

BAMBOO PLANNING PROJECT

AN ARTS AND HUMANITIES COMMUNITY PLANNING PROJECT TO DEVELOP
SHARED TECHNOLOGY SERVICES FOR RESEARCH

FOR THE ANDREW W. MELLON FOUNDATION

DECEMBER 28, 2007
VERSION 4.8.1S

PRINCIPAL INVESTIGATORS

Janet Broughton, Dean of Arts and Humanities and Professor of Philosophy, University of California, Berkeley
Gregory A. Jackson, Vice President for Information Technology and Chief Information Officer, University of Chicago

LEADERSHIP COUNCIL

Anthony Cascardi, Director, Doreen B. Townsend Center for the Humanities, and Professor of Comparative Literature, Rhetoric, and Spanish, University of California, Berkeley
James Chandler, Director, Franke Institute for the Humanities, and Professor of English Language and Literature, Committees on the History of Culture, Cinema and Media Studies, and Interdisciplinary Studies, and the College, University of Chicago
Charles Faulhaber, Director, Bancroft Library, and Professor of Spanish and Portuguese, University of California, Berkeley
Ian Foster, Director, Computation Institute, and Professor of Computer Science and the College, University of Chicago, and Associate Division Director of Mathematics and Computer Science, Argonne National Laboratory
Judith Nadler, Director and University Librarian, The University of Chicago Library, University of Chicago
Stuart Russell, Professor and Chair, Electrical Engineering and Computer Science, University of California, Berkeley
Shelton Waggener, Associate Vice Chancellor and Chief Information Officer, University of California, Berkeley

PROGRAM DIRECTORS

David A. Greenbaum, Director of Data Services, Information Services and Technology, University of California, Berkeley
Chad J. Kainz, Senior Director for Academic Technologies, Networking Services & Information Technologies, University of Chicago

CONTENTS

I	OVERVIEW OF BAMBOO PLANNING PROCESS.....	3
II	PERSPECTIVES OF FIVE COMMUNITIES	5
II.1	Arts and Humanities Scholars	5
II.2	Computer Science	7
II.3	Information Science	10
II.4	Library and Scholarly Communications.....	11
II.5	Central Information Technology Organizations	13
III	WHY BAMBOO AND THIS PLANNING EFFORT?	17
IV	WHAT DOES SUCCESS LOOK LIKE?	19
V	SCHEDULE AND SUMMARY OF DELIVERABLES	22
V.1	Chicago and Berkeley Bamboo Planning Project Leadership Meeting.....	23
V.2	Workshop One: Planning Process and Understanding Arts and Humanities Scholarship ..	23
V.3	Workshop Two: Identifying Services	24
V.4	Workshop Three: Selecting Services	24
V.5	Workshop Four: Orchestrating Services and Defining Partnerships	25
V.6	Workshop Five: Forming a Consortium	25
VI	LEADERSHIP COUNCIL, RELATED INITIATIVES AND PROPOSAL CONTRIBUTORS	27
VI.1	Leadership Council	27
VI.2	Related Initiatives	28
VI.3	Contributors.....	30

I OVERVIEW OF BAMBOO PLANNING PROCESS

How can we enhance arts and humanities research through the development of shared technology services?

Our near-term ambition is to answer this question by drawing upon the expertise and vision of people in five arenas: researchers in arts and humanities; computer scientists; information scientists; librarians; and campus information technologists. We aim to draw in representatives in these fields from the United States and other countries, involving people from research universities, liberal arts colleges, community college campuses, disciplinary societies, and national technology consortia.

Of course, it is not enough to understand how to enhance arts and humanities scholarship through the development of technology; the technology must then be developed. Our long-term ambition is thus to create shared technology services that will support arts and humanities research, as it both is and could be carried out.

The University of California and the University of Chicago are delighted to have this opportunity to submit a joint proposal to the Andrew Mellon Foundation for support of a community-driven planning process that will lead to a rich and well-informed answer to the key question above. Once the planning process has yielded an answer, we expect to be able to form a broader consortium of institutions to implement the specification and development of the technologies that researchers in arts and humanities need.

We are motivated to make this proposal and to undertake the work involved in the planning and implementation processes because we believe that researchers in arts and humanities are radically underserved and under-funded by comparison with researchers in other disciplines. Researchers in the arts and humanities who attempt to use technology in their work encounter significant barriers to access to digital content, work in isolation from systems of technological support, and spend far too much of their time on constructing tools when they should be able instead to focus on using them. Both of our universities have been leaders in the humanities and have seen the humanities as central to their academic missions. We are thus deeply concerned by the fact that the humanities are so conspicuously underserved at the beginning of what will clearly be a century of great technological innovation.

The fragmented way in which technology has so far been created and used in the arts and humanities means that researchers across these fields do not have a shared sense of what technology they need now and what technology they would like to have in the future. Still less do computer scientists or information technologists in academia understand what technologies would be helpful to researchers in the arts and humanities. That is why we believe our first ambition must be to complete a careful planning process, one that is driven by researchers in arts and humanities but involves key people in the technological realm from the start.

The planning process that we envision would proceed via a series of interconnected workshops carried out over eighteen months in a variety of locations. The workshops will help researchers in arts and humanities to articulate their needs for technology, which will enable the technologists to consider what kinds of technological solutions might work. The workshops will also help researchers in arts and humanities to understand what kinds of support digital technologies could

offer them, which would open up new research horizons. We thus see the workshops as an opportunity for an upward-spiraling conversation about how research in arts and humanities can advance with the help of shared technology services.

