

The Digital Future is Now:

A Call to Action for the Humanities¹

Christine L. Borgman
Professor & Presidential Chair in Information Studies UCLA
Borgman@gseis.ucla.edu

Article submitted to Digital Humanities Quarterly
Status: In review, September 16, 2009

Table of Contents

Abstract	2
Introduction	2
Problem Statement	3
Scholarly Information Infrastructure	4
Science [and, or, versus] the Humanities	5
Publication practices	6
Data in digital scholarship	8
What are data?	9
Data as evidence	9
Data sources	10
Research methods	12
Collaboration	14
Incentives to participate	15
Learning	17
Summary	19
A Call to Action	20
What are data?	20
What are the infrastructure requirements?	21
Where are the social studies of digital humanities?	21
What is the humanities laboratory of the 21 st century?	21
What is the value proposition for digital humanities in an era of declining budgets?	21
Acknowledgements	22
References	22

¹ Based on Keynote Presentation to Digital Humanities '09 Conference, College Park, MD, June 23, 2009

ABSTRACT

The digital humanities are at a critical moment in the transition from a specialty area to a full-fledged community with a common set of methods, sources of evidence, and infrastructure – all of which are necessary for achieving academic recognition. As budgets are slashed and marginal programs are eliminated in the current economic crisis, only the most articulate and productive will survive. Digital collections are proliferating, but most remain difficult to use, and digital scholarship remains a backwater in most humanities departments with respect to hiring, promotion, and teaching practices. Only the scholars themselves are in a position to move the field forward. Experiences of the sciences in their initiatives for cyberinfrastructure and eScience offer valuable lessons. Information- and data-intensive, distributed, collaborative, and multi-disciplinary research is now the norm in the sciences, while remaining experimental in the humanities. Discussed here are six factors for comparison, selected for their implications for the future of digital scholarship in the humanities: publication practices, data, research methods, collaboration, incentives, and learning. Drawing upon lessons gleaned from these comparisons, humanities scholars are “called to action” with five questions to address as a community: What are data? What are the infrastructure requirements? Where are the social studies of digital humanities? What is the humanities laboratory of the 21st century? What is the value proposition for digital humanities in an era of declining budgets?

INTRODUCTION

This is a pivotal moment for the digital humanities. The community has laid a foundation of research methods, theory, practice, and scholarly conferences and journals. Can we seize this moment to make digital scholarship a leading force in humanities research? Or will the community fall behind, not-quite-there, among the many victims of the massive restructuring of higher education in the current economic crisis? Much is at stake in the community’s ability to argue for the value of digital humanities scholarship and to assemble the necessary resources for the field to move from “emergent” to “established.”

The sciences, arts, and humanities have converged and diverged in various ways over the centuries. In the area of digital scholarship, many interests are in common across the disciplines. It is the pace of adoption that is divergent. The sciences, and to a lesser extent the social sciences, have been successful in developing the technical, social, and political infrastructure for digital scholarship under the rubrics of cyberinfrastructure – the term used in the U.S., and eScience – the term more widely used in the U.K. and elsewhere (U.K. Research Council e-Science Programme, 2009; Atkins et al., 2003). Digital humanities scholarship remains emergent, while eScience has become the norm in the sciences. The humanities need not emulate the sciences, but can learn useful lessons

by studying the successes (and limitations) of cyberinfrastructure and eScience initiatives.

While leaving definitions of “the humanities” to the reader, two complementary definitions of “digital humanities” provide a useful scope statement. Frischer’s definition (2009, p. 15) is “the application of information technology as an aid to fulfill the humanities’ basic tasks of preserving, reconstructing, transmitting, and interpreting the human record.” One resulting from the UCLA Mellon seminar claims that “Digital humanities is not a unified field but an array of convergent practices that explore a universe in which print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated” (Digital Humanities Manifesto, 2009). Taken together, the digital humanities is a new set of practices, using new sets of technologies, to address research problems of the discipline.

PROBLEM STATEMENT

Interest in the digital humanities has grown steadily for several decades. The Digital Humanities Conferences have occurred annually since 1989, sponsored by the Alliance of Digital Humanities Organizations. Constituent organizations of the Alliance have held conferences since 1973 (Alliance of Digital Humanities Organizations, 2009). MITH (Maryland Institute for Technology in the Humanities, 2009) celebrated its tenth anniversary, and IATH (Institute for Advanced Technology in the Humanities, 2009) at the University of Virginia its 17th anniversary. Academic research in the digital humanities at UCLA, Duke, Stanford, King’s College London, and elsewhere also appears to be thriving. Funding continues apace, with the Mellon Foundation, Council on Library and Information Resources, National Endowment for the Humanities, U.K. Arts and Humanities Research Council, and others focusing on infrastructure, tools, and services to support humanities scholarship in digital environments. Yet digital scholarship remains a backwater in much of the humanities. Concerns about publishing, tenure, and promotion for digital humanities scholars are a continuing theme in the conferences and in the literature of the field (Friedlander, 2008; 2009; Unsworth et al., 2006).

Despite many investments and years of development, basic infrastructure for the digital humanities is still lacking. Those who wish to gather and analyze digital data for humanities problems often find the overhead daunting, as exemplified by this emailed complaint from a history student in my scholarly communication course, who is pursuing a doctoral dissertation about the German enlightenment:

I’m finding that something as simple as constructing my maps of related concepts are not easily applied to primary sources in digital libraries. *So what use are the digital libraries, if all they do is put digitally unusable information on the web?* The digital libraries don’t offer a platform for

traditional note taking, much less for larger scale analysis, either quantitative or qualitative. (emphasis added; quoted with permission)

“Digital libraries,” the term used by my student, usually implies the existence of tools, services, and a library imprimatur of cataloging and curation. Her complaint is more about digital collections, which often lack basic capabilities for retrieval or analysis. This distinction is particularly relevant to the digital humanities. Content in digital collections may be “relatively raw;” others can add layers of interpretation, presentation, tools, and services, but these layers may be maintained separately from the content (Borgman, 1999; 2000; Lynch, 2002). The invisibility of essential infrastructure for digital scholarship in the humanities is one of many challenges to be addressed in growing the field. Until analytical tools and services are more sophisticated, robust, transparent, and easy to use for the motivated humanities researcher, it will be difficult to attract a broad base of interest within the humanities community.

Whose problem is it to improve the situation; that is, to design, develop, and deploy the scholarly infrastructure for digital humanities? As my UCLA colleague, Johanna Drucker, put it so well, “Leaving it to ‘them’ is unfair, wrongheaded, and irresponsible. Them is us.” (Drucker, 2009, p. B8). She believes that the digital humanities are at a “critical juncture,” and is concerned that her fellow scholars are deferring responsibility for action to librarians, computer scientists, technology developers, publishers, and others.

The operant terms in “digital humanities scholarship” are the latter two. Scholarly methods are as deeply seated in the humanities as they are in the sciences (Borgman, 2007). Only those who do the work and who require the infrastructure are in a position to take the field forward. Librarians and technology developers are essential partners, but those who conduct the research must take the lead.

This article, based on a keynote presentation to the most recent Digital Humanities Conference, reviews and reflects upon the differences between the approaches of the sciences and the humanities to digital scholarship (Borgman, 2009). First, I frame the notion of scholarly information infrastructure, then compare the approaches to digital scholarship of the sciences and the humanities. My analysis concludes with a call to action for the digital humanities community.

SCHOLARLY INFORMATION INFRASTRUCTURE

The term “scholarly information infrastructure” encompasses the technology, services, practices, and policy that support research in all disciplines. Cyberinfrastructure and eScience – both coined initially in reference to the sciences and technology, and both now used more broadly – refer to an infrastructure that enables forms of scholarship that are information- and data-intensive, distributed, collaborative, and multi-disciplinary. eResearch has become the collective term for variants such as eScience, eSocial Science,

and eHumanities (Borgman, 2007). The report of the *Commission on Cyberinfrastructure for the Humanities and Social Sciences* (Unsworth et al., 2006) was modeled on the strategy for science and technology (Atkins et al., 2003), while diverging to emphasize the humanities' motivations to make cultural heritage more widely available for teaching, research, and outreach. A similar argument is made by Todd Presner and Chris Johanson (2009) that digital humanities offers the opportunity to reconceptualize society as our cultural heritage migrates to digital formats, thus altering our relationship to knowledge and culture.

The technical and policy infrastructure for scholarship is being built rapidly, particularly for the sciences (*Cyberinfrastructure Vision for 21st Century Discovery*, 2007). Rare are the encompassing visions for scholarly infrastructure that originate in the humanities. Amy Friedlander (2009) provides a notable exception. She identified four research areas in digital scholarship where the interests of humanists, technology researchers, and others converge. These are scale, language and communication, space and time, and social networking. Issues of scale are of general interest because methods and problems must be approached much differently when one has, for example, the full text of a million books rather than a handful. Inspection is no longer feasible; only computational methods can examine corpora on that scale. Issues of language and communication, which are central to the humanities, are of broader interest for problems such as pattern detection and cross-language indexing and retrieval. Space and time encompass the new research methods possible with geographic information systems, geo-tagged documents and images, and the increased ability to make temporal comparisons. Social network analysis, long popular in sociology and bibliometrics, has become generalized to include patterns of social relationships in older texts or in online communication. Cross-cutting agendas such as these can be very influential in the design of an encompassing infrastructure. Humanities researchers need to be at the table as fundamental infrastructure decisions are being made.

