

**Promoting Digital Scholarship:
Formulating Research Challenges
In the Humanities, Social Sciences and Computation**

A Workshop
Co-Sponsored with the National Endowment for the Humanities
By the
Council on Library and Information Resources

1752 N Street, NW
Suite 800
Washington, DC 20036

HC-50002-08

Amy Friedlander, Ph.D.
Project Director

Draft Final
To be submitted spring 2009

Table of Contents

1.0 Narrative Description: Introduction and Purpose	2
2.0 Project Activities	2
2.1 Steering Committee	2
2.2 White Papers and Web Site	3
2.3 Participants	3
2.4 Agenda	4
3.0 Description of the Symposium and Outcomes	4
3.1. Issues in Digital Humanities Research	4
3.1.1 Data and Collections	4
3.1.2 Services and Tools	5
3.1.3 Research and Analysis	6
3.1.4 Teaching and Communication	7
3.1.5 Credentialing and Cultures	8
3.2 Research Challenges and Opportunities	8
3.2.1 Research Topics	9
3.2.2 Context of Research	11
3.2.3 Specific Projects	12
4.0 Evaluation and Next Steps	13
Appendices	14

1.0 Narrative Description: Introduction and Purpose

The Council on Library and Information Resources (CLIR) in cooperation with the National Endowment for the Humanities (NEH) convened a group of leading scholars for a one-day workshop on September 15, 2008. The workshop had two goals:

- i. promote digital humanities by identifying a series of long-term research challenges at the intersection of the humanities, social sciences, and computation; and
- ii. document the findings of the workshop in parallel print and electronic publications to support subsequent discussion from the concerned communities.

Organizing this workshop responds to recommendations set forth by the American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences that call upon concerned organizations to exhibit leadership and advance digital scholarship, and is consistent with goals articulated by NEH and other public agencies. This effort has been funded through two cooperative agreements. The first (and the subject of this report) was directed toward organizing and convening the symposium; the second cooperative agreement (HC-50004-08) supported a publication that consists of the proceedings of the day, white papers that were commissioned to support the event, and an interpretive essay that contextualized the results of the meeting together with recommendations for future programs. The resulting publication broadly reflects many of the current issues in digital humanities scholarship.

2.0 Project Activities

This project comprised the following activities:

- Convening a steering committee
- Identifying authors of white papers that would provide context for the meeting and then commissioning papers from these individuals
- Identifying and inviting participants
- Building a Web site
- Developing an agenda for the one-day symposium
- Organizing the logistics for the symposium
- Running the meeting

These activities are described in the following sections. As previously noted, a publication has resulted from this meeting and is separately reported.

2.1 Steering Committee

The purpose of organizing the steering committee was to ensure that the content of the symposium would reach the appropriate audiences and that those invited to participate would reflect an appropriate balance among computer scientists, humanities researchers, researchers with a deep interest in digital humanities, librarians, archivists, and information scientists as well as attention to women and under-represented groups. The composition of the committee had been proposed in the original request to NEH and with one exception those individuals did agree to participate. Participants included representatives from the Institute of Museum and Library

Services (IMLS), NEH, the National Science Foundation (NSF), the Andrew W. Mellon Foundation, and the Coalition for Networked Information. A representative from the ACLS was invited but did not participate. Monthly meetings were held by telephone from February through June 2008 and supplemented by more frequent e-mail communications to update members of the committee on progress concerning specific tasks. Notes were circulated after each telephone conversation to document the results and decisions and to maintain momentum in the planning process.

The steering committee participated directly in identifying authors of white papers that would provide context for the meeting, identifying the participants, and developing the agenda. The members provided review and comment for the Web site. CLIR undertook all logistics associated with inviting participants, commissioning the papers, and building the Web site, which included an online forum for discussion among the invitees.

2.2 White Papers and Web Site

Six white papers were commissioned, posted to the Web site in advance of the meeting, and edited for inclusion in the publication resulting from this symposium. They included:

- Tools for Thinking: ePhilology and Cyberinfrastructure, by Gregory Crane, Alison Babeu, David Bamman, Lisa Cerrato, and Rashmi Singhal
- Social Attention in the Age of the Web, by Bernardo A. Huberman
- The Changing Landscape of American Studies in a Global Era, by Caroline Levander
- Art History and the New Media; Representation and the Production of Humanistic Knowledge, by Stephen Murray
- A Whirlwind Tour of Automated Language Processing for the Humanities and Social Sciences, by Douglas W. Oard
- Information Visualization: Challenge for the Humanities, by Maureen Stone

These papers, three by humanists and three by computer scientists, represent a cross-section of the domains and capabilities potentially relevant to computationally intensive scholarship in the humanities. Humanities research was represented by classics/philology, history of art and architecture, and Latin American studies. Computer science research was represented by language and speech, visualization, and social networking. Papers were intended to provide overviews of the key issues in the field, opportunities for collaboration, and issues for future research.

After the workshop, it was decided that a seventh essay, a review of digital humanities centers, would be useful. Diane Zorich was invited to prepare this essay based on a prior study that she had conducted for the Scholarly Communication Institute (July 2008), thus effectively leveraging prior work. The Web site was also edited for release to the public. In addition to editorial revisions, the forum, which had been set up to facilitate discussion within the participants, was disabled.

2.3 Participants

A total of 30 people accepted the invitation to participate in the conference as a member of the steering committee, writer, or participant; 25 actually attended. A list of invitees and their biographical information is included in Appendix 2 of this report. Invitees were selected to represent the range of humanistic inquiry and computer science research together with

representatives of the library community, which is entrusted with the management of the source material on which such scholarship depends. Thirteen are women, and seventeen are men. Every effort was made to include broad representation of demographic groups, and the resulting profile is reflective of the membership of the constituent disciplines and domains. The domains represented in the group included archaeology, history of art and architecture, classics, philology, Latin American studies, medieval studies, Romance languages and literature, Tibetan/Himalayan cultures, visual arts, philosophy, geography and geographic information systems, political science, artificial intelligence, speech and language recognition, human-computer interaction, library and information science, Korean and Old Norse languages and cultures, and visualization and visual literacy.

All participants were given a list of attendees that included full contact information for the others to facilitate ongoing communication. Telephone numbers and email addresses have not been included in the material posted to the Web site or in the appendix of this document to protect the privacy of the participants.

2.4 Agenda

The agenda is included in Appendix 1 of this report. The morning's activities were structured around discussion of the white papers. This was a mechanism for framing an otherwise broad discussion of issues, methods, and concerns in the humanities and the opportunities afforded by technology and the collaboration with researchers in advanced technology. The afternoon was organized around articulating specific questions and challenges.

3.0 Description of the Symposium and Outcomes

The content of the symposium is expressed in three sources: the white papers, lengthy and thoughtful posts to the online forum in advance of the meeting, and the rich discussion during the day. Collectively, the outcomes fall into two broad categories: general discussion of issues in digital humanities and articulation of research challenges and opportunities.

3.1. Issues in Digital Humanities Research

Humanists address a very broad range of topics and use diverse methods and sources. Thus, isolating issues that transcend disciplinary boundaries represents a challenge. Based on the discussions, trans-disciplinary issues in digital humanities research can be grouped into five major topics: data and collections, services and tools, research and analysis, teaching and communications, and credentialing and cultures.

3.1.1 Data and Collections

The first and essential consideration is management of collections for use by multiple audiences. Collection development entails capturing, archiving, and preserving digital materials. This includes converted materials, which may be both legacy and the product of ongoing format conversion projects, as well as so-called "born digital" materials—information that has been created in digital form, such as Web sites, instrumented data from sensors, satellites, and similar recordation and measurement devices, databases of observations and transactions, and so on.

On the one hand, scholars are facing what one participant called a “tsunami” of digital information, which is both voluminous and heterogeneous in format, genre, language, and sources. Such data is harder to work with in the sense that it may not be easily authenticated or sourced, requiring future scholars to develop ways of obtaining “the best” data and challenging scholars to learn to develop ways to obtain the appropriate quality. But notions of data scarcity or abundance are actually context dependent. The same participant who cited the data tsunami also acknowledged that there is a relative paucity of “suitably tagged” data for machine learning algorithms. Nor are there many humanities disciplines adequately equipped with test collections for machine translation, summarization, and other natural language processing technologies, all prerequisite to certain kinds of computational research using these materials. Her point was echoed by other domain specialists who noted the relative scarcity of suitably marked up digital texts in medieval literature, for example, compared with the number of existing analog texts of potential interest to scholars. Many such texts have not even been cataloged and are hence undiscoverable through standard search techniques.