In addition to developing an answer to our lead question, these workshops would serve several other functions. They would help us identify partners in a consortium that could then move on to the implementation phase. They would also help us to raise questions about academic organization and funding. Nationwide, and on many individual campuses, there are systems of support and funding for research in the sciences but no models for providing researchers in arts and humanities with what they need. The arts and humanities are unlikely to flourish if these basic aspects of national and campus functioning do not change.

In what follows, we will use several terms in ways that we would like to explain here.

Humanities. We will use this as a shorthand way to refer to the following range of academic disciplines: practice of art; the study of languages, literatures, and arts, past and present; philosophy; history; and humanistic inquiry in social-science disciplines such as anthropology. Research in the humanities refers to artistic achievement (for example, musical composition; sculpture; or choreography) as well as scholarly work.

Shared Technology Services. Here we refer to three interrelated ideas about information technology. The need to make core and common tools such as collaborative environments, digitization technologies, media repositories, digital libraries, text mining, natural language processing, visualization, website development, virtual environments, and the like readily available to humanists. The need to provide and expose these technologies as re-usable *services* that can be easily found and employed as building blocks for research by the humanities community. And the belief that technology is only as good as the human infrastructure that supports it – that is, staff and organization matter.¹

Bamboo. In the natural world, bamboo is a highly flexible organic material that serves multiple purposes: it can live as a single stalk on a desk or grow quickly into renewable forests; be used for constructing buildings or decorating them; become as strong as hardwood or as flexible as cloth; and can be lashed together to keep water out as in a boat or transport water as in a pipe. We envision our approach for arts and humanities digital services to be similar: configurable, flexible, sustainable, and reliable – hence the name, Bamboo. Bamboo has two phases: a planning phase and an implementation phase. Our present proposal is for the planning phase.

¹ The term “cyberinfrastructure” is sometimes used to capture parts of these ideas about technology, as well as others. See, for example, American Council on Learned Societies, *Our Cultural Commonwealth: The final report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities & Social Sciences* (December 2006) <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>.

II PERSPECTIVES OF FIVE COMMUNITIES

Our long-term goal for Bamboo is to develop shared technology services *to enhance humanities research now and in the future*. This planning project is a means towards accomplishing that end. As we have explained, the success of both the planning project and the implementation project depends on bringing together members of five communities – humanities researchers, computer science researchers, information scientists, librarians, and campus technologists. Each community has distinctive practices, lingo, assumptions, and concerns; and clearly there is much diversity within each community as well. In this section we portray perspectives from each community on the question we seek to answer and explain what motivates the communities to work together on this project. By drawing together these different perspectives we hope to start building a common language for collaboration.

The joint Berkeley and Chicago project team that prepared this proposal itself included leaders from these five communities. We have spent significant time talking with colleagues at our institutions as well as other campuses, and in November 2007, we hosted an all day focus group at the Townsend Center for the Humanities at Berkeley that brought together approximately 50 humanities faculty, computer science faculty, campus technologists, library staff, and others². In the text below we draw upon and sometimes quote from these conversations and the focus group.

II.1 ARTS AND HUMANITIES SCHOLARS

While a one-day focus group can make no claim to being comprehensive, it did clearly point to some issues that we will need to explore more systematically in the series of workshops we envision. We include below in parentheses the disciplines of those quoted and paraphrased.

IS IT LIKE DRIVING A CAR IN THE 19TH CENTURY?

A computer science colleague noted that in the late 19th century to drive a car was a foreign activity – both for drivers and those who experienced driving from afar. Paved roads did not exist, driving practices were rudimentary at best, and, in some cases, laws required that anyone who drove a “horseless carriage” had to have a man with a red flag walk ahead of them so that horses were warned when an automated carriage was near³. *Is the state of arts and humanities technology akin to driving in the 1890s?* For many in the humanities, computers today are like horseless carriages of the late 19th century—the ecosystem (roads, gas stations, part stores, repair shops, highway code, DMV, driver education, consumer reports) has not evolved so that ordinary users can take full advantage of their potential; and sometimes it even feels like we must wave a red flag warning that technology is now going to be used. We must try to understand what kinds of uses of technology will in the future become as ordinary for humanities researchers as driving a car, sending e-mail, and exchanging word-documents are to us now (analogy presented by a faculty member in Philosophy).

² This meeting was jointly led by the by the Dean of Arts and Humanities, Director of the Townsend Center for the Humanities, Director of Bancroft Library, Chair of Computer Science, and Director of Data Services within central IT.

³ See J. Frank Duryea, *Carriages Without Horses*, pg. 129, Society of Automotive Engineers, Warrendale, PA, 1993.

DIY OR “DO IT YOURSELF”

Clearly there are “low level” barriers to digital research that must be identified and removed. These are exemplified by the “broken podium” (Art History) on which a laptop rests insecurely. For a number of faculty, using technology in research is often DIY. A faculty member who wants, for example, to search, create, present, or share a collection of images must often invent his or her own digital tools with little by way of outside guidance or support.

“WHO WILL DO THE DATA SHOVELING?”

For many humanists, (for example, Literature, Music, History) building large digital corpora is a preliminary and necessary step if genuinely productive research is to be supported by technology. Yet building these collections is extremely difficult; for example, after 30 years of effort and several million dollars of investment less than 2% of the available Tibetan Buddhist texts have been digitized. Faculty would like not only to build larger digital collections but to be able to have readily available and adaptable tools for searching, extracting, and annotating the collections.