SCIENCE [AND, OR, VERSUS] THE HUMANITIES

The humanities and the sciences each encompass broad swaths of scholarship, with much internal diversity. These two communities have significant commonalities, while differing in important ways. Identified here are six factors for comparison, selected for their implications for the future of digital scholarship in the humanities: publication practices, data, research methods, collaboration, incentives, and learning. The first five of these are drawn from longer analyses published elsewhere (Borgman, 2007); the last is drawn from the *NSF Task Force on Cyberlearning* (Borgman et al., 2008). The sequence of topics is cumulative to reflect how the boundaries are blurring between the sciences and the humanities.

Publication practices

Scholarly journal publication is shifting rapidly toward electronic formats, especially in the sciences. Some journals are dropping print publication altogether, others are declaring the online version (usually released several weeks to several months prior to the printed edition) to be the edition of record. Under pressure from authors, the majority of scholarly journals now appear to allow online posting of some form of pre-print or post-print (SHERPA/RoMEO: Publisher copyright policies & self-archiving, 2009).

For physics and related areas of computer science and mathematics, arXiv is the locus of scholarly communication. Monthly deposits of new papers now number more than 5,000; the site, which contains over 500,000 papers, typically receives 50,000 visits per hour (ArXiv.org e-Print archive, 2009). At least three iPhone applications are available for arXiv searching and retrieval. ArXiv, similar repositories in fields such as economics, and institutional repositories such as eScholarship – which serves all 10 campuses of the University of California – employ standard data structures that make their contents readily discoverable by search engines (Open Archives Initiative Protocol for Metadata Harvesting, 2009; Research Papers in Economics, 2009; University of California eScholarship Repository, 2009). It is little wonder that our science colleagues claim they never go to their campus libraries any more; their libraries come to them.

In the humanities, neither journal nor book publishing has moved rapidly toward online publication, despite pioneering efforts such as the 1990 launch of the *Journal of Post Modern Culture* as an electronic-only journal and the 2005 launch of *Vectors* as an online-only multi-media journal (Journal of Post Modern Culture, 2000; Vectors: Journal of Culture and Technology in a Dynamic Vernacular, 2009; Hamma, 2009; King et al., 2006; Whalen, 2009, forthcoming). A few of the established humanities journals have begun online versions that take advantage of digital technologies. Beginning in March, 2010, for example, the *JSAH* (Journal of the Society of Architectural Historians, 2009) will publish an online version that will support “zoomable images, video, GIS map integration, Adobe Flash VR, 3-D models, and online reference linking” – while continuing to publish its static print version.

The reasons for the slow adoption of digital publishing in the humanities are many, from not trusting online dissemination to a general reluctance to experiment with new technologies, even those well proven – “professionally indisposed to change” as Ken Hamma (2009) puts it. Monographic publishing, which is core to humanities scholarship, has begun a seismic shift toward digital publishing (Jaschik, 2008; 2009; Poe, 2001; Willinsky, 2006; 2009). A growing number of university presses are offering online access to monographs they publish in print, whether or not they also offer digital-only or print-on-demand formats. Other university presses are reinventing themselves in digital form (Rice University Press Mission Statement, 2008). The University of California Press recently announced a partnership with the California Digital Library, which hosts the university’s institutional repository, to offer “a suite of publishing services robust and flexible enough to support the complexities of content, format, and dissemination that

increasingly define scholarly communications” (University of California Publishing Services, 2009).

The “love affair with print” (Whalen, 2009, forthcoming) of art historians and other humanities scholars places not only “traditional” humanities scholarship at risk but also that of digital humanities. The distinction between print and digital publication is as much about epistemology as genre. Digital publishing is not simply repackaging a book or article as a computer file, although even a searchable pdf has advantages over paper. By incorporating dynamic multi-media or hypermedia, digital publishing offers different ways of expressing ideas and of presenting evidence for those ideas (Lynch, 2002; Presner, 2009, forthcoming; Presner & Johanson, 2009). When digital scholarship is published in print venues, much of its sophistication is lost.

Digital publishing differs from print publishing in several ways. One is the shorter time from submission to publication. While speed of publication is a much greater concern in the sciences than in the humanities, much of that time delay involves the physical production of the journal or book. Reviewing time varies little between print and digital formats. The humanities could benefit from faster turnaround, reaching audiences much sooner.

A second advantage of digital publishing – even more critical – is the larger audience for online publications. Anyone with an online connection and a subscription (in the case of fee-paid content), anywhere in the world, can read digital publications. Only those with access to a physical copy can read print-only publications. The number of titles and the number of copies of scholarly books and journals published in print form are decreasing rapidly, thus limiting both publishing outlets and readership. Maureen Whalen’s (2009, forthcoming) concern for art history, with its continuing reliance on print publishing, is that “the voices of authority ... will be talking amongst themselves.”

Two other consequences of the inexorable shift toward digital publication should be of concern to the humanities. One is that print material – including older material – becomes “widowed” as students and scholars alike search only online. The widowing problem was recognized early in the days of online catalogs, and was a major impetus for research libraries to digitize their entire back catalogs rather than only records of new material (Borgman, 2000; Lynch, 2003; Lynch & Garcia-Molina, 1995).

The other consequence is that easier access to online material frequently increases its rate of citation. Articles published in open access journals, open repositories, or dual-published by providing preprints or postprints online, tend to have a citation advantage over articles published only in closed-access journals, whether print or online. The degree of advantage varies by field and by a number of other factors, including how “open access” is defined (The Facts about Open Access: A Study of the Financial and Non-Financial Effects of Alternative Business Models on Scholarly Journals, 2005; Directory of Open Access Repositories, 2008; Directory of Open Access Journals, 2009; Open Content Alliance, 2009; The effect of open access and downloads ('hits') on citation impact: a bibliography of studies, 2009; Bailey, 2005).

While the details of these studies are much contested between authors, editors, librarians, and publishers, the simple tautology that easier discovery is associated with higher citation is difficult to dispute. As do authors in other fields, scholars in the humanities desire recognition in the form of citations to their work. Universities consider citation metrics in hiring and promotion decisions, despite known problems in their use for evaluating scholarly productivity (Bollen & Van de Sompel, 2008; Monastersky, 2005; Reedijk & Moed, 2008).

In sum, the sciences have benefited from online publication in ways that the humanities have not (yet). Digital publication is faster, reaches a wider audience, and tends to increase the citation rate over print-only publication. As the proportion of print-only publication continues to decrease, those for whom it is their only venue risk reaching an ever smaller and more closed community with their scholarship. Curation of digital objects is a concern in all fields, and is a topic that has the attention of management in libraries and archives. Nonetheless, digital publication has become the norm, and until the humanities community accepts that fact they will become increasingly out of the academic mainstream.

Data in digital scholarship

Central to the notion of cyberinfrastructure and eScience is that “data” have become essential scholarly objects to be captured, mined, used, and reused. This trend has been under way in science for many years, to varying degrees by field. As the technical and communications infrastructure became sufficiently robust to support large-scale data analysis and exchange, data became more valuable commodities. The availability of large volumes of data has enabled scientists to ask new questions, in new ways. Environmental scientists can conduct longitudinal analyses and make comparisons between locales using datasets compiled from multiple sources. Similarly, genome data offer analytical power at much finer granularity, and at larger scales.

While “data” is less familiar terminology in the humanities, the availability of large text, image, audio, and multi-media corpora has a similar result, enabling scholars in multiple fields to interrogate sources in new ways (Crane, Babeu & Bamman, 2007). Judging by presentations at the 2009 Digital Humanities Conference, *data* is becoming a popular term, whether framed in terms of “mining” or “cultural analytics.” Data mining “is the process of identifying patterns in large sets of data . . . to uncover previously unknown, useful knowledge.” (National Centre for Text Mining, 2009). Cultural analytics is a term that arose in the humanities as an analog to “visual analytics,” “business analytics,” and “web analytics,” and includes the use of “computer-based techniques for quantitative analysis and interactive visualization” to identify patterns in large cultural data sets (Manovich, 2009).

What are data?

The increasing value of data begs the question of “what are data?” Definitions associated with archival information systems offer a useful starting point: “A reinterpretation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen” (*Reference Model for an Open Archival Information System*, 2002, 1-9).