This tension between richly marked-up text and semi-processed data has been recognized primarily in the wake of the mass digitization projects now underway. In this meeting, emphasis was given, however, to the heterogeneity of the digital tsunami. In addition to the frequently mentioned sources of such heterogeneity from language (including non-Roman scripts), format (text, image), and display (e.g., print versus handwritten, archaic fonts), one of the participants pointed out that much of the modern content that will be of interest to the future is spoken and therefore presents layers of complexity beyond those inherent in information that is born, so to speak, as text, even if that text exists in many languages.

As the white paper by Levander argues, digital library technologies and digital library-based approaches enable organization of collections at a new scale that invites reconceptualizing research questions. In her case, the organization of the Our Americas Archives Project (OAAP) on a hemispheric scale involves three geographically separate institutions and enables a reconceptualization that is independent of modern geopolitical boundaries and that embraces underlying heterogeneity in languages, organization, culture, and format. Consequently, she argues, the structure of the collections helps scholars “to pry [their research] loose” from the self-limiting assumptions of the nation state. Among the responses to her paper were those who argued that it might be prudent to recast the boundaries differently, perhaps, for example, to see the Atlantic world as distinct from the Pacific world rather than defining the scope as the continental landmasses. Nevertheless, no one quarreled with her fundamental insight: that the organization of collections is inherent in the way that research is framed; that such organization of knowledge bounds the way that research is then undertaken, and that challenges to conceptual boundaries begin with the organization of source material.

3.1.2 Services and Tools

The digital tsunami itself affords opportunities to recast research but necessitates new strategies for managing the data, extracting relevant information, and understanding patterns within it. Huberman and his team have been able to show that information embodies social networks and have devised algorithms for identifying those relationships. Thus, one of the implications of that body of work is to focus attention on the Web technology and the web of digital information itself as an object of study. The white paper by Crane et al. outlines a set of challenges that arise because the data potentially available to researchers in the humanities is (1) vaster in size than that previously accessible to researchers and (2) vastly more heterogeneous. They argue for advances at the infrastructure level that would support two broad fronts: “On the one hand, we are extending the intellectual range of individual scholars, enabling them to pursue topics that

require analysis of more primary sources or more linguistic materials than was feasible with print. . . . At the same time, we want to increase the complementary effect and further extend the audiences that the products of particular cultures can reach” (Crane et al.: 2). It is telling that Crane and his coauthors articulate goals that are levels of abstraction below specific tools, a point of view that was implicitly and explicitly shared by the entire conference, albeit from different points of view, probably because a focus on tools is ultimately short-sighted and may not lead to creative or deep collaborations across domain/computer science boundaries.

One of the computer scientists suggested that humanities scholars need to “get to the next level of problem definition, perhaps talking about the tasks they need solved (such as finding something particular in text) rather than the system they need built.” This comment resonated with a recommendation from another computer scientist, a specialist in human-computer interfaces and design, who advised humanists to be able to answer the question, “What is it you are trying to do?” and to explain the kinds of evidence that would be necessary to adduce to answer a given question rather than focusing on the available technologies or the technologies they believe are available. Finally, a professor of romance languages asked rhetorically, “Are we letting our anxieties about tools and protocols, and methodologies obscure bigger questions?” observing that methods, protocols, and disciplines gradually evolve only after the need for a function or capability has been perceived.

For all that there was generally agreement on the need to look beyond the capabilities of known tools, there were also specific and concrete recommendations about ways to proceed within the framework of existing tools. First, existing tools should be identified, evaluated for their broad utility to a scholarly community, and maintained; these are among the goals of Project Bamboo (<http://projectbamboo.org/>). Tools and services should be integrated from the perspective of the user’s experience to avoid creating “multiple fragmented environments.” Second, ontologies are a useful device for organizing the conceptual structure of a field or discipline, and there is substantial work in progress of potential interest to humanists. This continues to be a vibrant research topic among computer scientists and represents one of the boundary areas where domain scholars might interact fruitfully with computer scientists. Third, tools do exist among advanced researchers that might be interesting to humanities scholars, but “we have a long way to go to make interacting with any of these tools anything other than abhorrent,” the HCI specialist said. Yet another computer scientist with a long track record in collaborations across many disciplines essentially concurred with this sentiment in his comment to the forum where he wrote about the divergence between the career path of a computer scientist and the needs of the domain scholar: “Often the work will include a prototype that is stable enough for performance measurements or usability testing. Very rarely will this prototype include all the details that would be required for practical use.”

3.1.3 Research and Analysis

The work by Huberman and his team offers one set of tools for understanding the Web or subsets of information that is internally referenced, like e-mail. This work points to an approach in which a set of computational methods is applied to a computational phenomenon; it is, in sense, wholly contained within the digital world. It is important, a geographer reminded the group, to stop thinking “of the computer as a black box.” This fairly innocent and ordinary image, in fact, captures and exposes several fault lines in digital scholarship. One set of fault lines reflects those who are at home with the technology and those who use the tools built by others; we will come back to this point in the next two sections. A second set, which is the subject of this section, concerns those who are comfortable—or at least recognize—the limitations with the technology and a style of iterative, computational analysis as distinct from those who implicitly assume a

model in which the data are submitted for processing and an answer is retrieved. Certainly the latter model was common in the days of mainframes when computational cycles were scarce and expensive. Personal computing upends that model of computational inquiry, and computational analysis becomes exploratory and iterative. Moreover, computers, one participant said, “aren’t quite as precise at parsing as many people might think.” More generally, Oard argues in his white paper, “humanities scholars are going to need to learn a bit of probability theory.” This is a different model of reasoning, where the arguments concern statistical likelihood, degrees of confidence, and the presence of error, terms that have precise and well-defined mathematical definitions.

Much of the discussion revolved around uses of computing that is exploratory, useful for detecting anomalies and patterns, and frankly accepting of a degree of uncertainty. Among such technologies are clustering (as a means of mathematically representing text or aspects of text), visualization (which is a means of communication as well as analysis), and social networking (the subject of the paper by Huberman). The scholar of old Norse suggested, for example, that the social networking approach laid out by Huberman could be used to “map the social network in [Icelandic] sagas over time and then perhaps integrate with GIS and use this to try and draw actual historic and geographic interpretations.”

As a scholar of the ancient Persia pointed out, though, this form of close analysis of the material does have repercussions in terms of the traditional disciplines. “Cuneiformists” are viewed as data driven and at his home institute, this label can be considered pejorative. In cross-disciplinary meetings like this one, he hastened to add, “people view such a label with esteem.” His cautionary observation and its implications for the organization of knowledge and the ways that the existing structure of disciplines infiltrates perception of significance resonates with the point that Levander made concerning the organization of archives and the ability of reconceptualizing their organization as a way to challenge inherited boundaries between disciplines. Nevertheless, the comment points to the importance of context and perception in the definition of research topics and the reality, shared by the group, that such changes in perception and culture will occur over future generations of students.

Iterative use of computation as a mechanism for detecting anomalies and patterns was one thread in this discussion. A second concerned certainty and trust. Some in the group were quite conservative, arguing that verification and trust were very important and that maintaining and developing methods for sustaining trust in the data and the systems should be paramount. Others, however, were more comfortable with models of probabilistic reasoning, arguing that the approach allowed for creativity and scaled to vast amounts of information even though it also permits a degree of uncertainty. And indeed, since much humanities scholarship rests on interpretation of ambiguous source material, scholars in the humanities are actually fairly comfortable with ambiguity if not with the specific mathematics of probability. Humanists, one art historian quipped, “are programmed to spin out ideas of a starry-eyed dream world if only they were given enough money to do the research and produce the tools.”

3.1.4 Teaching and Communication

The essays by Murray and Stone point to the importance of computation as a means of communication, particularly communication as educators although several participants pointed out that future generations of students will communicate using graphics, rather than text. An architectural historian of French gothic cathedrals, Murray uses a mix of capture and display technologies to re-create the three-dimensional spaces so that his students can also re-experience the soaring interiors at an otherwise inaccessible level of detail and to demonstrate relationships

among resources that are geographically separate. He argues that pedagogical technique removes the cathedral from its status as a fully formed and static object represented by a slide in a darkened lecture hall and allows students to understand these were works in progress over a period of decades, embodying countless choices and decisions. In a brief and vibrant demonstration, he used simulation techniques based on the underlying engineering principles to show the evolution of rounded Romanesque arches to slender Gothic pointed arches as an aesthetic response to engineering constraints.

As the world of the Web becomes increasingly graphics intensive, visual literacy becomes extremely important. This, Stone argues, has two dimensions: one is the ability to communicate effectively using shape, color, and other visual tools; the second is the ability to recognize miscommunication and bias, which may stem from several sources. The point is to teach students to look for bias rather than to accept the display. Indeed, the emphasis that she places on visual literacy goes to a theme that many participants echoed—the importance of understanding what the systems and tools could and could not do and hence the probability that true practitioners would be found in the next generation of scholars, what one writer called, “interdisciplinary natives.” This sentiment was voiced in different ways. Some participants simply acknowledged that the beneficiaries of these discussions lie in the future. Others were more specific about ways in which reasoning and communication would be different, notably the use of visualization and probability and comfort with technology more generally, as well as with interdisciplinarity and alternative ways of organizing knowledge.