ENVIRONMENTS TO SUPPORT COLLABORATION AMONG SCHOLARS

Berkeley’s center for new music on campus showed how they used Drupal – an open source content management system – to build a complete environment for blogging, software distribution of music players and technology, informal publication, discussion groups, taxonomies, folksonomic annotation and visualization via “tag clouds”, and a sense of virtual place for the center.

HEROIC EFFORTS OF SMALL TEAMS: “...AND WE ARE EXHAUSTED”

After hearing this exciting presentation about the Drupal environment, several other humanities center directors and faculty wondered how they might acquire something similar. The faculty presenter (Musicologist and Composer) cautioned, “*we did it ourselves, but we are exhausted.*”

WHAT IS THE “OBJECT” OF PUBLICATION?

Common refrains included the publication process, new forms of publication, the nature of the “object” to be published, the evolution of the book, and barriers to scholarly dissemination and communication, including financial models and issues of intellectual property. We benefited from comments by librarians and by humanities researchers on these complex issues.

SCHOLARLY CONTEXT DRIVES THE DISCUSSION OF TOOLS

Other technology topics surfaced as well: new multi-dimensional digitization technologies; repositories to manage, preserve, and share images, video, text, and audio; technologies to create “marginalia”; wikis and collaborative writing; semantic services and search tools; sophisticated sensors to capture music and movement (Dance). The humanities researchers were interested in these topics, but only to the extent that they could imagine some relevance to their research needs.

DESIRE TO LEARN TOGETHER

There was keen interest on the part of humanities researchers to learn more about how other researchers, centers, and collaborative projects are using digital tools in research, publication, and teaching. Again, examples were very important.

PHYSICAL PLACES FOR DISCUSSION AND COLLABORATION

A number of humanists (from English, Musicology, Rhetoric, and Theater, among others), noting the characteristic isolation that comes with the individualistic emphasis of much humanities scholarship, stressed the importance of physical “places” in which partners can come together to learn in a “spiral of conversation” (Philosophy) about how to enhance art practice and scholarship through the use of information technology.

COORDINATION BETWEEN PROJECTS AND LARGER COLLABORATIVE ECOSYSTEMS

There was an engaged, hearty collaboration in our focus group among people from the four broad communities, and as Berkeley’s CIO put it, we were in a position to start “aggregating demand” for significantly greater resources that could be directed to the humanities. Still, there were expressions of weariness (Anthropology, Computer Science) with overly complex, bureaucratic collaborative structures. Many participants in the focus group expressed the hope that the Berkeley campus could support “lightweight” partnerships and small projects that could be connected together as part of larger a ecosystem (Art History, central IT) of collaboration. In this sense, the Bamboo metaphor of organizational flexibility resonated with a number of faculty.

TRANSFORMATION OF RESEARCH

We heard from faculty in dance and music about striking ways in which media and technology are transforming artistic practice. For example, tele-immersion allows a dancer to dance with herself or other dancers located far away. Other researchers imagined transformation coming from new forms of archiving, curation, and memory, and from “mash-ups” that could be built from massive collections of humanities and cultural heritage artifacts: text, images, audio, and video.

DISCUSSION OF ORGANIZATIONAL MODELS BUBBLED UP REPEATEDLY

Several humanities faculty pointed to the need for a technology “go-to” person who would have an understanding of the humanities as well as technical acumen and the ability to reach out across campus for resources and solutions. Several people suggested that there may be humanities graduate students who are excited about digital scholarship and who might fill such a go-to role.

THE CRITIQUE OF TECHNOLOGY BY THE HUMANITIES

Finally, some humanists struck a cautionary note. Part of the responsibility of researchers in the humanities is to take a critical stance towards technology, inquiring into its impact on the humanities, on individual lives, and on society as a whole. We need to ensure that when technology is that it expands and deepens the humanistic vision rather than supplanting it with flashy but superficial capabilities.

II.2 COMPUTER SCIENCE

Computer science research can be described as a steady march *away* from its elemental foundations in operations on bit sequences. Layer on layer of abstraction has been built on top of these operations, enabling researchers to work on questions that are of direct interest to potential lay users. As one focus group attendee put it, “Computers are to computer science as paint is to art.”

Within subfields of computer science such as natural language understanding, computer vision, information extraction, and information integration, the questions of interest and the technologies being developed are beginning to have semantic overlap with the questions of interest to humanities researchers – the two groups are, at least to some extent, finally talking about the same things. The prospect of collaborating with these researchers to find uses for existing technologies and

guide the development of new methods is tremendously exciting for computer scientists. We hope that an expanded awareness of what is, or might become, technologically possible may also enable humanities researchers to develop entirely new modes of inquiry.

We have, as part of the preparation for this planning proposal, polled faculty in Berkeley's Computer Science Division and the School of Information to gauge their interest in computing for humanities research. We received detailed replies, many describing ongoing projects, from over 20 faculty members. The range of computer science research topics covered is too broad to enumerate here, so we provide one example scenario of what would be possible in the future if we can find better ways to structure and sustain partnerships:

Over the centuries Buddhist monasteries have housed thousands of "books" of Tibetan Buddhist literature, each composed of the print from several hundred woodblocks. As was common in the early phases of the digital revolution in the humanities, many of these have been scanned – approximately four million page images are available at www.tbrc.org – but many have not. First attempts at transcribing and collating these vast collections and creating a simple index, all done by hand, have proved expensive and time-consuming. With an investment estimated to be in excess of \$1M, less than 2% of Tibetan texts have been input.