Another way to think about data is by origin. In the context of cyberinfrastructure, the four categories of data identified in an influential U.S. policy report (*Long-Lived Digital Data Collections*, 2005), and incorporated in National Science Foundation strategy (*Cyberinfrastructure Vision for 21st Century Discovery*, 2007), are now widely accepted. *Observational* data include weather measurements and attitude surveys, either of which may be associated with specific places and times or may involve multiple places and times (e.g., cross-sectional, longitudinal studies). *Computational* data result from executing a computer model or simulation, whether for physics or cultural virtual reality. Replicating the model or simulation in the future may require extensive documentation of the hardware, software, and input data. In some cases, only the output of the model might be preserved. *Experimental* data include results from laboratory studies such as measurements of chemical reactions or from field experiments such as controlled behavioral studies. Whether sufficient data and documentation to reproduce the experiment are kept varies by the cost and reproducibility of the experiment. *Records* of government, business, and public and private life also yield useful data for scientific, social scientific, and humanistic research.

Data as evidence

The need to address categories and levels of data is a pragmatic concern for managing information. Yet data are often in the eye of the beholder. In Buckland’s terms, data are “alleged evidence” (Buckland, 1991; Edwards, Jackson, Bowker & Knobel, 2007). What counts as good data varies widely, as one person’s noise is often another person’s signal. Similarly, the choices of data depend heavily on the questions being asked (Scheiner, 2004).

Whether any given set of observation or records can be considered data depends on context, even in the sciences. In our research on science and technology researchers in the environmental sciences, we found differing views of data on concepts as basic as temperature. Some of the computer science and engineering researchers interviewed said roughly, “*temperature is temperature*,” whereas biologists gave much more nuanced descriptions of how temperature was measured: “*There are hundreds of ways to measure temperature. ‘The temperature is 98’ is low-value compared to, ‘the temperature of the surface, measured by the infrared thermopile, model number XYZ, is 98.’ That means it is measuring a proxy for a temperature, rather than being in contact with a probe, and it is*

measuring from a distance. The accuracy is plus or minus .05 of a degree. I [also] want to know that it was taken outside versus inside a controlled environment, how long it had been in place, and the last time it was calibrated, which might tell me whether it has drifted..." (Borgman, Wallis, Mayernik & Pepe, 2007). Thus these two groups of researchers, often working side-by-side in the field as collaborators, had very different perspectives on what were acceptable data for their evidentiary purposes.

Studies of scientific practice, such as our work in embedded sensor networks, is providing insights for the design of cyberinfrastructure and eScience. The social studies of science and technology is a large and burgeoning field, with multiple journals and book series, and a scholarly society established more than 40 years ago (Van House, 2004). No comparable body of research on scholarly practices in the humanities exists, with the exception of studies on information-seeking behavior (Anderson, 2004; Bates, 1996a; b; Bates, Wilde & Siegfried, 1993; 1995; Case, 2006; Siegfried, Bates & Wilde, 1993; Stone, 1982; Tibbo, 2003). Lacking an external perspective, humanities scholars need to be particularly attentive to unstated assumptions about their data, sources of evidence, and epistemology. We are only beginning to understand what constitute data in the humanities, let alone how data differ from scholar to scholar and from author to reader. As Allen Renear remarked, "in the humanities, one person's data is another's theory" (personal communication, June 22, 2009).

Data sources

The sciences and humanities differ greatly in their sources of data and the degree of control they have over those data (Borgman, 2007). Scientific data sources vary by discipline, as seen in these few examples:

- Medicine: x-rays
- Chemistry: protein structures
- Astronomy: spectral surveys
- Biology: specimens
- Physics: events, objects
- Ecology: weather, ground water, sensor readings, historical records

Scientists, generally speaking, use data that were created by and for scientific purposes. They usually generate their own data, as in field observations or laboratory studies, or may acquire data from collaborators or other scientists. They may also acquire data from repositories in their field or from government sites, such as records of rainfall or river flow. Scientific documentation such as laboratory and field notebooks is sometimes considered to be data and sometimes metadata.

The social sciences occupy the middle position between the sciences and humanities on a continuum of data sources and control. Those at the scientific end of the scale gather their own observations, whether opinion polls, surveys, interviews, or field studies; build models of human behavior; and conduct experiments in the laboratory or field. Other social scientists rely on records collected by others, such as economic indicators or demographic data from the census. Government and corporate records are often of

interest, as are the mass media. A number of important data repositories exist, especially for large social surveys (e.g., Survey Research Center, Institute for Social Research, 2009; Survey Research Center, UC-Berkeley, 2009; UK Data Archive, 2009).

The humanities and arts are the least likely of the disciplines to generate their own data in the forms of observations, models, or experiments. Humanities scholars rely most heavily on records, whether newspapers, photographs, letters, diaries, books, articles; records of birth, death, marriage; records found in churches, courts, schools, and colleges; or maps. Any record of human experience can be a data source to a humanities scholar. Many of those sources are public while others are private. Cultural records may be found in libraries, archives, museums, or government agencies, under a complex mix of access rules. Some records are embargoed for a century or more. Some may be viewable only on site, whether in print or digital form. Data sources for humanities scholarship are growing in number and in variety, especially as more records are digitized and made available to the public.

Lynch's (2002) dichotomy of raw material vs. interpretation has a number of implications for the digital humanities. Two are of concern here. One is that raw materials are more likely to be curated for the long term than are scholars' interpretations of those materials. It is the nature of the humanities that sources are reinterpreted continually; what is new is the necessity of making explicit decisions about what survives for migration to new systems and formats. Second is the implication for control of intellectual property. Generally speaking, humanities scholars have far less control over the intellectual property rights of their sources – these raw materials – than do scientists, whose data usually are original observations or specimens. Typically, scholars can read, view, and cite cultural records, but often need explicit permission to reproduce them – and frequently need to pay a fee, especially in the case of images, to include them in reports of their research.

Intellectual property constraints on publishing of digital humanities scholarship are much different than those that usually apply in other disciplines. Rights to reproduce material remain closely tied to a print model, specified by number of copies printed and by temporal rules on sale that are irrelevant to online publication. Even cultural institutions as sophisticated as the Getty Trust encounter structural barriers to online publication of humanities scholarship (Whalen, 2009, forthcoming). The policy shift toward data sharing, well under way in the sciences, generally presumes that those who produce the data have the authority to release or deposit them for reuse (Arzberger et al., 2004).

In sum, “what are data?” is an important question for the humanities. The answer will determine what data are produced, how they are captured, and how they are curated for reuse. Data sharing in the humanities is a complex set of issues – not that they are simple in the sciences – that must be addressed. The humanities community needs a critical mass of digital resources and needs common tools, services, and repositories if they are to move beyond “boutique projects” (Friedlander, 2009) to a solid foundation of theory and method.

Research methods

Questions of “what are data?” are inextricable from the choice of research method. Many of the sciences, especially those “big science” areas that require large scale instrumentation and produce vast volumes of data, are in transition to a data-driven paradigm (Bell, Hey & Szalay, 2009; Foster, 2009). As the analysis, modeling, and merging of data become more central to scientific research, partnerships between scientists and computer scientists are becoming the norm.

An important case example of the changing role of data in science is the Sloan Digital Sky Survey (*Sloan Digital Sky Survey*, 2006), begun in 1992 by Jim Gray, Alex Szalay, and others (Gray et al., 2005; Gray & Szalay, 2002; Szalay, 2008). It was the first major astronomical survey founded on the premise that the resulting data would be openly and freely available, both to the astronomy community and to the public at large. Not only did astronomers mine the Sloan datasets for research purposes – more than 1700 scholarly papers were published – but manifold more users of these data were students and amateur astronomers. Amateurs, whose backyard telescopes could never yield data of such quality, also made important discoveries.

The Sloan Digital Sky Survey is significant for its openness, research productivity, and community engagement, and because it instantiates the “value chain” of scholarship (Borgman, 2007). On the SDSS site, papers are linked to the datasets on which they are based and datasets are linked to papers about them. One can enter the chain from either point and follow the relationships. While the project has ceased collecting new observations, the Sloan data remain available for use and are a canonical experiment in curation of large-scale datasets (Choudhury et al., 2008; Choudhury & Stinson, 2007). Astronomers and computer scientists are now engaged in the next generation project, Panoramic Survey Telescope and Rapid Response System, which is yielding about twenty times as much data as Sloan (PAN-STARRS, 2009).

Humanities scholars are more likely to find their data sources in the library – their traditional laboratory – than in the skies. While the library continues to be more central to scholarship in the humanities than it is to other fields, the characteristics of that relationship are changing. The use of physical space and of library staff has changed radically in the last two decades, largely in response to flat or declining university library budgets. Campus libraries have been consolidated in efforts to minimize the number of public service points to be staffed. Books, journals, and other physical materials have been moved to remote facilities, paged from the stacks upon request. Professional librarians are a smaller proportion of library staffs, and their attention is devoted to collections and curation. Thus fewer librarians are available to assist users in discovering, interpreting, and using the physical materials that remain – a trend likely to accelerate in the current economic crisis of higher education. The sciences are placing less demand on the physical library, allowing university libraries to reconfigure their spaces to benefit faculty and students in the humanities. Prime floor space previously

devoted to card catalogs, journals, and book stacks is now available for groups to work together with physical and digital resources.