3.1.5 Credentialing and Cultures

Threaded through the discussions was an acknowledgement of the role of academic cultures, the boundaries and expectations that exist within those cultures, and the career trajectories that the current culture imposes on both students and faculty. Differing views over the value of data-driven research have already been cited. Appropriate credit in collaborative efforts is another constraint, particularly in the humanities where the single-authored work is the norm. The potential expansion in authorship is a third consideration. Although it has perhaps not been fully realized, the technology offers the possibility of enlarging the pool of those who can “become involved with the production of knowledge” and building a cyberinfrastructure to support such democratization, in the view of one participant, should be a goal. Closer to home, undergraduate students are already participating in major research projects by contributing to projects and by publishing results. However, not all institutions and their faculties are equally provisioned and expanding the infrastructure domestically and internationally remains a concern.

3.2 Research Challenges and Opportunities

There was some discussion of the notion of a new environment, with some questioning how new or innovative much of the research is and others questioning the meaning of the term “environment.” CLIR had intentionally used the word without defining it, in part to see the response and in part to avoid connotations associated with “cyberinfrastructure,” which supports the environment, and in part, to avoid the more limiting notion of “tools.” In fact, the computer scientists in the conference consistently urged participants not to think in terms of tools but rather in terms of what they wanted to do. The latter might be thought of as functions that a research infrastructure should support rather than the specific tools that accomplished a set of tasks. That said, the environment, or medium, consists of digital data; the network; local or remote access to facilities and computational resources that allow users to capture, discover, and manipulate data; and the ability to communicate and publish findings in digital form together with the standards,

services, protocols, codes of conduct, best practices, and so on that are necessary to enable the systems to function smoothly.

Not all resources will be equally accessible, just as not everyone can use advanced instrumentation on Mauna Kea or at CERN. But cyberinfrastructure, broadly conceived, contains those capabilities. We used the term “environment”, however, to underline the point that the shared suite of facilities and services that constitute the shared cyberinfrastructure may support a series of environments within which individuals may work. And those communities of like minded scholars may have specific requirements, as well as culture, that are built on top of the infrastructure. Our focus was the environment to support humanities scholarship, ways computation does and might affect that scholarship, collaborations that might result, and strategies to influence the future of such research.

In the afternoon session, participants were asked to identify three major challenges or opportunities and three major issues or barriers to achieving those challenges. The responses were rich and fall into three broad categories: topics that might generate research, that is, the substance or content of humanities research; the infrastructure and social context of work, which includes both the academic values that define importance as well as their accustomed modes of conducting research; and, finally, some concrete suggestions for activities and programs.

3.2.1 Research Topics

Research can be grouped into four broad topics: scale, language, space and time, and social networking. The boundaries among these bleed into one another, but they individually represent significant clusters of research and obtain a sufficient level of abstraction that enables researchers of many stripes to situate their research and potentially to discover future partners.

3.2.1.1 Scale and the poverty of abundance

The overwhelming characteristic of contemporary scholarship is digital information of varying quality in quantities that were unimaginable in prior generations. This has strained notions of carefully managed collections with rich mark-up and organized around well defined topics and individuals, but also offers the tantalizing promise of a new style of research and new questions appropriate to this new environment of information abundance. The paradox, of course, is that the abundant information is rife with problems and seems inaccessible given current tools. Moreover, the questions that humanists traditionally ask do not always seem to tap this potential of intellectual riches.

Although participants did not propose questions analogous to the grand challenge questions that characterize science, they did collectively emphasize the importance of finding ways to identify pattern and anomalies. This subsumes text analysis (for example, clustering); representation (“more than just a bag of words,” according to one participant), and modeling; and visualization. Also related to scale are issues of information discovery and retrieval as well as data management, archiving, preservation, and sustainability. “We need more sophisticated search and discovery tools, particularly when searching across interdisciplinary collections; users need tools to find resources they did not know existed yet turn out to be very relevant,” one participant said.

3.2.1.2 Language

Because language is fundamental to humanities research, CLIR ensured that a number of linguists and scholars of literature were present. In addition, language is a major source of heterogeneity, particularly where the notion of language encompasses spoken as well as written languages. While linguistic heterogeneity is an attribute of scale, we have separated out language as a distinct topic, in part because issues of scale are not necessarily always linguistic issues and in part because there are rich intellectual traditions in both computer science and language studies that can be brought to bear. Indeed, Oard's white paper was invited precisely to provide a context for understanding some of the research that takes places among computer scientists. That said, there was clearly unspoken consensus among participants on the importance of language, and linguistic and cross-language studies, but few specific recommendations.

3.2.1.3 Space and time

Like language, notions of space and time inhere in humanities. A geographer observed that while dealing with space is difficult, dealing with space and time is "far tougher." This leads to challenges in dealing with items that need to be analyzed both geographically (that is, spatially) and temporally. Although she was speaking from a computational perspective, the central place of exactly such questions for many humanities disciplines suggests a fruitful area of future research and collaboration.

A second aspect of space concerns the reconstruction of space and its representation in digital form. Murray's paper and demo showed both the analytical and pedagogical power and implications of the technology for studying and teaching art and architecture. Reconstructions of archaeological sites and simulations of past conditions are also a rich area of research as evidenced by the Persepolis project. Such projects are data and computationally intensive and typically require relatively large and interdisciplinary research teams (geography, climate, architecture and engineering, social history, archaeology, and so on) that challenge many conventional models of the research process, a point that will be discussed further in a later section. From the perspective of the research process, there does exist an important question of the value of visualization and simulation as an analytical strategy as distinct from a communication or pedagogical strategy and how the results may be captured, documented, reported, archived, preserved, and ultimately replicated or reused in future studies.

3.2.1.4 Social networking

In the white paper contributed to this project, Huberman describes work undertaken at his lab in the area of social networking, arguing that the web of information represents a network of social relationships as well as a technological network. The information can be read backwards, then, to expose relationships that might not be otherwise evident and to illustrate how the specific technologies affect the allocation of human attention. Based on his work first with e-mail and then with material provided by Amazon.com, Huberman and his team developed a series of algorithms that teased out the relationships and then a set of equations to capture the effects. They showed a winner-take-all effect on the diffusion of ideas such that a handful of ideas within a community may attract the lion's share of the attention, given the referral behavior among readers, and then that information decays fairly slowly. Similar phenomena have been identified in citation practices in scholarly journals, and the significance of this work is in its scale, rigor, and level of abstraction. That is to say, the algorithms can be applied in any body of work where the links can be established and, unlike citation counts, it is not necessarily confined to scholarly literature. Rather, the findings go more generally to how attention is allocated and to the ways

that the interactions between social and technological networks mediate allocation of human attention. As a result, the Web ceases to be a neutral technology but instead affects the outcomes by amplifying and instantiating certain behaviors. Huberman's paper excited substantial comment both for its rigor and its ability to evoke new ways of analyzing material. It was seen both as evidence of the way that the Web becomes an object of study as well as a set of analytics that could be used to characterize text corpora; Icelandic sagas was the example given.

3.2.2 Context of Research

The context in which scholars work is shaped by their academic and disciplinary cultures as well as by their resources and facilities. Several participants mentioned the academic culture and its system of prestige and reward, which prescribes a fairly rigid path to tenure and promotion. Posts to the forum also pointed to inconsistencies between the career path of a graduate student in computer science who might work on an interdisciplinary project and the needs of the domain scholar for a more finished product. A medievalist responded, "Typical trajectory of a humanities computing project: It receives start-up funding from, let us say, NEH. The funding eventually goes away, leaving the project incomplete, usually with software that is neither robust to start with nor with the resources necessary to maintain it over time." Moreover, he continued, "Most humanists don't know what computing capabilities are, so when they talk about what they want, they tend to project their current scholarly practices as a straight line, seeking to replicate what they know." This is not unusual. For example, scholars often start with duplications of what they know how to do and then gradually move from putting up static PDFs to creating hyperlinked documents.

In broad terms, several participants called for initiatives that might use technology to broaden scholars' outlook and perspectives, to democratize access and participation, and to expand the current model of publication. Several spoke to the importance of breaking down barriers and of creating "lively models to convince people that computational tools will help them." Concrete suggestions of ways to accomplish that were case studies and experiments in which such results were showcased, workshops that engaged students, and building collaborative teams. Two specific clusters of challenges stand out: sustainability and collaboration.