Now in our scenario, faculty members in Computer Science, East Asian Languages and Cultures, and South and Southeast Asian Studies team up to address the problem. First, the digital images are stored in the campus archiving repository, which provides improved speed of access, reduced costs, and a guarantee of permanence. Achieving the requisite level of accuracy will itself require the development of new OCR techniques by Computer Science Professor 1 (CS-Prof1) guided by syntactic and semantic models co-developed with East Asian Language and Cultures Professor 1 (EALC-Prof1). Metadata on authorship, woodblock location, etc., is added to the corpus.

Then, CS-Prof2 and EALC-Prof2 work together to develop a digital lexicon for the various styles of Tibetan used in the corpus. CS-Prof2's automated grammar learning system is used to create a probabilistic context-sensitive free grammar for Tibetan. As the proper semantic rendering of Tibetan is highly dependent on the mastery of a vast number of contexts and idiomatic usages, CS-Prof2's automated system enables the development of translation tools that dramatically reduce the amount of time required for scholars to master the language, and thus significantly increases the quality and quantity of translations.

EALC-Prof1 is excited to discover systematic patterns in the evolution of grammatical styles over time. Longstanding debates regarding the existence of "Old Tibetan" dialects are resolved by means of grammatical analysis of 1000s of texts in the corpus. With the help of a number of Sanskrit-Tibetan, Tibetan-English, and Tibetan-Chinese parallel texts, CS-Prof2, EALC-Prof1, and a linguistics researcher use machine learning techniques to create rough translation systems that enable automated translation among most of the canonical languages of the Buddhist tradition. These make possible the identification of several thousand cases where passages from the literature in one language (e.g. Tibetan Buddhist literature) turn out to have been borrowed wholesale from another (e.g. Sanskrit Saiva literature). These borrowings clarify many previously unresolved questions in the development of the Buddhist tradition in Asia, as well as providing a much larger set of parallel texts that enable more accurate translations.

The resolution of many pertinent historical questions lies in the identification and cross-correlation of key historical figures across a range of literature. EALC-Prof2 is interested in using the texts as a historical source, and works with CS-Prof2 and CS-Prof3 to apply their information-extraction technology to pull out basic historical assertions from the

corpus, as well as from other related corpora including those containing the writings of diverse ethnic groups of Silk Route travelers in Central Asia from the same period.

The technology is able to create a multilingual glossary of names and places, giving reliable identification of the many different ways in which the same name is rendered at different times and in different languages. For example, by compiling a large collection of assertions made about Padmasambhava, who is popularly thought to have “brought” Buddhism to Tibet in the 8th century, EALC-Prof2 is able to resolve long-standing questions as to the historicity and influence of this individual. Examination of other historical sources further contextualizes the development of Buddhism in Tibet.

Along with these content-related technologies, computer science researchers are interested in new, generic capabilities that cut across content areas and provide robust, usable, general-purpose services. Examples include the following:

- *Why can't I load this into that?* One of the most important usability problems is the proliferation of data formats and associated application software. This presents a huge problem of interoperability. For example, there are over 100 standards for encoding image data, many of them proprietary. The typical user experiences the interoperability problem many times daily when images fail to open in a browser or PowerPoint presentation and painstakingly transcribed foreign alphabets are replaced by nonsense symbols in someone else's version of Word. Generic operations (such as finding the absolute red level of the top left pixel of an image) and generic types (such as “image”, rather than “JPEG⁴ 2000 RGB 12-bit [with extensions] image”) are feasible, particularly using the ideas embodied in the XML data format, and can be made *extensible* to handle new formats and operations that arise over time. Collaborative research is clearly needed between computer science and the humanities to understand what types and operations are required.
- *What's the right software for doing this to that to get the other, and how do I apply it?* This is a traditional problem in software engineering where the noble goal of reusing existing (correct) software, instead of writing everything from scratch (with all new errors), has mostly foundered on the lack of effective support for describing the required functionality and finding the piece of software that provides it (or something like it). For users wishing to find the right applications and web services for their needs, and perhaps to compose them painlessly into new systems, the only current answer is to type something into Google and hope for the best. New techniques such as “specification mining and retrieval” work out what a piece of software does, for what types of objects, and make the results of the analysis easily queryable. It is likely that these will form an important part of the humanities computing ecosystem.
- *What if I don't know the answer for sure?* Optical character recognition systems, word identification systems, automatic translators, named-entity recognizers, information extraction systems, and databases can be tremendously useful in, for example, the task of developing a historical timeline from a multilingual corpus of original sources. These technologies are traditionally organized, however, as a pipeline, each stage of which must produce a definite answer – is this a kappa? Does this say Themistocles? Is this the same Themistocles? Was he archon in 493 BCE? In the real world of humanities research, uncertainty is often ubiquitous and locally (although perhaps not globally) irresolvable.

⁴ From the JPEG Wikipedia page: “The most common file extension for this format is .jpg, though .jpeg, .jpe, .jiff and .jif are also used.” As they say, you can't make this stuff up.

Computer scientists in the database area (with so-called *dataspace* models⁵) and in artificial intelligence (with so-called *open-universe* probability models⁶) have begun to create a radical generalization of classical information systems, one capable of simultaneously entertaining many uncertain hypotheses and their connections to the original evidence. With such capabilities, humanities researchers may be able to synthesize information from wider and more disparate sources than previously believed possible.