At most universities today, humanities scholars and students are the primary constituency for physical books, journals, and records. This community also makes the finest distinctions among editions, printings, and other variants – distinctions that are sometimes overlooked in the transition from print to digital form. For general reading, any edition may suffice, and some degradation in image quality may be an acceptable tradeoff for access to large corpora of books and journals. Scholars are much more dependent on metadata to identify and compare variants, and may require physical copies to examine characteristics of printing and paper, annotations, and other details.

Differences in the methods of using print and digital objects are being thrown into sharp relief by mass digitization projects, most recently by the intense public debate over Google's book-scanning project. Concerns include not only the quality of scanning and of metadata, but the possibility that libraries will discard physical copies of books for which scans are available (UC and the Google Book Settlement: Frequently Asked Questions, 2009; Duguid, 2007; Nunberg, 2009; Samuelson, 2009). Also lost in most of these discussions is the distinction between scanning for search and access purposes (the Google approach) and scanning for preservation purposes, which has higher standards for image quality and for metadata (NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials, 2002; Mass Digitization: Implications for Information Policy, 2006; Greenstein, Ivey, Kenney, Lavoie & Smith, 2004).

Digital humanities projects have yet to achieve the scale of data, audience, or participation as the Sloan Digital Sky Survey. However, several long-lived digital humanities projects have made important contributions to research methods and data quality. Perseus is usually considered the first digital library in the humanities, with planning begun in 1985 and services available by 1987 (Perseus Digital Library, 2009; Crane et al., 2001; Marchionini & Crane, 1994). The initial collections of Perseus cover the history, literature, and culture of the Greco-Roman world. They have since expanded into other areas, and conducted significant research on the classification, management, and use of visual and textual materials (Crane, 2006; Mahoney, 2002; Smith, Mahoney & Crane, 2002). Rome Reborn, begun in 1997 at UCLA, was first concerned with digital library problems such as metadata, organization of historical and architectural periods, and representing relationships between textual sources and visual models (Rome Reborn, 2009; Frischer, 2004; 2009). Now the system exists in multiple manifestations, supports three-dimensional "fly-throughs," audio typical of the time period (including spoken Latin), and gladiator fights in the amphitheater using the latest computer graphics technology. Perseus, Rome Reborn, and newer projects such as HyperCities integrate map layers from Google Earth and other sources, which broadens their scope, audience, and interoperability with other components of the scholarly information infrastructure (HyperCities, 2009; Presner, 2009, forthcoming).

In sum, choices of data sources, research methods, and research problems are inextricably linked. Research methods in the sciences and in the humanities are becoming more data-driven. The key to “better” data – that is, data suitable for curation, reuse, and sharing – is capturing data as cleanly as possible and as early as possible in its life cycle. Agreements about data sources, structures, and formats will further the development of information infrastructure for digital humanities scholarship.

Collaboration

The size of collaborations is increasing in all fields, as measured by the number of co-authors on papers, and at the fastest rate in the sciences (Cronin, 2005). In sciences that rely heavily on instrumentation, such as astronomical observatories and particle accelerators, collaborations are large, diverse, and essential. Sciences that are more inductive and are conducted in field settings, such as habitat biology, tend to work in smaller groups. Sciences of all sizes are grappling with data management issues, as data are the glue – and often the product – of collaboration.

As noted above, the new forms of scholarship characterized by eResearch are information- and data-intensive, distributed, collaborative, and multi-disciplinary. Collaborations, when effective, produce new knowledge that is greater than the sum of what the participating individuals could accomplish alone. In fields where collaboration is the norm, graduate students learn teamwork, whether in the laboratory, the field, or in group work on data collection and analysis. Science dissertations frequently are carved out of larger group projects, with the student identifying a research problem worthy of sustained investigation. Funding agencies in the sciences consider dissertations to be important products of awards to faculty investigators. Dissertations and theses are listed explicitly in National Science Foundation annual reports, for example.

While the digital humanities are increasingly collaborative, elsewhere in the humanities the image of the “lone scholar” spending months or years alone in dusty archives, followed years later by the completion of a dissertation or monograph, still obtains. Students often are discouraged from conducting dissertation research under a faculty grant. Instead, they are expected to spend yet more time identifying funding for solo research. When one is groomed to work alone and does so for the years required to complete the doctorate, collaborative practices do not come easily.

Friedlander (2009, p. 6) argues that for digital humanities to thrive, “one component must be a set of organizational topics and questions that do not bind research into legacy categories and do invite interesting collaborations that will allow for creative cross-fertilization of ideas and techniques and then spur new questions to be pursued by colleagues and students.” As she suggests, the digital humanities need to move beyond large numbers of small, uncoordinated projects. Collaborative projects attract more resources and more attention. If properly designed, they also may be more sustainable, creating platforms on which new projects can be constructed. The plethora of boutique digital humanities projects risks the same fate as most digital learning objects. While intended for general use, they lack a common technical platform, common data

structures, and means to aggregate or decompose modules to a useful level of granularity (Borgman et al., 2008).

An indicator of collaboration in the digital humanities community is the shift over the last two decades from a focus on the audience – those who might learn or appreciate the cultural content presented – to a focus on participation, in which scholars, students, and the public can contribute content or conduct their own investigations (Electronic Cultural Atlas Initiative, 2009; *Ivanhoe: a game of critical interpretation*, 2009; Tibetan and Himalayan Library, 2009; Presner, 2009, forthcoming). The latter approach is more readily sustainable, as more people have vested interests in its capabilities and availability, and because it reflects current technical practices for Internet architecture (Architecture of the World Wide Web, 2004; Semantic Web Activity: W3C, 2009).

Scholarly collaboration is much studied but little understood. Among the predictors of success are the ability to achieve a common vocabulary and shared knowledge (Kanfer et al., 2000; Olson & Olson, 2000). The more disciplines involved, the more effort is required to achieve common ground. Investments must be made in learning enough about each other's disciplines that at least a pidgin language is established (Galison, 1997). Relationships take time, and must be nurtured. One important measure of success, and a worthwhile goal in eResearch, is that papers suitable for publication in each of the participating disciplines arise from a joint project.

In sum, the digital humanities community could benefit from more collaborative partnerships within the field and between the humanities and disciplines such as computer science. Collaboration requires investment in listening skills, always being alert to nuanced differences in assumptions, theories, definitions, and methods. Lessons and skills learned from these partnerships can enhance the scholarship of all participants. Common technology platforms also are important to achieve interoperability and sustainability, and can be leveraged as investments across projects.

Incentives to participate

Constructing a critical mass of data sources for scholarship in any field presumes that people will share the products of their research. Because data and collaboration are so central to the methods of digital scholarship, data sharing is an important indicator of success for eResearch, although practices are somewhat different in the sciences and in the humanities.

The public nature of scholarship has deep roots. Notions of “open science” date back at least to Francis Bacon, with scientific findings being accepted only after peer review. Scholars' incentives to share their results include recognition and acceptance of their work, which in turn drives hiring and promotion. In the sciences, authors may be required to release data as a condition of publishing the papers on which they are based. Funding agencies also are becoming more assertive about the release of data that result from

grants. However, publishing data is a far less mature practice than is publishing books and articles. Releasing a major dataset rarely brings as much recognition as releasing a major paper or book, but that balance is shifting, at least in the sciences (Borgman, 2007).

Scholars compete as well as collaborate, and thus have reasons *not* to share their data sources. The following are disincentives that apply to all disciplines, albeit to varying degrees (Borgman, 2007): (1) Faculty get more rewards for publishing papers and books than for releasing data; (2) the effort of individuals to document their data for use by others is much greater than the effort required to document them only for use by themselves and their research team; (3) data and sources offer a competitive advantage and are essential to establishing the priority of claims; and (4) data are often viewed as one's own intellectual property to be controlled, whether or not the data (or their sources) are legally owned. Means exist to address each of these concerns, but all are complex responses to a complex environment.

The first disincentive is the most universal across disciplines. The sciences and medicine are under the greatest pressure to release their data. In these disciplines the reward structure is adapting, and repositories and data structures exist. While humanities scholars are under less pressure to release their data and sources, they are contributing models, modules, and tools to participatory projects and shared collections.

Data documentation is an issue in all fields, but as the volume of data increases, consistent documentation becomes progressively more necessary. Once data are captured cleanly, sharing them later becomes less of a problem. Humanities scholars are acutely aware of the importance of metadata and finding aids in discovering sources. Metadata are equally important for data curation. Given this level of awareness and the trend toward more participatory projects, digital humanities scholars should be positioned well to address the documentation requirements necessary for the sustainability of their research products.