3.2.2.1. Sustainability

Sustainability embraces archiving, preservation, and reuse of collections and tools. As such, it is tied to the way that projects are conceived, managed, and eventually shut down when they are concluded. As several participants pointed out, there is substantial work to be done simply identifying what resources exist; whether a permanent repository should be found for them, and if so, then the means to preserve the material; and in the case of tools, resources to sustain the continued development of that tool. The previously cited Project Bamboo is one example of a major initiative to identify tools. Other barriers to sustainability (and to collaboration) include the following:

- Scholars may be reluctant to release their work.
- Preservation of context can be very important as well as very difficult.
- Finding and re-using existing literature corpora is also difficult, requiring substantial work to normalize the data before even text mining can be done.
- There is little consensus among scholars and hence little motivation to sustain materials.

It is hard to overstate the importance of managing data. As one participant asked rhetorically, "What is evidence in the digital world?"

3.2.2.2 Collaborations

Given the purpose of the meeting, the nature and structure of collaborations excited substantial attention. Participants acknowledged both the difficulties in creating incentives for collaborations but also proposed a number of creative responses. For example, one metric that a computer scientist who has undertaken many collaborative projects uses for defining a joint project is the ability of the project to support publication in peer reviewed journals by both the technology and domain experts. Nevertheless, he also acknowledged significant problems with authoring systems designed for humanities scholars, a serious issue when credit and attribution loom large in current systems of promotion and tenure and in particular in the single investigator model that characterizes traditional humanities research. Specific actions and projects included the following:

- Improve connectivity between the larger and smaller institutions and among a variety of potential user groups, including possible citizen scholars who might actively contribute to projects.
- Learn the technology and pose problems, in Oard's terms, that are at the boundaries between disciplines.
- Organize workshops, summer schools, and paper sessions that highlight promising, collaborative research.
- Develop systems that collaboratively track contributions to joint projects
- Turn learning to use tools into a collaborative opportunity.
- Use computers to simulate a counter-factual scenario, which, as the participants discovered by watching Murray's presentation on Romanesque and Gothic arches, can be a transformative experience.
- Explore models for collaboration with students and encourage graduate students to experiment with collaboration, project-based learning, and critical synthesis rather than deep originality.

Model this behavior for them.

3.2.3 *Specific Projects*

Three specific projects either captured ideas that resonated with others or were explicitly endorsed by others. They are described in the following sections.

3.2.3.1 Test collections

Test collections were proposed in several contexts, affording researchers the opportunity to learn and to experiment. Crane described the most ambitious version of this idea when he laid out the model of putting existing large corpora in the public domain up on a large, powerful computer system, such as a teragrid, where researchers could experiment with clustering, text mining, mapping, and so on. On the basis of that experience, he argues, the innovative questions that several people called for might emerge. At the same time, the shared resource becomes central to the structure of a discipline or set of disciplines whose research depends on it. As one participant asked, "What is the Protein Data Bank for the humanities?"

3.2.3.2 Ontologies

Ontologies were also proposed. Although some work has been done, no large teams have formed and there is substantial interdisciplinary potential in such collaborations between domain specialists and computer scientists. Ontologies can be used to capture the formalization of basic concepts and can then inform more sophisticated tools and systems.

3.2.3.3 Hard questions about tools

Finally, many participants referred to tools in many contexts. Hard questions must be asked of these tools: Are they working and if so, why? If they are not working, why not? What is frustrating for users? And where is automation appropriate? These are obviously rhetorical questions, but the tenor of this aspect of the day was a recognition that tool development should be generalized and subjected to systematic evaluation that takes into account all aspects of rigorous software development contextualized in an understanding of the user communities. As a general proposition, humanists were advised to avoid general purpose tools but to see tool development as an iterative and learning process.

4.0 Evaluation and Next Steps

No formal evaluation was included as part of this program. Anecdotal responses from participants indicate that the day was successful, an impression that is strengthened by the speed with which invitees accepted, their participation in the online forum that was maintained in preparation for the conference, and the depth to which participants had prepared for the seminar. At least collaboration has been formed between two participants, and a third participant plans a publication based on his contribution to the online form and the interest it engendered. It was obvious from the discussion that the papers had been read in detail, that participants came prepared with their own comments and contributions, and that there was interest in developing future collaborations.

No concrete next steps were formulated beyond a commitment to preparing both the final report for the NEH and the anthology for public distribution. In addition, Kathlin Smith, CLIR's Director of Communications, has prepared a news item for *CLIR Issues*, the organization's bi-monthly newsletter, reporting on the seminar and building interest in the forthcoming anthology.

Appendices

1. Agenda

**Promoting Digital Scholarship:
Formulating Research Challenges
In the Humanities, Social Sciences and Computation**

Co-Sponsored by:
Council on Library and Information Resources
National Endowment for the Humanities

September 15, 2008
Washington, DC

Agenda

7:30 am – 8:30 am	Breakfast
8:30 am – 9:00 am	Welcome: Goals and Scope of the Meeting and Plan for the Day <i>Amy Friedlander, CLIR</i> <i>Joel Wurl, NEH</i> <i>Charles Henry, CLIR</i>
9:00 am – 10:30 am	Review of White Papers and Questions for Authors
10:30 am – 11:00 am	Break
11:00 am – 12:30 pm	How does the new environment (or medium) shape evidence, methods and questions by discipline or across disciplines?
12:30 pm – 2:00 pm	Lunch
2:00 pm – 3:30 pm	What are the questions and where are the opportunities?
3:30 pm – 4:00 pm	Break
4:00 pm – 4:30 pm	Summary (all)
4:30 pm – 5:00 pm	Wrap up <i>Charles Henry</i>

II. List of Participants and Biographical Sketches

**CLIR/NEH Symposium
Participants
September 15, 2008**

Alison Babeu

Research Coordinator
Perseus Project
Tufts University
Medford, MA 02155

Anthony Beavers

Professor of Philosophy
Director of Cognitive Science
University of Evansville
1800 Lincoln Avenue
Evansville, IN 47722

Selmer Bringsjord

Chair, Department of Cognitive Science
Director, RAIR (Rensselaer AI &
Reasoning) Lab
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY 12180

Gregory Crane

Professor of Classics
Tufts University
Medford, MA 02155
Winnick Family Chair of Technology and
Entrepreneurship
Editor-in-Chief, Perseus Project

Robert Darnton

Carl H. Pforzheimer University Professor
Director of the University Library
Office of the Director
Harvard University Library-Wadsworth
House
1341 Massachusetts Ave.
Cambridge, MA 02138

Dr. Charles B. Faulhaber

James D. Hart Director
The Bancroft Library
University of California
Berkeley, CA 94720-6000

Rachel Frick

Senior Program Officer
Institute of Museum and Library Services
1800 M Street, 9th Floor
Washington, DC 20036

Amy Friedlander

Director of Programs
Council on Libraries and Information
Resources
1755 Massachusetts Avenue, NW Suite 500
Washington, DC 20036-2124

David Germano

Associate Professor
Director, Digital Humanities Center
(provisional - being named)
Director, Tibetan and Himalayan Digital
Library
University of Virginia
PO 400126
Charlottesville, VA 22904

Stephen M. Griffin

Program Director
Division of Information and Intelligent
Systems (IIS) National Science Foundation
4201 Wilson Boulevard, Room 1125
Arlington, VA 22230

Charles Henry

President
Council on Libraries and Information
Resources
1755 Massachusetts Avenue, NW Suite 500
Washington, DC 20036-2124

Bernardo Huberman

Senior Fellow and Director
Social Computing Lab
HP Laboratories
Palo Alto, CA 94304

Caroline Levander
Professor of English
Director of the Humanities Research Center
Rice University
P.O. Box 1892
Houston, TX 77005-1892

Marilyn M. Lombardi, Ph.D.
Director
Renaissance Computing Institute (RENCI)
Center at Duke University
Duke University Senior IT Strategist & ISIS
Senior Research Scholar
EDUCAUSE Learning Initiative (ELI)
Scholar-in-Residence
Duke University
OIT-Telecommunications Bldg
390 Science Drive Extension
Durham, NC 27708

Clifford Lynch
Executive Director
Coalition for Networked Information
21 Dupont Circle NW
Washington, DC

Deanna Marcum
Associate Librarian for Library Services
Library of Congress
Washington, D.C. 20540-4900

Tara McPherson
Associate Professor
School of Cinematic Arts
USC Editor, /Vectors/, Critical Studies/SCA
University of Southern California
Los Angeles, CA 90089-2211

David T. Moore
Technical Director
Office of the NSA/CSS Senior Intelligence
Authority
National Security Agency
Currently at the
Center for the Advanced Study of Language

Stephen Murray
Professor of Art History and Archaeology
Columbia University
New York, NY 10027

Fraser Neiman
Lecturer and Director of Archaeology,
Monticello
University of Virginia
Department of Archaeology
Monticello
Box 316
Charlottesville, VA 22902

