquired. Winning substantial awards from grant-giving bodies has, however, been a difficult proposition for many humanities and social science departments, especially when compared with the funding patterns of disciplines in the physical and natural sciences, engineering, and medicine. Indeed, a typical humanities or social science department is unlikely to receive significant income from grant awards, with graduate students paid through state allocations and many scholars fortunate to receive even teaching released time in support of their research. This is a model that is clearly unsustainable for work in the digital realm, with its challenge to the paradigm of scholars working largely in isolation. Fortunately, funding agencies and private foundations have shown considerable interest in investing money in digital scholarship; among recent developments are initiatives from the National Endowment for the Humanities and the MacArthur Foundation, while the Mellon Foundation has long been a key backer of digital humanities and social science research. The potential funding available can be maneuvered for use by individual scholars on their own research projects and by institutions seeking to establish and expand national centers such as I-CHASS.

## Assessment

Technology must not widen the gap between those who have access to new information and tools and those who do not. We in the academy must not allow it to happen and must remain ever vigilant in that regard. Faculty must insist upon input in the growth of information technology. The 21st century demands a much broader approach to leaning, and new technology can help us meet that demand. But technology has the potential to become a tool of the elite. Our goal must be to democratize education. Especially those of us fortunate enough to teach in research universities must commit ourselves to expanding the community of learning.

The promise of computer-based instructional materials has not been realized. Advertisements in education periodicals, websites devoted to education, and administrative policy on many campuses suggest that computers and the Internet improve education by providing tools, resources, and online teaching and learning. These are all, from the faculty member's point of view, tools analogous to pen and paper, chalk and chalkboard—they are not analogous to the analytical and interpretive works from which our students should be learning. The computer-based instructional materials we need are similar to the research-based articles and books we write, not to the instruments we use to produce them. Only when faculty researchers and writers produce computer-based instructional materials—and discuss them and argue over them in professional venues—will they be more than syllabi marked up in Hypertext Markup Language (HTML), more than text with attractive illustrations and hyperlinks, more than links to others' websites.

## Education

For teachers who use computers and the Internet, the current model of education involves two elements. First is marking up traditional course material such as a syllabus and weekly readings in HTML and making them available online along with hyperlinks to websites deemed relevant by the instructor. Second is crafting an "online environment" with asynchronous discussion (bulletin boards, conferences, listservs, threaded discussions) and opportunities for private interactions between teacher and student, discussing work and posting grades, and among students responding to each other's work.

To facilitate teaching with computers, campuses offer workshops and consultations for faculty members who wish to avail themselves of the first two elements. Yet common institutional response to the computer revolution in higher education is actually hindering progress because it has led many professors to believe that the role for faculty in this process is in the adoption of tools made by others. The result has been that some individual professors lag far behind their students in the techniques used to analyze and synthesize information, and also that higher education itself is failing to keep up with the young people it should be serving. An analogous situation is not one in which students know how to use a web browser while their teachers do not, but one in which students have read Karl Marx when their teachers in historical studies have not or in which students have read T. S. Eliot when their teachers in literature have not. The problem is real; higher education in HASS disciplines is failing to participate in the digital revolution.

## The New Digital Scholarship

Humanists and social scientists often work by themselves, in isolation. The new digital scholarship requires a different model. Much of the best digital scholarship is done collaboratively, and that is likely to be the way of the future. Collaboration is characteristic of digital scholarship—a point made often in discussions of the impact of new media on the humanities. Because the problems one can encounter with networked computers are large and complex, tackling them often requires cooperation in substantial and coordinated ways. Until recently, most humanities and even social science research has been cooperative, in the sense that scholars disseminate

information in the form of books, articles, and papers—a knowledge flow that is one-way, not a dialogue. But with the advent of networked computing, a new system of collaborative work requires a diverse team to identify a common research theme or scientific problem and plan and carry out a coordinated effort to solve that problem or to create new knowledge. The humanities and social sciences are badly in need of models beyond the monograph and article for the demonstration of excellence, and the scholarship itself is in need of new genres and new strategies for reaching new audiences. The older generation has laid the foundation. We need a new generation of scholars to lead the humanities and social science disciplines into the future.

## Key Questions to Ask

- What support for humanities, arts, and social science research do we currently provide on our campus?
- In what ways do we support faculty who wish to obtain research support from centers or organizations external to our institution?
- To attract and retain the strongest faculty, what support for humanities, arts, and social science research should we provide on our campus?
- What measures do we use to assess the value of research projects in the humanities, arts, and social sciences?
- What is needed in the short run to address the gap between our current research support and our desired level of research support? What is needed in the long run?

## Where to Learn More


- Burton, Orville Vernon. “American Digital History,” *Social Science Computer Review* 23, no. 2 (Summer 2005): 206–220.
- Burton, Orville Vernon, ed. *Computing in the Social Sciences and Humanities*. Urbana: University of Illinois Press, 2002.
- Harrison, Teresa M., and Timothy Stephen, eds. *Computer Networking and Scholarly Communication in the Twenty-First Century*. Albany: State University of New York Press, 1996.
- Institute for Computing in Humanities, Arts, and Social Science. A partnership among the University of Illinois at Urbana-Champaign, the National Center for Supercomputing Applications, and the Illinois Informatics Initiative. http://www.chass.uiuc.edu/.
- McCarty, Willard. *Humanities Computing*. New York: Macmillan, 2005.

- Schreibman, Susan, Ray Siemens, Raymond George Siemens, and John Unsworth. *A Companion to Digital Humanities*. New York: Blackwell Publishing, 2004.
- "Summit on Digital Tools for the Humanities: Report on Summit Accomplishments." University of Virginia, September 2005. http://www.iath.virginia.edu/dtsummit/SummitText.pdf.
- Unsworth, John M. "What Is Humanities Computing and What Is Not?" http://computerphilologie.uni-muenchen.de/jg02/unsworth.html.


## Endnote

1. Welshons, Marlo, ed., *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences* (New York: American Council of Learned Societies, 2006), 10, http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf.

## About the Authors


*At the University of Illinois at Urbana-Champaign, Orville Vernon Burton (vburton@ncsa.uiuc.edu) is University Distinguished Teacher/Scholar; Professor of History, African American Studies, and Sociology; and Director of the Institute for Computing in Humanities, Arts, and Social Science (I-CHASS). He is also the Associate Director for Humanities and Social Science at the National Center for Supercomputing Applications. At I-CHASS, Simon Appleford (sapplefo@ncsa.uiuc.edu) is Project Manager.*


### Copyright



### Citation for This Work


Burton, Orville Vernon, and Simon Appleford. "Cyberinfrastructure for the Humanities, Arts, and Social Sciences" (Research Bulletin, Issue 1). Boulder, CO: EDUCAUSE Center for Applied Research, 2009, available from http://www.educause.edu/ecar.