The third disincentive – competitive advantage – is often addressed in the sciences through embargoes, whereby the investigators have a set period of time (from a few months to a few years, depending on the field) after the end of a grant before being required to share their data. Embargoes serve two complementary purposes: they protect the scholars' control over data, and they ensure that others will have access to the data within a reasonable time period. In the humanities, scholars are similarly concerned about controlling access to the sources of their data, whether the Dead Sea Scrolls or a set of manuscripts in a university archive, until they have published their research. As data sources such as manuscripts and out-of-print books are digitized and made publicly available, individual scholars will be less able to hoard their sources. This effect of digitization on humanities scholarship has been little explored, but could be profound. Open access to sources promotes participation and collaboration, while the privacy rules of libraries and archives ensure that the identity of individuals using specific sources is

not revealed. Libraries and archives endeavor to maintain privacy in the use of digital as well as print sources. However, when digital content is controlled by commercial entities, protecting the privacy of users is a greater concern (Mass Digitization: Implications for Information Policy, 2006; Hoofnagle, 2009).

Intellectual property, the fourth disincentive to share data and sources, is the most intractable. The need to establish data sharing agreements in collaborative projects arose early in eScience initiatives and is far from resolved (David, 2003; David & Spence, 2003). In the case of the sciences, ownership – or at least control – usually can be clarified through negotiation. If the research depends upon material acquired from others, such as cell lines, rules on data release will be governed by contract. The reliance of humanities scholarship on cultural records, as discussed above, creates particularly complex intellectual property challenges in the sharing of data. For example, an art historian usually can publish his or her notes, but not the paintings on which the research is based. In the case of cultural models such as digital cities, it can be difficult to distinguish between data that represent an individual city and the model in which those data are incorporated. Difficulties in separating data from models (a problem in the sciences and in the humanities) plague both curation and data release efforts (HyperCities, 2009; Rome Reborn, 2009; Serving and Archiving Virtual Environments, 2009).

In sum, the digital humanities encounter most of the same incentives and disincentives for sharing data and sources faced by the sciences and by other disciplines. The details play out somewhat differently, of course. The need to build critical masses of cultural sources and interoperable technology platforms affirms the need to broker agreements about data. If the infrastructure for the digital humanities errs toward openness, as is the norm in much of the sciences, the field will advance more quickly.

Learning

The last comparison between the sciences and humanities, but by no means the least, is the role of information technology in learning. “Cyberlearning,” as argued by the National Science Foundation’s Task Force, can leverage the nation’s investment in cyberinfrastructure to benefit learning at all ages – “from K to grey” (Borgman et al., 2008). This argument was made earlier in the humanities, claiming that cyberinfrastructure could serve the humanities both for scholarship and for making cultural material more readily accessible for learning and outreach (Unsworth et al., 2006). Cyberlearning is defined as the use of networked computing and communications technologies to support learning. The scope of cyberlearning concerns in the Task Force report was necessarily constrained to the U.S. and to the domains funded by the NSF, which do not include the arts and humanities. However, the Task Force noted explicitly that the value of cyberlearning encompasses the sciences, social sciences, humanities, and arts, and is an important international initiative.

Several of the recommendations for advancing the state of cyberlearning have analogies for advancing the state of digital humanities. One is the need to build a vibrant field by promoting cross-disciplinary communities, publishing best practices, and recruiting diverse talents. The Cyberlearning Task Force made a careful distinction between cyberlearning as learning *with* distributed computing technologies and workforce development as teaching people *about* cyberinfrastructure. The latter is also a concern of the National Science Foundation (Cyberinfrastructure Vision for 21st Century Discovery, 2007). In the humanities as in the sciences, people need to learn *about* cyberinfrastructure before they can learn *with* it – or can use it for their research and teaching.

Another analogous recommendation from the cyberlearning report is the need to instill a “platform perspective.” As noted earlier, the takeup rate of digital learning modules has been limited by reliance on unique tools, proprietary software, and general lack of interoperability. Unless products are easily adapted to new uses, others have little incentive to invest in them. Both cyberinfrastructure and cyberlearning initiatives are constructing common technical platforms that will improve the sustainability and reuse of tools, services, and content. Some of these technical platforms can be leveraged for digital humanities scholarship. Where capabilities are lacking, the community can work in concert to construct them. Common platforms and standards are among the goals of the Mellon-funded Bamboo project, for example (Project Bamboo, 2009).

The Cyberlearning Task Force also recommended initiatives to enable students to use data. By embedding data skills early in the science curriculum – in the primary grades where feasible – students can learn to “think like scientists” early on. Hands-on science approaches endeavor to engage students in “real” science, making it more interesting and exciting than purely textbook approaches (Pea, Wulf, Elliott & Darling, 2003). Projects like the Sloan Digital Sky Survey and eBird encourage individuals to contribute their observations – whether about the sky or about birds in their backyards – for use in scientific investigations (Sloan Digital Sky Survey, 2006; eBird, 2009). The same promise applies to the humanities. If students can explore cultural records from the early grades and learn to construct their own narratives, they may find the study of humanities more lively. By the time they are college students, they will have learned methods of collaborative work and the use of distributed tools, sources, and services. Projects such as Perseus, HyperCities, and the Valley of the Shadow already enable students in humanities courses to engage in new forms of collaborative discourse (Perseus Digital Library, 2009; Ayres, 2004; Presner, 2009, forthcoming).

Lastly, the Task Force made a strong recommendation to the NSF to promote open educational resources. Educational content resulting from cyberlearning grants should be made available online with permission for unrestricted use and recombination. New proposals for research and development in cyberlearning should include plans to make their materials available and sustainable. These recommendations are relevant to all disciplines. Open educational resources are growing rapidly in variety and number (Atkins, Brown & Hammond, 2007; Baker, 2009; Thierstein, 2009). Licensing models

such as Creative Commons (Creative Commons, 2009) now include specific capabilities for licensing learning materials (ccLearn, 2009) and scientific data (Science Commons, 2009). Digital humanities projects, whether or not they include a learning component, also can benefit from Creative Commons licenses. The owners of intellectual property retain their copyright; they simply license it for reuse under publicly stated conditions. Intellectual property owned by others must not be appropriated, of course, but usually it can be linked if not specifically licensed.

Openness matters for the digital humanities for reasons of interoperability, discovery, usability, and reusability. Open resources – that is, those that can be used under license or are in the public domain – are more malleable for research and for learning. They can be mixed up and mashed up, and others can add value to them. Resources that are available via open repositories also are more readily discovered than those posted on local websites (OER Commons, 2009; Open Education, 2009; The Case of the Textbook: Open or Closed?, 2009; Atkins et al., 2007).

In sum, cyberlearning is important for the digital humanities for a number of reasons. One is the need to learn how to use and how to evaluate digital cultural materials early; graduate school is rather late. Second is the need to build common technology platforms for digital humanities scholarship, which will advance the field by leveraging efforts and resources and by increasing interoperability. Third is the value of open access to resources, which then become more malleable for research and for learning. Last is the need to build a strong community of digital humanities scholars, one that represents a much larger portion of the humanities than is the case today.

SUMMARY

My student's complaint, "So what use are the digital libraries, if all they do is put digitally unusable information on the web?" nicely captures the challenges facing the digital humanities today. Digital content, tools, and services all exist, but they are not necessarily useful or usable. Much work remains to build the scholarly infrastructure necessary for digital scholarship to become mainstream in the humanities. Humanities scholars must lead the effort, because only they understand the goals and requirements for what *should* be built. Librarians, archivists, programmers, and computer scientists will be important collaborators.

A number of developments in cyberinfrastructure, eScience, and eResearch offer guidance to the digital humanities community in the quest to become a more established field with a broader base of infrastructure. One is in the area of publication practices. The humanities lag in digital publication of journals and books. Digital publishing, while far from a panacea, offers a number of advantages in the speed, scope, and format of communication. Scholarly print publishing is on the decline, and those who publish only in print form risk being isolated, talking only to each other. More digital-only venues are needed, where dynamic and visual work can be published in its vernacular form.

Another area is the dissemination and use of data. The humanities community should continue to clarify their choices of data and data sources, for these will drive what content is produced, captured, managed, and available for reuse. Questions of data are closely related to research methods, which also are evolving. Data-driven research methods are most valuable when they enable scholars to ask new questions in new ways.

Collaboration is essential in digital humanities projects. Few individuals have the range of expertise required to execute these projects alone. Humanists should continue to seek out complementary partners and encourage people to listen and learn from each other. Working together is also more likely to lead to common platforms and other means of reducing the overhead of technical projects.

In both the sciences and the humanities, incentives to share one's writing are more obvious than are incentives to share one's data and sources. In the sciences, data release is being encouraged (or required) by journals and funding agencies, and data-driven research methods can draw upon large corpora that grow as new observations are contributed. In the humanities, data release is less of an issue, but the availability of common technical platforms, tools, and services will promote the sharing of data and sources. The disincentives to share are complex in both the sciences and the humanities, but are being addressed. As the sciences learn how to share data and to share credit for their findings, the humanities can build upon their best practices. Intellectual property constraints remain a major stumbling block, and the considerations vary between the sciences and the humanities.