II.3 INFORMATION SCIENCE

One of the challenges of building an ecosystem of collaboration between humanities researchers and computer scientists, is how to share the development of advanced technologies by small research teams so that they are more broadly accessible. The focus of faculty in computer science is generally directed to the discovery and development of core technologies, rather than the processes needed to make them applicable and productive for customers. The necessary complements to the strengths of computer science faculty are professional experience, an understanding of enterprise workflows, and appreciation of the issues that arise with the dissemination of information technologies. These are precisely the emphases of schools of information science, where faculty and graduate students combine technology, sociology, law and business perspectives to provide deployable solutions for a broad range of audiences.

A promising model is thus likely to combine focused collaborative research to explore the feasibility of applying state of the art technology for a given problem, together with an information management and systems perspective on how to deploy the technology more widely. This will often involve some creative compromises to hide technical details, abstracting key aspects into configuration tools that are much easier to adapt than the underlying software. In other cases, a certain amount of performance or flexibility will be sacrificed to generalize the technology and make it work for a broader range of applications. This may be a multi-part collaboration among computer science, humanities, information science, and campus IT working to recognize what can and cannot be generalized, and how to make this productive for the various partners.

Another area of potential collaboration emphasizes recent advances in the deployment of technology for large online communities. Various referred to as “social media”, “community annotation” or more generally the *community production of public goods*, these approaches define architectures and infrastructure that allow a community to productively collaborate to generate metadata, to construct and manage knowledge-bases, or to curate expert resources ranging from software to ontologies. Examples range from Wikipedia to Freebase to open-source software to collaborative tagging of museum collections. The associated communities are often depicted as open, public and largely unstructured, however the same ideas can also be productively applied for more focused expert communities such as the faculty and students in a given academic discipline. Technologists can provide tools and infrastructure that allow communities in the humanities to construct and curate knowledge-bases (e.g., ontological models of historical language and culture, semantic indexes of related corpora, et al.).

⁵ See, e.g., Alon Y. Halevy, Michael J. Franklin, David Maier, “Principles of Dataspace Systems,” pp. 1-9, In Stijn Vansummeren (Ed.), *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, ACM Press, Chicago, Illinois, June 2006.

⁶ See, e.g., Brian Milch, Bhaskara Marthi, David Sontag, Stuart Russell, Daniel L. Ong and Andrey Kolobov, “BLOG: Probabilistic Models with Unknown Objects.” In *Proc. IJCAI-05*, Edinburgh, Scotland, 2005.

Finally, we note another point of view, often emerging from information schools, which we believe will be critical to our planning effort. This is the large-scale sociological, historical, and technological view of the developing international system of digital scholarship and scholarly information infrastructure. As Christine Borgman writes in her recently published book, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*:

Scholars in all fields are taking advantage of the wealth of online information, tools, and services to ask new questions, create new kinds of scholarly products, and reach new audiences. The Internet lies at the core of an advanced scholarly information infrastructure to facilitate distributed, data- and information intensive collaborative research. These developments exist within a rapidly evolving social and policy environment, as relationships shift among scholars, publishers, librarians, universities, funding agencies, businesses, and other stakeholders. Scholarship in the sciences, social sciences, and humanities is evolving, but at different rates and in different ways. While the new technologies receive the most attention, it is the underlying social and policy changes that are most profound and that will have the most lasting effects on the future scholarly environments. This is an opportune moment to think about what we should be building. Let the conversation begin.⁷

II.4 LIBRARY AND SCHOLARLY COMMUNICATIONS

Humanists (and librarians) have repeatedly drawn the analogy between the library and the laboratory. For the humanist, the library is his or her laboratory, the place in which is found, or hidden, the raw materials for research⁸. Therefore, libraries, both on campus and in the form of national and international consortia, need to be key players in conversations about digital scholarship and research.

Until recently, the laboratory setting of the library has meant collecting printed monographs and journal literature, but it has been clear for at least a dozen years that the Web and the creation of large digital libraries will change scholarship and learning dramatically. Hitherto resources to digitize large corpora (texts, books, images, recordings) have been modest in comparison to the size of the corpora. The availability of large-scale commercial funding is already having an enormous impact, driving down the cost of digitization to pennies per page and making available vast quantities of information in ways almost unimaginable even five years ago.

We do not yet know how the availability of such digitized corpora will change research in the humanities or to what extent such content will transform it. What we do know is that the analytical tools currently at our disposal are utterly inadequate for current and future researcher. For texts, word searches won't cut it and for images, we've barely developed even the most primitive of tools. We must find ways to manipulate these digital corpora in the aggregate, working with them not at the level of the individual text, book, image (at ground level), but with a view from 20,000 feet studying entire collections or cross-sections of content across multiple media types.

Among the numerous issues surrounding the use of digital resources, three seem to rise to the top within the context of this planning program. First, we have to acknowledge and clearly articulate a subset of the myriad and complex copyright issues surrounding digitized content. There are

⁷ Christine Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press, 2007; pages xvii-xvix.

⁸ John Unsworth, "The Value of Digitization for Libraries and Humanities Scholarship." Innodata Isogen Symposium, The Newberry Library, May 17, 2004; page 6.

many more humanities scholars focused on the period since 1923, and copyright is a major hurdle that needs to be addressed. Obviously, this is much larger than the scope of Bamboo, but the issues surrounding copyright must be included in a portion of discussion.

Second, we have to solve the persistence problem. It does scholarship no good to use or refer to a text, image, tool or system that may disappear from a website tomorrow. To solve the persistence problem, one needs to address the concept of trusted repositories and tools. Persistence is a dimension that should be considered as a core element of the planning program.