Stephen G. Nichols
James M. Beall Professor of French &
Humanities Chair, German & Romance
Languages and Literatures Johns Hopkins
University
Baltimore, MD 21218-2687

Lucile T. Nowell
Program Director
Data, Data Analysis & Visualization
National Science Foundation
Office of Cyberinfrastructure
4201 Wilson Boulevard, Room 1145
Arlington, Virginia 22230

Douglas Oard
Associate Dean for Research,
College of Information Studies
Associate Professor, College of Information
Studies Institute for Advanced Computer
Studies
University of Maryland
College Park

Andreas Paepcke
Senior Research Scientist
Stanford University

Donna Peuquet
Professor of Geography
Pennsylvania State University
306 Walker Building
University Park, PA 16802

Mark Schiefsky
Harvard University
Department of the Classics
204 Boylston Hall
Cambridge, MA 02138

Kathlin Smith

Director of Communications
Council on Libraries and Information
Resources
1755 Massachusetts Avenue, NW Suite 500
Washington DC 20036-2124

Matthew W. Stolper

Professor of Assyriology
John A. Wilson Professor of Oriental
Studies
University of Chicago
Oriental Institute
1155 East 58th St.
Chicago, IL 60637

Maureen Stone

StoneSoup Consulting
15340 162nd Ave. NE
Woodinville, WA 98072

Timothy Tangherlini

Professor/Chair, The Scandinavian Section,
UCLA
Professor, Asian Languages and Cultures
Dept., UCLA
212 Royce Hall, UCLA
Box 951537
Los Angeles, CA 90095-1537

Donald J. Waters

Program Officer, Scholarly Communications
The Andrew W. Mellon Foundation
140 East 62nd Street, New York, NY 10065

Joel Wurl

Senior Program Officer
Division of Preservation and Access
National Endowment for the Humanities
1100 Pennsylvania Ave. NW
Washington, DC 20506

CLIR/NEH Symposium
September 15, 2008
Participant Biographical Sketches

Alison Babeu

Research Coordinator
Perseus Project
Tufts University

Alison Babeu has worked as the research coordinator for the Perseus Digital Library since 2004. Before coming to Perseus, she worked as a librarian at both the Harvard Business School and the Boston Public Library. She has a BA in History from Mount Holyoke College and an MLS from Simmons College. Her current research interests include the relationship between digital libraries and mass digitization projects, and how libraries will need to evolve in order to provide the more sophisticated access and tools that scholars will need to mine the growing wealth of digital materials.

Anthony Beavers

Professor of Philosophy
Director of Cognitive Science
University of Evansville

Anthony F. Beavers (<http://faculty.evansville.edu/tb2/>) is Professor of Philosophy and Director of Cognitive Science at the University of Evansville in southern Indiana. Since 1995, he has been involved on several projects relating to the use of the Internet in humanities scholarship. In 1996, he created (with Hiten Sonpal) the Argos search engine, dedicated to the ancient and medieval world, followed soon after by Hippias, a similar initiative in philosophy. Both Argos and Hippias were devoted to emergent organization and quality control of open access resources found online. In 1998, Beavers created Noesis (<http://noesis.evansville.edu>), another prototype for emergent organization and quality control in philosophy that continues to undergo development. Beavers has served as Executive Director of the International Association for Computing and Philosophy and continues to be active in the organization. As program chair for its 2007 conference, he oversaw a program dedicated to the free software and open access movements. This year he is serving again as chair for a conference dedicated to the limits of computation. Beavers is an affiliated researcher on the Indiana Philosophy Ontology Project (<http://inpho.cogs.indiana.edu/index.php>) and was recently awarded a 2008-2009 Digital Humanities Fellowship from the National Endowment for the Humanities to continue his work on Noesis at Indiana University.

Selmer Bringsjord

Professor of Cognitive Science and Computer Science
Director, Rensselaer AI and Reasoning (RAIR) Laboratory
Department of Cognitive Science (Chair)
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)

Selmer Bringsjord specializes in the logico-mathematical and philosophical foundations of artificial intelligence (AI) and cognitive science and in building AI systems on the basis of computational logics, including systems that assist intelligence analysts (e.g., the Slate system),

systems that are at least *apparently* creative (e.g., the Brutus system), and robots with human-level capacity (e.g., the robot PERI). He received the bachelor's degree from the University of Pennsylvania, and the PhD from Brown University in 1987, where he studied under Roderick Chisholm. Since 1987, he has been on faculty in the Departments of Cognitive Science and Computer Science at Rensselaer Polytechnic Institute (RPI) in Troy, New York, where as a Full Professor he teaches AI, symbolic logic, human and machine reasoning, and philosophy of AI. Dr. Bringsjord's publications range from science fiction to technical monographs, notably *What Robots Can & Can't Be* (1992, Kluwer), which is concerned with the future of attempts to create robots that behave as humans, and his most recent book, *Superminds: People Harness Hypercomputation, and More* (2003, Kluwer) as well as papers on such areas as AI, robotics, logic, gaming, philosophy of mind, and ethics. <http://www.rpi.edu/~brings>.

Gregory Crane

Professor of Classics

Tufts University

Gregory Crane's interests are twofold. On the one hand, he has published on a wide range of ancient Greek authors (including books on Homer and Thucydides). At the same time, he has a long-standing interest in the relationship between the humanities and rapidly developing digital technology. He began this side of his work as a graduate student at Harvard when the Classics Department purchased its first TLG authors on magnetic tape in the summer of 1982 and he has worked continuously on aspects of digital humanities ever since. His current research focuses on what a cyberinfrastructure for the humanities in general and classics in particular would look like. He is especially interested in how technology can extend the intellectual range of researchers moving through very large collections and working with more languages than was ever possible in print culture.

Robert Darnton

Carl H. Pforzheimer University Professor

Director of the University Library

Office of the Director

Harvard University Library

Much of my research concerns publishing, bookselling, writing, and reading in eighteenth-century Europe, mainly France. I found the archives so rich (50,000 unpublished letters from every corner of the book trade) that I decided to circumnavigate the subject in an e-book, which readers will be able to use in many unconventional ways. While president of the American Historical Association, I launched a project, Gutenberg-e, which was designed to promote the publication of prize-winning Ph.D. dissertations in electronic form. After seven years, it has been acclaimed as a success in some respects (breaking down barriers against e-publishing and setting scholarly standards for this new kind of book) and criticized as a failure in others (the business plan was not strong enough to make the project self-sustaining after the expiration of a grant from the Mellon Foundation.) Now that I have become director of the Harvard University Library, I must face many digital issues. I hope the symposium will help me cope with them.

Dr. Charles B. Faulhaber

James D. Hart Director

The Bancroft Library

University of California

Charles Faulhaber is Professor of Medieval Spanish Literature and Director of The Bancroft Library, Berkeley's rare book and special collections library. He has been involved with humanities computing for over thirty years, starting with the design of a data base to catalog the medieval manuscripts of the Hispanic Society of America (New York), a project that led to a comprehensive bio-bibliographical data base of medieval Spanish literature which has been available on the World Wide Web since 1997 (<http://sunsite.berkeley.edu/PhiloBiblon/phhm.html>). He also collaborated with the Spanish Quincentennial Commission's project to publish on CD-ROM digitized facsimiles and facing machine-readable texts of early Spanish printed texts (ADMYTE: Archivo Digital de Manuscritos y Textos Espanoles, 1992-93). He also taught one of the first distance learning classes at Berkeley (1995), on Old Catalan language and literature, with students at UC Irvine and UC Santa Barbara as well as at Berkeley.

At Bancroft he has actively promoted the mass digitization of primary source materials, manuscripts, archives, rare books, and pictorial collections. Currently Bancroft is working with the Internet Archive and expects to join the Google project next year.

He has been involved with academic computing on the Berkeley campus since 1980, having chaired the Academic Senate Computing and Communications Committee and served on numerous other standing and ad hoc committees, including, currently, the Leadership Council of Project Bamboo and its Berkeley counterpart.

Amy Friedlander

Director of Programs

Council on Library and Information Resources

Amy Friedlander is Director of Programs at the Council on Library and Information Resources where she is primarily engaged in projects involving cyberinfrastructure, preservation, and digital scholarship, encouraging partnerships and cross-fertilization of ideas across disciplines, agencies, and institutional boundaries. She is the founding editor of D-Lib Magazine and subsequently SAIC's now defunct iMP: The magazine on information impacts. She has also participated in the organizational phases of the Library of Congress' National Digital Information Infrastructure and Preservation Program. Since joining CLIR in 2007, Ms. Friedlander has been appointed to the National Science Foundation's Blue Ribbon Task Force on Economically Sustainable Digital Preservation and Access, guest-edited a special issue of the Journal of Electronic Publishing on communication and cyberinfrastructure, and organized a workshop with Gregory Crane on the implications of large-scale text digital corpora for humanities scholarship. She is the author of five short monographs on the history of large-scale, technology-based infrastructures in the United States. She holds the A.B. from Vassar College, the M.A., and Ph.D. in U.S. history from Emory University, and the M.S.L.I.S. from The Catholic University of America.