Opportunities for using cyberinfrastructure for learning exist in all disciplines. Distributed access to scholarly content, common technical platforms, and open resources will advance the humanities as well as the sciences.

A CALL TO ACTION

In the process of developing the keynote presentation for the 2009 Digital Humanities Conference and in writing this paper, I consulted many individuals in the digital humanities community for their thoughts on the issues facing the field. From these discussions and my analyses above, five pressing problems emerged.

What are data?

What constitute data in the humanities? What are data sources? How are they made, shared, valued, used, and reused? Answering these questions will enable the digital humanities community to be more articulate about its scope and its goals, and better positioned to identify their requirements for infrastructure.

What are the infrastructure requirements?

The sciences have struggled with this question for a decade or two already. They have convened workshops and study panels, and launched funding initiatives addressed specifically at defining, designing, and deploying the necessary infrastructure for eScience. The humanities have tackled this question on a much smaller scale, leaving them in the position of building upon the infrastructure constructed by and for other disciplines. As Johanna Drucker (2009) put it so well, “them is us.” It is time for the community to articulate its own requirements and to act upon them.

Where are the social studies of digital humanities?

Why is no one following digital humanities scholars around to understand their practices, in the way that scientists have been studied for the last several decades? This body of research has informed the design of scholarly infrastructure for the sciences, and is a central component of cyberinfrastructure and eScience initiatives. Given how rapidly scholarship in the humanities is evolving, it is fertile ground for behavioral research. The humanities community should invite more social scientists as research partners and should make themselves available as objects of study. In doing so, the community can learn more about itself and apply the lessons to the design of tools, services, policies, and infrastructure.

What is the humanities laboratory of the 21st century?

This is a question of great concern to research libraries as well as to humanities scholars. The library continues to be a laboratory for the humanities, but not the only laboratory. Humanities scholars run computing laboratories and may work in distributed virtual environments for research and for learning. Humanists need to partner both with librarians and with the information technology planning and policy groups on their campuses. These communities urgently need to “think together” about the common challenges faced in a time of shrinking budgets for collections, physical space, staffing, and technology services.

What is the value proposition for digital humanities in an era of declining budgets?

For universities, the current economic recession is like no other. Public and private universities alike are re-examining core principles as budgets are slashed by 10% to 30% from one year to the next. Nothing is sacred, and “because it’s beautiful” is not a viable economic argument. The sciences have been remarkably effective at making the argument for their value in economic and political terms, whether to university administrations, legislatures, funding agencies, or the general public. While the humanities will have difficulty making parallel arguments in terms of economic competitiveness and medical advances, they have plenty to offer in terms of cultural

understanding, writing and design skills, and critical thinking. Digital scholarship also promotes technical skills, which can be highlighted.

Digital projects require resources in the form of computers, software, staff, and content. Non-digital scholarship also costs money, of course, but more often in the form of travel and subsistence expenses for research in remote archives. Tradeoffs in travel and digitization can be made more explicit. The number of people who will use and benefit from any given project also can be made clearer. Investments in common technical platforms and standards that leverage resources across larger numbers of people and projects are easier to justify.

The digital humanities community has produced some beautiful work and made many advances in technology, design, and standards. Now is the moment to consolidate that knowledge and to articulate the community's requirements and goals. Go forth and do great things...

ACKNOWLEDGEMENTS

I am grateful to the colleagues who provided thoughtful commentary on an earlier draft of this paper, including Murtha Baca, Gregory Britton, and Maureen Whalen of the Getty Trust; Johanna Drucker, Alberto Pepe, Todd Presner, and Katie Shilton of UCLA; Amy Friedlander, Council on Library and Information Resources; Bernard Frischer, University of Virginia; and Alexander Parker, Harvard University.

Many other people were very generous with their time in response to my inquiries about the past, present, and future of the digital humanities, including (in alphabetical order) William Dutton, Oxford Internet Institute; Neil Fraistat, University of Maryland; Richard Furuta, Texas A&M; Kimberly Garmoe, Anne Gilliland, UCLA; Charles Henry, Council on Library and Information Resources; Jason Hewitt, UCLA; Jieh Hsiang, National Taiwan University; Marina Jirotko, Oxford University; Matthew Kirschenbaum, University of Maryland; Clifford Lynch, Coalition for Networked Information; Lev Manovich, University of California, San Diego; Ann O'Brien, Loughborough University; Susan Parker, UCLA; Allen Renear, University of Illinois; David Robey, Oxford University; Ben Shneiderman, University of Maryland; Harold Short and Paul Spence, King's College, London; Joshua Sternfeld, UCLA; Sarah Thomas, Bodleian Library; Sharon Traweek, UCLA; Anne Trefethen, University of Oxford; John Unsworth, University of Illinois; Sarah Watstein and Robert Winter, UCLA.

REFERENCES

- Alliance of Digital Humanities Organizations. (2009). Retrieved from <http://www.digitalhumanities.org/> on 16 August 2009.
- Anderson, I. (2004). Are you being served? Historians and the search for primary sources. *Archivaria*, 58: 81-129.

- Architecture of the World Wide Web. (2004). Retrieved from <http://www.w3.org/TR/webarch/> on 2 June 2009.
- ArXiv.org e-Print archive. (2009). Retrieved from <http://arxiv.org/> on 12 August 2009.
- Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P. & Wouters, P. (2004). An International Framework to Promote Access to Data. *Science*, 303(5665): 1777-1778.
- Atkins, D. E., Brown, J. S. & Hammond, A. L. (2007). A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities. William and Flora Hewlett Foundation. Retrieved from <http://www.hewlett.org/oer> on 9 September 2009.
- Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messina, P., Messerschmitt, D. G., Ostriker, J. P. & Wright, M. H. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon panel on Cyberinfrastructure. National Science Foundation. Retrieved from <http://www.nsf.gov/cise/sci/reports/atkins.pdf> on 18 September 2006.
- Ayres, E. L. (2004). The Valley of the Shadow. University of Virginia. Retrieved from <http://valley.vcdh.virginia.edu/> on 28 September 2005.
- Bailey, C. (2005). Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals. Washington, DC: Association of Research Libraries. Retrieved from <http://info.lib.uh.edu/cwb/oab.pdf> on 28 September 2006.
- Baker, J. (2009). It Takes a Consortium to Support Open Textbooks. *EDUCAUSE Review*, 44(1): 30-33.
- Bates, M. J. (1996a). Document familiarity, relevance, and Bradford's law: The Getty online searching project report no 5. *Information Processing & Management*, 32(6): 697-707.
- Bates, M. J. (1996b). The Getty end-user online searching project in the humanities: Report No 6: Overview and conclusions. *College & Research Libraries*, 57(6): 514-523.
- Bates, M. J., Wilde, D. N. & Siegfried, S. L. (1993). An analysis of search terminology used by humanities scholars -- The Getty Online Searching Project Report No.1. *Library Quarterly*, 63(1): 1-39.
- Bates, M. J., Wilde, D. N. & Siegfried, S. L. (1995). Research practices of humanities scholars in an online environment - The Getty Online Searching Project Report No. 3. *Library & Information Science Research*, 17(1): 5-40.
- Bell, G., Hey, T. & Szalay, A. (2009). Beyond the data deluge. *Science*, 323: 1297-1298.
- Bollen, J. & Van de Sompel, H. (2008). Usage Impact Factor: the effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, 59(1): 136-149.
- Borgman, C. L. (1999). What are digital libraries? Competing visions. *Information Processing & Management*, 35(3): 227-243.
- Borgman, C. L. (2000). From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. Cambridge, MA: The MIT Press.
- Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.

- Borgman, C. L. (2009). Scholarship in the Digital Age: Blurring the Boundaries between the Sciences and the Humanities. In. Digital Humanities '09. College Park, MD, Maryland Institute for Technology in the Humanities. Retrieved from <http://works.bepress.com/borgman/216/> on 12 August 2009.
- Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., Lynch, C. A., Oblinger, D. G., Pea, R. D., Salen, K., Smith, M. S. & Szalay, A. (2008). Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge. A 21st Century Agenda for the National Science Foundation. Report of the NSF Task Force on Cyberlearning. Office of Cyberinfrastructure and Directorate for Education and Human Resources. National Science Foundation. Retrieved from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf08204 on 12 August 2008.
- Borgman, C. L., Wallis, J. C., Mayernik, M. S. & Pepe, A. (2007). Drowning in Data: Digital Library Architecture to Support Scientists' Use of Embedded Sensor Networks. Proceedings of the 7th Joint Conference on Digital Libraries, Vancouver, BC, Association for Computing Machinery. 269 - 277.
- Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5): 351-360.
- Case, D. O. (2006). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior* (2nd ed.). San Diego: Academic Press.
- ccLearn. (2009). A Project of Creative Commons. Retrieved from <http://learn.creativecommons.org/> on 6 August 2009.
- Choudhury, S., DiLauro, T., Szalay, A., Vishniac, E., Hanisch, R., Steffen, J., Milkey, R., Ehling, T. & Plante, R. (2008). Digital data preservation for scholarly publications in astronomy. *International Journal of Digital Curation*, 2(2): 20-30. Retrieved from <http://www.ijdc.net/index.php/ijdc/issue/view/3> on 17 August 2009.
- Choudhury, S. & Stinson, T. (2007). The Virtual Observatory and the Roman de la Rose: Unexpected Relationships and the Collaborative Imperative. *Academic Commons*. Retrieved from <http://www.academiccommons.org/commons/essay/VO-and-roman-de-la-rose-collaborative-imperative> on 22 July 2008.
- Crane, G. R. (2006). What do you do with a million books? *D-Lib Magazine*, 12(3). Retrieved from <http://www.dlib.org/dlib/march06/crane/03crane.html> on 17 August 2006.
- Crane, G. R., Babeu, A. & Bamman, D. (2007). eScience and the humanities. *International Journal on Digital Libraries*, 7(1-2): 117-122.
- Crane, G. R., Chavez, R. F., Mahoney, A., Milbank, T. L., Rydberg-Cox, J. A., Smith, D. A. & Wulfman, C. E. (2001). Drudgery and deep thought: Designing a digital library for the humanities. *Communications of the Association for Computing Machinery*, 44(5): 35-4018 April 2006.
- Creative Commons. (2009). Retrieved from <http://www.creativecommons.org> on 18 April 2009.
- Cronin, B. (2005). *The Hand of Science: Academic Writing and its Rewards*. Lanham, MD: Scarecrow Press.