Finally, we have to solve the “unknown item” problem. Looking for a book (or service or tool) that you know exists is quite different from looking for information that may or may not exist. Library of Congress subject headings are, by today’s standards, inadequate as finding aids and the breadth and range of metadata surrounding objects and collections has become increasingly difficult to manage, especially within an interdisciplinary context. Libraries are tentatively beginning to realize that they have a role in the scholarly communications process that goes beyond merely receiving the end product of that process, the journal article or the scholarly monograph, and require standardized ways to share and exchange both data and metadata of objects deposited, entered or referenced within and beyond the collection. In addition, libraries are increasingly being asked to archive the results of the campus’s research activities that may not have been published elsewhere. Faculty who have sometimes devoted their entire careers to the creation of a corpus of materials (a database of medieval manuscripts, a corpus of linguistic raw material for a lexicography of languages from the Caucasus) must be able to place those materials, the raw materials of scholarship, in a repository for the benefit of their colleagues around the world.

The Library has much to offer Bamboo. The ethos of a librarian is quite different from that of the information technologist, much more focused on service and much more used, on the one hand, to the creation of sophisticated systems for information retrieval, and, on the other, to working one-on-one with the patron to find the information resources needed to carry out a particular research project. Like discipline- or domain-specific academic technologists, librarians act as bridges and translators between the world of the scholar and that of service and resource provider.

The Library also has much to learn from Bamboo. Historically, libraries have been focused around content, selection, preservation, and access, and have been seen as more of as being “service counters” than working laboratories. In this rapidly transforming world filled with digital information, how will this role change and what will stay the same? As digital collections extend beyond campus boundaries and evolve into cross-institutional meta-digital collections, what will be the role of the library? Will the validation of external content (content beyond collections and licensed resources) become a much more critical role for libraries? What will the new services be and how should they be implemented? For example, should libraries become active collaborators with researchers, beyond the traditional subject bibliographers, and engage more deeply as invested academic partners? Should the domain of libraries extend beyond content and into tools and services that act upon the content? Tools for searching and browsing are integrated into libraries today, but what about data mining and visualization? Do these more advanced tools and capabilities need to be as common as search is today? Are new services for the digital domain merely analogues of current services or will these new services need to be “invented” as something new to support tomorrow’s researcher?

Ubiquitous information access and open sharing of content, tools, materials, and services among scholars and researchers changes the role of the Library. To what extent is uncertain, but now is the time to engage in conversations and pilot processes and workflows among librarians, tech-

nologists, scholars and researchers in order to chart a path toward the next generation of digital research and scholarship.

II.5 CENTRAL INFORMATION TECHNOLOGY ORGANIZATIONS

Central information technology organizations at a number of universities and colleges are searching for better models to provide scalable technology services to enhance scholarship, but this is a challenge for these organizations⁹. The predominant focus of their resources and expertise in recent memory has been on campus infrastructure such as networking, data centers, and large enterprise-wide administrative systems. How to work with faculty from the arts to the sciences to add value to research in ways that can scale and be sustained – that is, without having to craft individual, almost artisanal solutions for each academic partner as short-term projects – has not been clear as a strategy or culture for most central IT organizations.

From the point of view of academic technologists, the landscape of scholarly projects for the humanities appears populated by many discrete technology projects large and small. At the highest level, academic technology projects tend to focus on domain-specific tools and resources (e.g. ARTFL Project¹⁰, Dictionary of Old English¹¹), integrated applications such as learning/collaboration management (e.g. Sakai¹², Bodington¹³), higher-education administrative systems (e.g. Kualu¹⁴), or digital asset collection, management and sharing (e.g. Fedora¹⁵, DSpace¹⁶). Beyond these projects are hundreds if not thousands of individual databases, websites, and web-based tools created by students, faculty and researchers around specific topics, interests, projects and/or initiatives. In all, these tools and resources are and were created to meet the specific needs of a particular community.

For a number of these projects the principal investigator from humanities becomes, at least partially, a software developer (or manager of software developers) who spends valuable time sorting through technology options, negotiating rights, and developing, testing, and maintaining code, rather than discovering new knowledge and advancing his/her discipline. In addition, the PI often needs to secure funding for one time software development and on-going maintenance of the project, as well to manage technical staff.

Complicating matters, the notion of sharing tools and/or data with other unrelated projects or systems is often foreign, and at best, is made possible through complex system- or tool-specific application programming interfaces (APIs). As a result, a scholar who has a specific research interest and wants to integrate tools and data from other projects is frequently left to figure out what resources might be available; and if some are identified, to negotiate the right to use the information and/or tools. In addition, there is a growing problem that, because of changing software

⁹ EDUCAUSE Center for Applied Research (ECAR) study, “IT Engagement in Research – Key Findings”. Harvey Blustain with Sandra Braman, Richard N. Katz, and Gail Salaway. EDUCAUSE, July 2006. <http://www.educause.edu/ir/library/pdf/EKF/EKF0605.pdf>

¹⁰ University of Chicago and Centre National de la Recherche Scientifique (CNRS), Project for American and French Research on the Treasury of the French Language (ARTFL), <http://humanities.uchicago.edu/orgs/ARTFL/>

¹¹ University of Toronto, Dictionary of Old English Project, <http://www.doe.utoronto.ca/index.html>

¹² Sakai Foundation, Sakai Learning and Collaboration Environment, <http://sakaiproject.org/>

¹³ Bodington Project, <http://www.bodington.org/>

¹⁴ Kualu Foundation, <http://kualu.org/>

¹⁵ Fedora Commons, <http://www.fedora-commons.org/>

¹⁶ DSpace, <http://www.dspace.org/>

definitions and standards and lack of ongoing maintenance, many humanities technology projects that succeed in reaching the status of a working system become non-working within a year or two and are never restored to full functionality. Most of the rest, while usable, are not re-usable as components of other projects that might want to build on previous work instead of reinventing it.