David Germano

Associate Professor

History of Religions, Buddhist Studies

University of Virginia

David Germano teaches and researches Tibetan and Buddhist Studies at the University of Virginia, directs the Tibetan and Himalayan Library (www.thdl.org, THL), and directs the new digital humanities center at UVa (in the process of being named). His personal scholarship focuses on the history of Tibetan culture and religion from the ninth to fourteenth century with a special focus on esoteric religious movements. At present, he is working with other faculty and

staff at UVa to establish a faculty governed center focused on integrating and coordinating various digital initiatives with a humanistic focus at UVa, as well as strengthening the dynamic interconnections between innovation and mainstream activity involving digital technology amongst faculty and students.

Since the mid 1990s he has explored digital technology as a means to facilitate interdisciplinary, collaborative, and engaged scholarship in Tibetan Studies under the umbrella of THL. The Library covers a wide range of types of materials (images, audio-video, texts, immersive objects, maps, etc.) and disciplines (geography, literature, religious studies, musicology, art, archaeology, etc.), but especially is focused on building spatial and temporal resources as an interdisciplinary and collaborative nexus for diverse projects and researchers. In this capacity, he has built strong partnerships with institutions, scholars, students and local communities in North America, Europe and Asia to work in collaborative and mutually beneficial projects. The initiative also stresses supporting Tibetan scholars and educators, and aims to encourage and facilitate engaged scholarship that actively cares about how academic work benefits local communities around the world. This includes using technology creatively to help support bridges between academics and development projects, and to enable local communities to use modern tools as vehicles for their own self-expression and empowerment. This participatory movement aims to redefine the notion of "scholars" and "scholarship" to include local communities across the world in a truly distributed production of knowledge. To this end, THL facilitates communication, coordination, and collaboration among individuals, organizations and projects of all types with a common interest in knowledge and education pertaining to Tibet and the Himalayas.

Stephen M. Griffin

Program Director

Division of Information and Intelligent Systems (IIS) National Science Foundation

Charles Henry

President

Council on Library and Information Resources

Charles Henry is President of the Council on Library and Information Resources. Prior to this appointment, he was vice provost and university librarian at Rice University, a position he had held since 1996. In that role, he was responsible for library services and programs, including the Digital Library Initiative and the Digital Media Center. He continues also publisher of Rice University Press, which has recently been reborn as the nation's first all-digital university press.

Mr. Henry has been a leader in the digital libraries and digital humanities movements. He is a trustee of the Digital Library Federation and chair of the advisory committee for the Information Resource Center at the International University of Bremen. He serves on the advisory board of Stanford University Libraries and on the ACLS Commission on Cyberinfrastructure in the Humanities and Social Sciences. He received a Fulbright senior scholar grant for library sciences in New Zealand, a Fulbright award for the study of medieval literature in Vienna, Austria, and recently completed a third Fulbright award in China. Mr. Henry has a Ph.D. in comparative literature from Columbia University.

Bernardo Huberman (not attending meeting)

Senior Fellow and Director

Social Computing Lab

HP Laboratories

Bernardo A. Huberman is a Senior HP Fellow and the Director of the Information Dynamics Lab at Hewlett Packard Laboratories. He is also a Consulting Professor in the Department of Applied Physics at Stanford University. For the past eight years his research has concentrated on the phenomenon of Web, with an emphasis on the understanding of its implications for social dynamics and the design of novel mechanisms for discovering and aggregating information. He is the author the book, *The Laws of the Web: Patterns in the Ecology of Information*, published by MIT Press. More information on his work is available at: <http://www.hpl.hp.com/research/scl/people/huberman>.

Caroline Levander

Professor of English

Director of the Humanities Research Center

Rice University

Caroline Levander is Professor of English and Director of the Humanities Research Center at Rice University. She is currently writing *Laying Claim: Imagining Empire on the U.S. Mexico Border* (under contract, Oxford University Press) and *The Idea of American Literature* (for Wiley-Blackwell's Manifesto Series), and co-editing *Engaging the Americas* (Palgrave Macmillan). She is author of *Cradle of Liberty: Race, the Child and National Belonging from Thomas Jefferson to W.E.B. Du Bois* (Duke University Press, 2006) and *Voices of the Nation: Women and Public Speech in Nineteenth-Century American Culture and Literature* (Cambridge University Press 1998) as well as co-editor of *Hemispheric American Studies* (Rutgers University Press, 2008) and *The American Child: A Cultural Studies Reader* (Rutgers University Press, 2003).

She is co-editor of a new book series, *Imagining the Americas*, with Oxford University Press, co-founder of the Americas Colloquium at Rice University and has developed the Rice Americas Archive. In collaboration with University of Maryland's Early Americas Digital Archive, the Americas Archive has generated the Our Americas Archive Partnership, which was awarded a 3-year National Leadership Grant from the Institute of Museum and Library Services for \$979,578. She has recently led an NEH Summer Seminar on the topic of hemispheric American literature and a National Humanities Center Dupont seminar on the globalization of American literary studies.

Her research begins with the acknowledgment that nineteenth-century US literature and politics were integrally blended. Most broadly, her work considers the dual questions of American literature's political impact and American political culture's literary effects. Using a wide-range of archival and literary sources, Levander explores how the writing of prominent Americans as well as those historically disenfranchised within the United States—women, children and racially diverse citizens—reconstitutes conceptual frameworks of nation formation and literary heritage.

Marilyn M. Lombardi, Ph.D.

Director

Renaissance Computing Institute (RENCI) Center at Duke University

Duke University Senior IT Strategist & ISIS Senior Research Scholar

EDUCAUSE Learning Initiative (ELI) Scholar-in-Residence

Duke University

Marilyn M. Lombardi is director of the Renaissance Computing Institute (RENCI) Center at Duke University. The Center engages Duke faculty within the physical and material sciences, the humanities, and the social sciences in multidisciplinary and multi-institutional collaborations that

leverage the advanced visualization, processing, and networking infrastructure and expertise of the statewide RENCi organization. As a former associate professor of English turned research computing strategist, Marilyn has made cross-disciplinary collaboration the hallmark of her professional agenda for many years. Much of Marilyn's research and strategic activities focus on the realization of a 3D "metamedium" for deeply collaborative digital scholarship, learning and discovery based on a scalable, open-source architecture. Last year, Marilyn served as a member of the advisory panel for the National Endowment for the Humanities (NEH)'s new grant program in Digital Humanities Scholarship. This year, she was awarded a planning grant from the National Science Foundation (NSF) Office of Cyberinfrastructure and NSF Directorate for Computer & Information Science (CISE) to enlist thought leaders from across the diverse human-computer interaction (HCI) research community in the development of a coordinated vision and set of strategic recommendations for the future of human-computer interaction in support of 21st century discovery. Most recently, she has taken on a leadership role in a Kauffman Foundation planning initiative aimed at developing and disseminating a robust infrastructure for the assessment of learning within virtual worlds. For the past several years, Marilyn has written white papers on transformative learning practices in higher education as scholar-in-residence for the EDUCAUSE <<http://www.educause.edu/>> Learning Initiative (ELI), and she continues to serve as senior research scholar in the Information Science and Information Studies program at Duke University and senior strategist for Duke's Office of Information Technology. Her recent publications include a contribution to the Carnegie Foundation book "Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge" (MIT Press, 2008). A former associate professor of English, she is also the author of a book, "The Body and the Song: Elizabeth Bishop's Poetics"; an edited volume, "Elizabeth Bishop: The Geography of Gender"; and numerous articles in scholarly publications.

Clifford Lynch

Executive Director

Coalition for Networked Information

Deanna Marcum

Associate Librarian for Library Services

Library of Congress

I am the Associate Librarian for Library Services at the Library of Congress. Before joining the Library in 2003, I was the president of CLIR. My interest in digital scholarship dates to those days at CLIR when we were collaborating with the scholarly community to determine how digital technology would affect the production of scholarship and the provision of library services. At the Library of Congress, we are developing a set of services to offer in the facilitation and promotion of digital scholarship. I hope to learn more at the symposium about how we can develop such services when there is no set community of users. Understanding others' expectations for the offerings of the national library would be most helpful.