- Cyberinfrastructure Vision for 21st Century Discovery (2007). National Science Foundation. Retrieved from <http://www.nsf.gov/pubs/2007/nsf0728/> on 17 July 2007.
- David, P. A. (2003). The economic logic of 'Open Science' and the balance between private property rights and the public domain in scientific data and information: A primer. In. The Role of the Public Domain in Scientific Data and Information. Washington, D.C., National Academy Press: 19-34. Retrieved from <http://siepr.stanford.edu/papers/pdf/02-30.html> on 30 September 2006.
- David, P. A. & Spence, M. (2003). Towards Institutional Infrastructures for E-Science: The Scope of the Challenge. Oxford Internet Institute Research Reports: University of Oxford. 92 Retrieved from <http://129.3.20.41/eps/le/papers/0502/0502002.pdf> on 30 September 2006.
- Digital Humanities Manifesto (2009). UCLA. Retrieved from <http://dev.cdh.ucla.edu/digitalhumanities/2008/12/15/digital-humanities-manifesto/> on 3 August 2009.
- Directory of Open Access Journals. (2009). Open Society Initiative, Scholarly Publishing and Academic Resources Coalition. Retrieved from <http://www.doaj.org> on 16 August 2009.
- Directory of Open Access Repositories. (2008). University of Nottingham, UK and University of Lund, Sweden. Retrieved from www.opendoar.org on 16 August 2009.
- Drucker, J. (2009). Blind Spots: Humanists must plan their digital future. Chronicle of Higher Education, 55(30): B6. Retrieved from <http://chronicle.com/free/v55/i30/30b00601.htm> on 25 June 2009.
- Duguid, P. (2007). Inheritance and loss? A brief survey of Google Books. First Monday, 12(8). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847> on 3 September 2009.
- eBird. (2009). Cornell Lab of Ornithology and Audobon Society. Retrieved from <http://ebird.org/content/ebird/> on 9 September 2009.
- Edwards, P. N., Jackson, S. J., Bowker, G. C. & Knobel, C. P. (2007). Understanding Infrastructure: Dynamics, Tensions, and Design. National Science Foundation: University of Michigan. NSF Grant 0630263. Retrieved from <http://hdl.handle.net/2027.42/49353> on 26 July 2007.
- Electronic Cultural Atlas Initiative. (2009). Retrieved from <http://www.ecai.org> on 15 September 2009.
- Friedlander, A. (2008). Head in the Clouds and Boots on the Ground: Science, Cyberinfrastructure and CLIR. Kanazawa Institute of Technology Library Roundtable, Kanazawa, Japan. Retrieved from <http://www.clir.org/pubs/resources/articles.html> on 15 August 2008.
- Friedlander, A. (2009). Asking Questions and Building a Research Agenda for Digital Scholarship. In. Working Together or Apart: Promoting the Next Generation of Digital Scholarship. Washington, DC, Council on Library and Information Resources. CLIR Publication No. 145: 1-15. Retrieved from <http://www.clir.org> on 15 June 2009.

- Frischer, B. (2004). Testimony to the Commission on Cyberinfrastructure for the Humanities and Social Sciences. American Council of Learned Societies. Retrieved from http://www.acls.org/cyberinfrastructure/cyber_meeting_notes_october.htm#frischer_summary on 6 August 2006.
- Frischer, B. (2009). Art and Science in the Age of Digital Reproduction: From Mimetic Representation to Interactive Virtual Reality. I Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación, Sevilla 17-20 de Junio de 2009.
- Galison, P. (1997). Image and Logic: A Material Culture of Microphysics. Chicago: University of Chicago Press.
- Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. & Heber, G. (2005). Scientific data management in the coming decade. CT Watch Quarterly, 1(1). Retrieved from <http://www.ctwatch.org/quarterly/articles/2005/02/scientific-data-management/> on 25 August 2006.
- Gray, J. & Szalay, A. (2002). The world-wide telescope. Communications of the ACM, 45(11): 51-55.
- Greenstein, D., Ivey, B., Kenney, A. R., Lavoie, B. & Smith, A. (2004). Access in the Future Tense. Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub126/contents.html> on 4 October 2006.
- Hamma, K. (2009). Professionally indisposed to change. EDUCAUSE Review, 44(2): 8-9.
- Hoofnagle, C. J. (2009). Beyond Google and Evil: How policy makers, journalists, and consumers should talk differently about Google and privacy. First Monday, 14(4). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2326/2156> on 15 September 2009.
- HyperCities. (2009). UCLA. Retrieved from <http://www.hypercities.com> on 17 August 2009.
- Institute for Advanced Technology in the Humanities. (2009). University of Virginia. Retrieved from <http://www.iath.virginia.edu> on 6 August 2009.
- Ivanhoe: a game of critical interpretation. (2009). Retrieved from <http://www.speculativecomputing.org/ivanhoe/index.html> on 3 September 2009.
- Jaschik, S. (2008). Publishing and values. Inside Higher Ed. Retrieved from <http://www.insidehighered.com/news/2007/08/22/anthro> on 10 March 2009.
- Jaschik, S. (2009). Farewell to the printed monograph. Inside Higher Ed. Retrieved from <http://www.insidehighered.com/news/2009/03/23/michigan> on 24 March 2009.
- Journal of Post Modern Culture. (2000). Retrieved from <http://pmc.iath.virginia.edu/> on 2 September 2009.
- Journal of the Society of Architectural Historians. (2009). Retrieved from <http://www.sah.org/index.php?src=gendocs&ref=JSAH&category=Publications> on 2 September 2009.
- Kanfer, A. G., Haythornthwaite, C., Bruce, B. C., Bowker, G. C., Burbules, N. C., Porac, J. F. & Wade, J. (2000). Modeling distributed knowledge processes in next generation multidisciplinary alliances. Information Systems Frontiers, 2(3-4): 317-331.

- King, C. J., Harley, D., Earl-Novell, S., Arter, J., Larence, S. & Perciali, I. (2006). Scholarly Communication: Academic Values and Sustainable Models. Andrew W. Mellon Foundation, Center for Studies in Higher Education, University of California, Berkeley. Retrieved from <http://cshe.berkeley.edu/publications/publications.php?id=230> on 28 July 2006.
- Long-Lived Digital Data Collections. (2005). National Science Board. Retrieved from <http://www.nsf.gov/pubs/2005/nsb0540/> on 18 April 2009.
- Lynch, C. A. (2002). Digital collections, digital libraries and the digitization of cultural heritage information. *First Monday*, 7(5). Retrieved from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/949/870> on 8 September 2009.
- Lynch, C. A. (2003). Life after graduation day: Beyond the academy's digital walls. *EDUCAUSE Review*, 38(5): 12-13.
- Lynch, C. A. & Garcia-Molina, H. (1995). Interoperability, scaling, and the digital libraries research agenda. IITA Digital Libraries Workshop. Retrieved from <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html> on 1 October 2006.
- Mahoney, A. (2002). Finding texts in Perseus. *New England Classical Journal*, 29(1): 32-34.
- Manovich, L. (2009). Cultural Analytics. Software Studies Initiative, University of California, San Diego. Retrieved from <http://lab.softwarestudies.com/2008/09/cultural-analytics.html> on 1 September 2009.
- Marchionini, G. & Crane, G. R. (1994). Evaluating hypermedia and learning: Methods and results from the Perseus project. *ACM Transactions on Information Systems*, 12(1): 5-34.
- Maryland Institute for Technology in the Humanities. (2009). University of Virginia. Retrieved from <http://mith.umd.edu/> on 6 August 2009.
- Mass Digitization: Implications for Information Policy (2006). Report from "Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects" -- Symposium held on March 10-11, 2006, University of Michigan: National Commission on Libraries and Information Science. ED495775. Retrieved from <http://www.eric.ed.gov/> and <http://www.lib.umich.edu/mdp/symposium/NCLIS-report.pdf> on 10 September 2009.
- Monastersky, R. (2005). The number that's devouring science. *Chronicle of Higher Education*, 52(8): A12-A17.
- National Centre for Text Mining. (2009). Retrieved from <http://www.nactem.ac.uk/> on 2 September 2009.
- NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials (2002). National Initiative for a Networked Cultural Heritage. Retrieved from <http://www.nyu.edu/its/humanities/ninchguide/> on 8 September 2009.
- Nunberg, G. (2009). Google's Book Search: A Disaster for Scholars. *Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Googles-Book-Search-A/48245/> on 3 September 2009.