Reflecting on the current landscape of humanities computing, a number of IT strategists believe we now can find new methods to make the process of creating digital scholarship projects easier. Simplifying the way in which resources and tools can be reused across projects, disciplines, and institutions (whether such tools are locally developed or incorporate web services available on the Internet today) is fundamental to this endeavor. These campus technologists believe that three interrelated strategies are needed to make an evolutionary leap in digital scholarship:

- 1) The development of sustainable partnerships among researchers, instructors, libraries, domain specialists, and information technology professionals (a partnership that extends well beyond the boundaries of any one institution, domain, or region)¹⁷;
- 2) The creation of a set of core capabilities and services built upon common technology frameworks; and
- 3) The ability to embrace and use a blend of open-source and commercially-provided tools, resources, and services.

A successful partnership model is built on a foundation of common understanding about the problem space to be addressed. For many campus technologists (and for others well) a comprehensive understanding of what the nature of the work is – humanities scholarship – and where the work may be going does not exist. In the case of the humanities, this may not just be a problem of insufficient understanding on the part of technologists, but rather a broader absence of analytic models for scholarly practices and workflows in humanities research.

A key methodological approach (as well as set of technologies) that may serve as the impetus for evolution in humanities digital scholarship is to start from a perspective of services and service architectures. Service based approaches can be seen as coming from two quite different worlds:

- 1) the world of the large enterprise (whether corporate, governmental, or educational) with a set of service-oriented architecture¹⁸ (SOA) practices that emphasize scale, management, cost-effectiveness, and long-term stability; and

¹⁷ The basic notion of interdisciplinary partnerships to realize research goals through information technology extends back a number of years. In 1987 a panel report entitled *Visualization in Scientific Computing* recommended that interdisciplinary teams of computational scientists and engineers, visualization specialists, systems support personnel, artists, and cognitive scientists work together to tackle scientific visualization problems, in turn making “it much more likely the tools developed will be reused by other scientists and engineers in other fields, and that their use will then diffuse through their respective communities.” (“Visualization in Scientific Computing.” Bruce McCormick, Thomas A. DeFanti, and Maxine D. Brown, eds. *Computer Graphics*, ACM SIGGRAPH, vol. 21 no. 6, p11. November 1987)

¹⁸ In his 2005 article in *Science* magazine, Ian Foster defines service-oriented architectures as being *standard interfaces and protocols that allow developers to encapsulate information tools as services that clients can access without knowledge of, or control over, their internal workings. Thus, tools formerly accessible only to the specialist can be made available to all; previously manual data-processing and analysis tasks can be automated by having services access services.* See “Service-Oriented Science,” Ian Foster, *Science*, 308, pp 814-17.

<http://www.sciencemag.org/cgi/content/full/308/5723/814?ijkey=aqCCmCFix8Ll.&keytype=ref&siteid=sci>

- 2) what might be called the *Wild West* of data “mash-ups”¹⁹ coming from the Web 2.0²⁰ and cloud computing²¹ worlds in which ease, flexibility, and fast innovation (with a focus on the individual “consumer” rather than the organizational citizen) are paramount.

Common to both approaches is the idea of being able to re-use and weave together loosely-coupled, discrete, specialized technology services that come from other providers and projects rather than building and managing all on one’s own.

From the enterprise point of view, some in central IT believe the time is right to develop a light-weight, common, and easy-to-implement services architecture that could enable current and future scholarly tools and resources created across institutions and disciplines to be shared, reused, and maintained²². Critical to such an approach is the implementation of a web services framework. Such a framework is not a vertical application that focuses on a single in-depth function or a self-contained software tool used directly by a user, but rather a horizontally integrating set of technologies and set of core shared capabilities that enable the creation, aggregation, and reuse of services and resources among scholars, projects, and institutions. Frameworks of these sorts are appearing in a number of large-scale administrative domains within higher education.

Many questions abound regarding the design of such a services based approach. For example, how might campuses connect to and take advantage of the specialized services that come from digital asset management tools such as Fedora or DSpace, data analysis and mining tools such as SEASR, learning management systems such as Sakai or Moodle, digital library collections from Aquifer and the Open Content Alliance, and collaborative environments from a range of providers, such as Google and Six Apart? How can a services architecture help us to move towards a package of core and common services that can be provided to all scholars and/or to disciplinary or functional clusters? And how could these services be woven together with mapping, news feed, bibliographic, blogging, collaboration, and other social tools available from the Web 2.0 world?