Tara McPherson

Associate Professor

School of Cinematic Arts

University of Southern California

Tara McPherson teaches courses in new media, television, and popular culture in the School of Cinematic Arts at the University of Southern California (USC). She is author of the award-winning *Reconstructing Dixie: Race, Gender and Nostalgia in the Imagined South* (Duke UP: 2003). She is co-editor of the anthology *Hop on Pop: The Politics and Pleasures of Popular*

Culture (Duke UP: 2003) and editor of *Digital Youth, Innovation and the Unexpected*, part of the MacArthur Foundation series on Digital Media and Learning (MIT Press, 2008). She is currently co-editing an anthology on digital narrative and politics and working on a book manuscript on the racial epistemologies of new media. Her new media research focuses on issues of convergence, gender, race, and representation, as well as upon the development of new tools and paradigms for digital publishing, learning, and authorship. She is the founding editor of *Vectors*, the multimedia peer-reviewed journal sponsored by the Institute for Multimedia Literacy at the University of Southern California. *Vectors* pushes far beyond the "text with pictures" format of much online scholarly publishing, encouraging work that takes full advantage of the multimodal and networked capacities of computing technologies. She was recently selected as one of three editors for the new MacArthur-supported *International Journal of Learning and Media* (forthcoming from MIT Press in 2009), a hybrid online/print journal that will also explore new forms of online publishing. Co-organizer of the 1999 conference, *Interactive Frictions*, Tara is among the founding organizers of *Race in Digital Space*. She is a member of the Academic Advisory Board of The Academy of Television Arts and Sciences Archives, has served as an AFI juror, is a member of HASTAC, and is on the boards of several journals and of the Scholarly Communication Institute. Her research activities have been funded by the Rockefeller, Mellon, MacArthur, Ford, and Annenberg Foundations. <http://www.vectorsjournal.org>

David T. Moore

Technical Director

Office of the NSA/CSS Senior Intelligence Authority

National Security Agency

Currently at the Center for the Advanced Study of Language

Stephen Murray

Professor of Art History and Archaeology

Columbia University

Stephen Murray was educated at Oxford and London Universities. He has taught as professor and has served as chairman at Indiana and Columbia Universities and has held grants and fellowships from the Guggenheim Foundation, the Stanford Center for Advanced Studies in the Behavioral Sciences, the National Humanities Center, the National Endowment for the Humanities, and the Andrew Mellon Foundation. He was Founding Director of the Media Center for Art History at Columbia.

In his research and publications he has explored the life of the great Gothic cathedrals of France (Notre-Dame of Paris, Amiens, Beauvais and Troyes). He believes that it is important to consider all aspects of the cathedral including design, construction, social context and liturgical function: this inclusive agenda inspired his most recent book, *A Gothic Sermon* University of California Press, 2004.

In order to animate the cathedral and to make it available to as wide an audience as possible he has, most recently, experimented with the digital media, including the Internet, three-dimensional computer modeling and video. He has taught numerous cathedral seminars in Europe, especially through the National Endowment for the Humanities and Columbia University. Under the auspices of a grant from the Andrew Mellon Foundation currently directs a project to create a database for French Gothic Cathedrals.

Fraser Neiman

Director of Archaeology, Monticello, Virginia

Lecturer, Departments of Anthropology and Architectural History, University of Virginia

Fraser Neiman directs ongoing archaeological research at Monticello (www.monticello.org) into the ecological and social dynamics of the early-modern Chesapeake and the larger Atlantic world of which it was a part. His lab is home to the Digital Archaeological Archive of Comparative Slavery (www.daacs.org). DAACS is an experiment in the use of IT and the internet to share detailed archaeological data, encourage comparative analysis, leverage collaboration, and accelerate progress in understanding the evolution of slave societies of the Chesapeake, Carolinas, and the Caribbean. He teaches courses in quantitative methods, historical archaeology, and archaeological theory at the University of Virginia (www.people.virginia.edu/~fn9r).

Stephen G. Nichols

James M. Beall Professor of French & Humanities

Chair, German & Romance Languages and Literatures, Johns Hopkins University

Stephen G. Nichols, James M. Beall Professor of French and Humanities, heads the Department of German and Romance Languages and Literatures at Johns Hopkins University. He specializes in medieval literature in its relations with history, philosophy, and history of art. One of his books, *Romanesque Signs: Early Medieval Narrative and Iconography*, received the Modern Language Association's James Russell Lowell Prize for an outstanding book by an MLA author in 1984. Another, *The New Philology*, was honored by the Council of Editors of Learned Journals in 1991. In 1992, the University of Geneva conferred on him the title of Docteur ès Lettres, honoris causa, while the French Minister of Culture made him Chevalier de l'Ordre des Arts et Lettres in 1999, and Officier in 2007.

A Fellow of the Medieval Academy of America, he is also a Senior Fellow of the School of Criticism and Theory, which he directed from 1995-2001. Author, editor, and co-editor of twenty-four books, Nichols conceived and is co-director of a project creating digital surrogates of medieval manuscripts at the Milton S. Eisenhower Library of Johns Hopkins. The project is currently working with the Bibliothèque Nationale de France to ingest more than 130 manuscripts of the Romance of the Rose to complement those already on the site. He has lectured and written on digital scholarship in the Humanities, e.g. "From Parchment to Cyberspace," "Digital Scholarship, What's all the Fuss?" "Born Medieval: Manuscripts in the Digital Scriptorium," "Manuscripts and Digital Surrogates: Sibling or Counterfeit?", "There's an Elephant in the Room: Digital Scholarship and Scholarly Prejudice."

Lucile T. Nowell

Program Director

Data, Data Analysis & Visualization

National Science Foundation

Office of Cyberinfrastructure

Now on assignment as a Program Director in the area of Data, Data Analysis and Visualization for the Office of Cyberinfrastructure at National Science Foundation (NSF), Lucy Nowell is a Chief Scientist from the Information Analytics group at Pacific Northwest National Laboratory (PNNL). At PNNL, her research focused on helping users find items and patterns of interest in large collections of documents. She is also an alumna of the Virginia Tech Digital Libraries Research Laboratory, where she designed one of the first information visualization user interfaces

for the Envision Project. Her research interests include long-term data preservation/archiving, user interaction with information in the context of massive data, usability engineering for information exploitation systems and digital electronic libraries, cognitive issues in user interface design, information visualization, intelligent user modeling and intelligent user interfaces, and information storage and retrieval.

As Program Director for Data, Data Analysis and Visualization in NSF's Office of Cyberinfrastructure, her program responsibilities include:

- Sustainable Digital Data Preservation and Access Network Partners (DataNet)
- Community-based Data Interoperability Networks (INTEROP)
- Software Development for Cyberinfrastructure (SDCI)
- Strategic Technologies for Cyberinfrastructure (STCI)

Doug Oard

*Associate Dean for Research,
College of Information Studies*

*Associate Professor, College of Information Studies Institute for Advanced Computer Studies
University of Maryland*

Douglas W. Oard is Associate Dean for Research at the University of Maryland's College of Information Studies. He holds joint appointments as an associate professor in the College of Information Studies and the Institute for Advanced Computing Studies. Dr. Oard earned his Ph.D. in 1996 in Electrical Engineering from the University of Maryland, College Park, a Master of Electrical Engineering degree from Rice University in 1979, and a B.A. in Electrical Engineering and Mathematical Sciences, also in 1979. His research is focused on the design and evaluation of interactive systems to support search and sense-making in large collections of character-coded, scanned, and spoken language. He is best known for his work on cross-language information retrieval, but his current interests also include support for e-discovery in litigation (as a coordinator for the TREC Legal Track) and investigating application of computational linguistics for social science research (as a Co-PI for the NSF-funded PopIT Human Social Dynamics project).

Andreas Paepcke

*Senior Research Scientist
Stanford University*

Andreas Paepcke is a Senior Research Scientist and Director of the Digital Library Project at Stanford University. Dr. Paepcke has served on numerous program committees, including as Program Chair for the Joint Conference on Digital Libraries 2008, and Vice Program Chair of the World-Wide Web conference's 'Browsers and User Interfaces' program track. He served on several National Science Foundation proposal evaluation panels and is associate editor of ACM Transactions on the Web. Dr. Paepcke received BS and MS degrees in applied mathematics from Harvard University, and a Ph.D. in Computer Science from the University of Karlsruhe, Germany. Previously, he worked as a researcher at Hewlett-Packard Laboratory, and as a research consultant at Xerox PARC. He serves on the technical advisory board of Center'd.com.

Dr. Paepcke's interests include user interfaces for small devices, novel Web search facilities, and browsing facilities for digital artifacts that are difficult to index. With his group of students he has designed and implemented WebBase, an experimental storage system for Web contents. He is currently working on a Web Sociologists Workbench. The result of this work will be tools that

allow social scientists and historians to analyze large time-series Web snapshot archives without knowledge of computing intricacies.