- OER Commons. (2009). Retrieved from <http://www.oercommons.org/> on 18 April 2009.
- Olson, G. M. & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction*, 15(2-3): 139-178.
- Open Archives Initiative Protocol for Metadata Harvesting. (2009). Retrieved from <http://www.openarchives.org/pmh/> on 12 August 2009.
- Open Content Alliance. (2009). Retrieved from <http://www.opencontentalliance.org/> on 16 August 2009.
- Open Education. (2009). A Project of Creative Commons. Retrieved from http://opened.creativecommons.org/Main_Page on 6 August 2009.
- PAN-STARRS. (2009). Panoramic Survey Telescope & Rapid Response System. Retrieved from <http://pan-starrs.ifa.hawaii.edu/public/> on 14 September 2009.
- Pea, R., Wulf, W. A., Elliott, S. W. & Darling, M. A. (2003). *Planning for Two Transformations in Education and Learning Technology: Report of a Workshop*. Washington, D.C.: National Academies Press. Retrieved from <http://www.nap.edu/catalog/10789.html> on 8 September 2009.
- Perseus Digital Library. (2009). Tufts University. Retrieved from <http://www.perseus.tufts.edu/hopper/> on 3 September 2009.
- Poe, M. (2001). Note to self: Print monograph dead; invent new publishing model. *The Journal of Electronic Publishing*, 7(2). Retrieved from <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.204> on 16 August 2009.
- Presner, T. S. (2009, forthcoming). *HyperCities: Building a Web 2.0 learning platform*. In Natsina, A. & Tagialis, T. (Eds.). *Teaching Literature at a Distance*. London and New York, Continuum books.
- Presner, T. S. & Johanson, C. (2009). *The Promise of Digital Humanities: A White Paper*. 1-19. Retrieved from <http://www.digitalhumanities.ucla.edu/> on 12 August 2009.
- Project Bamboo. (2009). Retrieved from <http://projectbamboo.org/> on 15 September 2009.
- Reedijk, J. & Moed, H. (2008). Is the impact of journal impact factors decreasing? *Journal of Documentation*, 64(2): 183-192.
- Reference Model for an Open Archival Information System. (2002). Recommendation for Space Data System Standards, Consultative Committee for Space Data Systems Secretariat, Program Integration Division (Code M-3), National Aeronautics and Space Administration. Retrieved from <http://public.ccsds.org/publications/archive/650x0b1.pdf> on 4 October 2006.
- Research Papers in Economics. (2009). University of Connecticut. Retrieved from <http://www.repec.org/> on 12 August 2009.
- Rice University Press Mission Statement. (2008). Retrieved from <http://rup.rice.edu/about/mission> on 1 September 2009.
- Rome Reborn. (2009). Retrieved from <http://www.romereborn.virginia.edu/> on 17 August 2009.
- Samuelson, P. (2009). The Dead Souls of the Google Book Search Settlement. *Communications of the Association for Computing Machinery*, 52(7): 28-30.
- Scheiner, S. M. (2004). Experiments, observations, and other kinds of evidence. In Taper, M. L. & Lele, S. R. (Eds.). *The Nature of Scientific Evidence: Statistical*,

- Philosophical, and Empirical Considerations. Chicago, University of Chicago Press: 51-66.
- Science Commons. (2009). A Project of Creative Commons. Retrieved from <http://sciencecommons.org/about/index.html> on 6 August 2009.
- Semantic Web Activity: W3C. (2009). Retrieved from <http://www.w3.org/2001/sw/> on 14 April 2009.
- Serving and Archiving Virtual Environments. (2009). Serving and Archiving Virtual Environments. Retrieved from <http://www3.iath.virginia.edu/save/> on 13 August 2009.
- SHERPA/ROMEO: Publisher copyright policies & self-archiving. (2009). Retrieved from <http://www.sherpa.ac.uk/romeo/> on 12 August 2009.
- Siegfried, S. L., Bates, M. J. & Wilde, D. N. (1993). A profile of end-user searching behavior by humanities scholars - The Getty online searching project report no. 2. *Journal of The American Society for Information Science*, 44(5): 273-291.
- Sloan Digital Sky Survey. (2006). Retrieved from <http://www.sdss.org/> on 15 August 2006.
- Smith, D. A., Mahoney, A. & Crane, G. R. (2002). Integrating harvesting into digital library content. 2nd ACM IEEE-CS Joint Conference on Digital Libraries, Portland, OR, New York: ACM. 183-184. Retrieved from <http://www.perseus.tufts.edu/Articles/oaishort.pdf> on 4 October 2006.
- Stone, S. (1982). Humanities scholars: Information needs and uses. *Journal of Documentation*, 38(4): 292-313.
- Survey Research Center, Institute for Social Research. (2009). University of Michigan. Retrieved from <http://www.isr.umich.edu/src/> on 16 August 2009.
- Survey Research Center, UC-Berkeley. (2009). University of California, Berkeley. Retrieved from <http://srcweb.berkeley.edu/> on 16 August 2009.
- Szalay, A. (2008). Jim Gray, astronomer. *Communications of the ACM*, 51(11): 59-65.
- The Case of the Textbook: Open or Closed? (2009). *EDUCAUSE Review*, 44(1): 13.
- The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. (2009). The Open Citation Project - Reference Linking and Citation Analysis for Open Archives. Retrieved from <http://opcit.eprints.org/oacitation-biblio.html> on 16 August 2009.
- The Facts about Open Access: A Study of the Financial and Non-Financial Effects of Alternative Business Models on Scholarly Journals (2005). Kaufman-Wills Group LLC: Association of Learned and Professional Society Publishers. Retrieved from <http://sippi.aaas.org/Pubs/> on 27 September 2007.
- Thierstein, J. (2009). Education in the Digital Age. *EDUCAUSE Review*, 44(1): 33-34.
- Tibbo, H. R. (2003). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *The American Archivist*, 66(Spring-Summer): 9-50.
- Tibetan and Himalayan Library. (2009). University of Virginia. Retrieved from <http://www.thlib.org/index.php> on 1 September 2009.
- U.K. Research Council e-Science Programme. (2009). Retrieved from <http://www.rcuk.ac.uk/escience/default.htm> on 13 August 2009.

- UC and the Google Book Settlement: Frequently Asked Questions. (2009). University of California, Office of Scholarly Communication. Retrieved from <http://osc.universityofcalifornia.edu/google/faq.html> on 3 September 2009.
- UK Data Archive. (2009). Retrieved from <http://www.data-archive.ac.uk/about/about.asp> on 16 August 2009.
- University of California eScholarship Repository. (2009). Retrieved from <http://repositories.cdlib.org/escholarship/> on 16 April 2009.
- University of California Publishing Services. (2009). Retrieved from <http://www.ucpress.edu/pubservices/> on 2 September 2009.
- Unsworth, J., Courant, P., Fraser, S., Goodchild, M., Hedstrom, M., Henry, C., Kaufman, P. B., McGann, J., Rosenzweig, R. & Zuckerman, B. (2006). Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for Humanities and Social Sciences. American Council of Learned Societies. Retrieved from <http://www.acls.org/cyberinfrastructure/cyber.htm> on 17 July 2007.
- Van House, N. A. (2004). Science and technology studies and information studies. In Cronin, B. (Ed.). *Annual Review of Information Science and Technology*. Medford, NJ, Information Today. 38: 3-86.
- Vectors: Journal of Culture and Technology in a Dynamic Vernacular. (2009). Retrieved from <http://www.vectorsjournal.org/> on 8 September 2009.
- Whalen, M. (2009, forthcoming). What's wrong with this picture? An examination of art historians' attitudes about electronic publishing opportunities and the consequences of their continuing love affair with print. *Art Documentation*.
- Willinsky, J. (2006). *The Access Principle: The Case for Open Access to Research and Scholarship*. Cambridge, MA: MIT Press.
- Willinsky, J. (2009). Toward the Design of an Open Monograph Press. *Journal of Electronic Publishing*, 12(1). Retrieved from <http://dx.doi.org/10.3998/3336451.0012.103> on 1 April 2009.