Donna Peuquet

*Professor of Geography
Pennsylvania State University*

Donna Peuquet is Professor of Geography at The Pennsylvania State University and a faculty associate of the GeoVISTA Center and was acting Director of the GeoVISTA Center during the 2007-2008 academic year. Dr. Peuquet performs research on the theory of geographic knowledge representation, spatio-temporal data representation, spatial cognition, geocomputation, geographic database design, and the use of GIScience in epidemiological research. Her book, entitled *Representations of Space and Time*, develops an integrated perspective on philosophical, cognitive, database and visualization issues on spatial and space-time representation. She was lead PI (with Alan MacEachren as co-PI) on a recently completed project to develop an integrated database and visualization environment for space-time information exploitation, called STNexus. Recent research includes increasing emphasis on representation of complex geographic processes.

Joyce Ray

*Associate Deputy Director for Library Services
Institute of Museum and Library Services*

Mark Schiefsky

*Harvard University
Department of the Classics*

Mark Schiefsky took his Ph.D. degree in Classical Philosophy from Harvard University in 1999 and has been an Assistant Professor in the Department of the Classics since January 2000. His research interests are centered on the interaction of ancient philosophy and science, especially medicine and mechanics. His publications include a commentary on the Hippocratic treatise *On Ancient Medicine* (Brill 2005), along with several articles on ancient medicine and mechanics; he is currently working on a book that will explore the connections between ancient mechanics, mechanical technology, physics, and mathematics. Professor Schiefsky is also collaborating closely with scholars at the Max Planck Institute for the History of Science in Berlin, Germany on the Archimedes Project, an international initiative funded by the National Science Foundation to create a digital library for the history of mechanics and mechanical technology (<http://archimedes.fas.harvard.edu>). He has taught courses on Plato and Lucretius in the original languages, as well as Ancient Greek Medicine, Introduction to Ancient Philosophy, and Ancient Cosmology and Mechanics in translation.

Kathlin Smith

*Director of Communications
Council on Library and Information Resources*

Kathlin Smith is director of communications at the Council on Library and Information Resources (CLIR) in Washington, D.C., where she oversees CLIR's publications program and sponsor communications. Before joining CLIR in 1997, she worked for nine years at the Committee on Scholarly Communication with China, sponsored by the National Academy of Sciences, American Council of Learned Societies, and Social Science Research Council. She also served as a consultant to the World Bank on projects in China. She holds a B.A. in International

Relations from the Pennsylvania State University, and an M.A. in International Development from American University.

Matthew W. Stolper

Oriental Institute

Matthew W. Stolper (Professor of Assyriology, Oriental Institute & Department of Near Eastern Languages and Civilizations, University of Chicago) is the director of the Persepolis Fortification Archive Project (described at <http://oi.uchicago.edu/research/projects/pfa/>; related postings at <http://persepolistablets.blogspot.com/>). The Persepolis Fortification Archive consists of tens of thousands of clay tablets and fragments, remains of an administrative archive compiled around 500 BC and rediscovered in 1933 by archaeological excavations at Persepolis, the palace complex in southwestern Iran built by the Achaemenid Persian king Darius I and destroyed by Alexander the Great. Thousands of tablets have texts in Elamite language; hundreds have texts in Aramaic language; a few have texts in other languages (Greek, Old Persian, Phrygian, Babylonian); almost all have impressions of one or more seals. Continuing access to this vast, unique source of information on Achaemenid Persian languages, art, society, and history is in grave peril from litigation. The Persepolis Fortification Archive Project at the Oriental Institute of the University of Chicago brings together an inter-institutional and international team of editors and collaborating projects in an emergency effort to record, catalogue and analyze as much of the Archive as possible, and to distribute the results through at least four co-operating on-line sites: the On-Line Cultural Heritage Resource Environment (OCHRE) at the University of Chicago (<http://ochre.lib.uchicago.edu/>); InscriptiFact, the Web site of the West Semitic Research Project at USC (<http://www.inscriptifact.com/>); the Cuneiform Digital Library Initiative (CDLI) at UCLA (<http://cdli.ucla.edu/>); and the paired sites of [achemenet.com](http://www.achemenet.com/) and the Musée Achéménide Virtuel et Interactif (MAVI) at the Collège de France (<http://www.achemenet.com/> and <http://www.museum-achemenet.college-de-france.fr/>). Like many project directors, Stolper is a well-meaning user of modest ability and incomplete knowledge of his resources.

Maureen Stone

StoneSoup Consulting

Maureen Stone has been working in digital color, graphics, perception and the tools for information display for almost 30 years. At Xerox PARC in the 1980's, she participated in the desktop publishing revolution, creating tools for illustration, typography, and color selection. She and her colleagues created some of the first color management systems for digital prepress, uniquely focused on purely digital imagery (as opposed to scanned photographs). At the end of her tenure at PARC, she was a member a small group exploring the relationship between technology and design called RED (Research in Experimental Design), where she worked on digital sound, 3D Web graphics, and a walk-through comic strip. Since founding StoneSoup Consulting in 1999, she has worked on a wide range of research and development activities, from building multi-projector display walls at Stanford to designing color palettes for Tableau Software to teaching Information Visualization in the University of Washington iSchool. She is an adjunct professor at Simon Fraser University's School for Interactive Arts and Technology, and editor in chief of IEEE Computer Graphics & Applications. Her book, *A Field Guide to Digital Color*, was published by A.K. Peters in 2003. She received a BS and MS degrees in Computer Engineering from the University of Illinois, and a MS in Computer Science from Caltech. She is a member of ACM, IEEE and IS&T.
<http://www.stonesc.com>

Timothy Tangherlini

Professor / Chair, The Scandinavian Section, UCLA

Professor, Asian Languages and Cultures Dept., UCLA

Timothy Tangherlini is Professor of Scandinavian and Korean folklore at the University of California, Los Angeles. His current work focuses on applying machine learning techniques (supervised and unsupervised learning) to Danish folklore and Old Icelandic literature corpuses. He is also leading an effort at UCLA to develop an automated morphological analyzer for Old Icelandic (<http://dev.cdh.ucla.edu/~newmedia/ICEmorph/>). Other work related to digital humanities scholarship includes a growing archive of Korean and Korean American folklore based on student collections (<http://projects.cdh.ucla.edu/koreanfolklore>), as well as a born-digital project that presents storyteller repertoires from the collections of the nineteenth century Danish folklorist, Evald Tang Kristensen (<http://dev.cdh.ucla.edu/~newmedia/DFL2>). Along with the UCLA digital library, he is currently in the process of making all 79 volumes of Tang Kristensen's folklore collections, along with a deeply tagged edition of his memoirs, freely available in digital form. He is also a consultant on a project within the Electronic Cultural Atlas Initiative at UC Berkeley, and Co-PI on Mapping Nordic Literary Culture sponsored by the Nordic Council of Ministers.

Donald J. Waters

Program Officer, Scholarly Communications

The Andrew W. Mellon Foundation

Donald J. Waters is the Program Officer for Scholarly Communications at The Andrew W. Mellon Foundation. Before joining the Foundation in 1999, he served as the first Director of the Digital Library Federation (1997-1999), as Associate University Librarian at Yale University (1993-1997), and in a variety of other positions at the Computer Center, the School of Management, and the University Library at Yale. Waters graduated with a Bachelor's degree in American Studies from the University of Maryland, College Park in 1973. In 1982, he received his Ph.D. in Anthropology from Yale University. Waters conducted his dissertation research on the political economy of artisanry in Guyana, South America. He has edited a collection of African-American folklore from the Hampton Institute in a volume entitled *Strange Ways and Sweet Dreams*. In 1995-96, he co-chaired the Task Force of the Commission on Preservation and Access and the Research Libraries Group on Archiving of Digital Information, and was the editor and a principal author of the Task Force Report. He was a member of the Section 108 Study Group. He is a fellow of the American Association for the Advancement of Science, and serves on the Steering Committee of the Coalition for Networked Information, the National Digital Strategy Advisory Board of the Library of Congress. He is also the author of numerous articles and presentations on libraries, digital libraries, digital preservation, and scholarly communications.

Joel Wurl

Senior Program Officer

Division of Preservation and Access

National Endowment for the Humanities

Joel Wurl is a Sr. Program Officer in the Division of Preservation & Access, National Endowment for the Humanities, where he also serves on the inter-divisional working group for the Office of Digital Humanities. He is also an Adjunct Instructor in the Applied History program at George Mason University. Prior to joining NEH in October, 2006, he worked for 20 years with

University of Minnesota's Immigration History Research Center ending there as Head of Research Collections and Associate Director. From 2002 to 2005, he served on the council and executive committee of the Society of American Archivists and as editor of the Midwest Archives Conference journal *Archival Issues*. He co-chaired the program committee for the 2008 SAA annual meeting in San Francisco. Wurl's publications have appeared in both archival and immigration/ethnic history journals, and he is general editor for "North American Immigrant Letters, Diaries, and Oral Histories," an online publication of Alexander St. Press. Wurl was named a Distinguished Fellow of SAA in 2007.