From the commercial and Web 2.0 domains, applications are being created daily by “mashing-up” tools to create new derivative works. Examples are legion²³, and include among many options, layering shared community photos on top of Google maps or extracting crime data from city databases and related news stories and blending these together in neighborhood blogging sites. These mash-ups are based on a model of simplicity. Individuals create simple interfaces to complex systems, which in turn process information fed to them and transform results into equally simple outputs transmitted over the Web. The result is thousands of new applications and

¹⁹ A mashup is a lightweight tactical integration of multisourced applications or content into a single offering. Their primary business benefit is that they can quickly meet tactical needs with reduced development costs and improved user satisfaction. Quoted from the DMReview glossary. <http://www.dmreview.com/glossary/m.html>

²⁰ Today, Web 2.0 tends to refer to an online experience that is interactive, social, and data-focused. Tim O’Reilly defines Web 2.0 as software that adheres to seven basic principles, quoted here: 1) *The Web as platform*, 2) *harnessing collective intelligence*, 3) *data is the next Intel inside*, 4) *end of the software release cycle*, 5) *lightweight programming models*, 6) *software above the level of a single device*, and 7) *a rich user experience*. See O’Reilly’s article at <http://www.oreilly.com/lpt/a/6228>

²¹ The term “Cloud computing” has increasingly been used to refer to software as a service provided via the web by such companies as Google, Yahoo, Microsoft, as well as range of smaller firms. See, for example, John Markoff, New York Times: “Why Can’t We Compute in the Cloud?”, August 2007, <http://bits.blogs.nytimes.com/2007/08/24/why-cant-we-compute-in-the-cloud/>

²² For a more detailed perspective on the transformation of information technology from a services perspective, see “Describing the Elephant” by Ian Foster and Steven Tueke. *ACM Queue*, July/August 2005, pp 26-34. http://www.ogf.org/documents/Diff_Faces_foster.pdf

²³ As of December 8, 2007, 1,839 Web 2.0 mash-up websites were listed on Go2Web2.0. <http://www.go2web20.net>

services that use the Web as a *platform*. This new class of applications depends entirely on services provided by others to deliver niche functionality.

For all of the innovation and rapid deployment advantages, there are problems with the current incarnation of mash-ups that limit long-term use in research and education. Among the numerous issues, three come to the forefront:

1. Service providers have little knowledge of how or why a developer is using their tool, resource, or service. As a result, a revision to a service may adversely affect an application and they (the service provider) would be unaware of the impact;
2. In general, data exchanged between Web 2.0 services is open and free to be viewed and/or intercepted in transit. Data that is private or may have copyright restrictions associated with it becomes vulnerable; and
3. Locating services that may provide what you need can be difficult, especially if one does not register a service or API through commonly used websites.

In the end, today's mash-ups provide little to the scholar beyond enabling innovation; the problems around sustainability, reliability, data reuse, and tool discovery still exist. But when the idea of the mash-up is folded into the discussion around the academic enterprise, a number of questions emerge, some of which include:

- What if one could tie together the reliability of enterprise services with the rapid development and innovation model of mash-ups into a single framework?
- What if that framework could enable discovery of services and resources without needing to publish details within a single, centralized registry?
- What if the entire environment could include resources from across higher education and tap services and content available commercially?
- What if the needs of scholars came before technology?

These questions once again gives rise to an overarching issue for central IT: how can the organization bring value to the core mission of a university – research, teaching, and service – and do it in a way that blends enterprise scale and sustainability with the local partnerships and flexibility needed to work with the humanities?

III WHY BAMBOO AND THIS PLANNING EFFORT?

Reflecting on the five community perspectives, we believe now is the right time to proceed with the Bamboo planning project. By bringing together arts and humanities scholars, computer scientists, libraries, central IT, national scholarly technology projects, and others across our institutions, and by bringing together multiple interested campuses, we believe we can do a far better job in support of arts and humanities digital scholarship than is currently the case at many of our schools. Our experience is that a significant number of arts and humanities scholars, technologists, librarians, and others want to and are *excited* to learn and work together. In addition, we see a number of campuses and national consortia that are also looking for the opportunity to partner to make a difference in developing shared technology services for the arts and humanities. We may be at a natural tipping point for this kind of partnership.

The challenge, of course, is how to structure and encourage these collaborative efforts in the right way at our campuses and between campuses. The process and collaborative structures must both build richer understanding and community *and* lead to tangible benefits at the individual faculty, small project, campus, and national levels. It should create a situation where the whole is greater than the sum of the parts without injecting bureaucratic sluggishness, complexity, and collaborative fatigue. These are challenging balances, to say the least. We suggest that a community-driven planning process will help us grapple with and address these issues together.

What is needed is a deep, systematic, and organic understanding of humanities scholarship across multiple disciplines. Our collective assessment should chart scholarly practices and workflows, future evolutions, barriers, visions, and needs within and across disciplines. From this assessment we believe we can derive a number of commonalities, scholarly primitives, and natural clusters of activity across disciplines, and develop a deeper understanding of needs and services. Although there is some relevant literature in this area²⁴, and obviously rich understanding within disciplines and specific technology projects, we believe that the core set of practices, commonalities, and then needs is insufficiently understood; certainly insufficiently understood by technologists. A major goal of our planning will be to create a stronger analytic framework and documented understanding of the humanities to guide the development of a common services development.

The time is right to explore how to build an ecosystem of art and humanities and computer science innovation projects. Small teams of collaborative researchers could focus on solving important and potentially transformational problems for humanities scholarship, but do this as part of a lightly coordinated system of collaboration that helps to transition successful projects into longer-term sustainable and shareable services.

The time is also right to investigate how services-oriented architectures at the enterprise level (where the enterprise is both a campus and the higher education community as a whole) can provide a fundamental methodological and technological platform for exposing, sharing, and re-using tools and data services across arts and humanities projects internationally. Within this con-

²⁴ Again, see *Our Cultural Commonwealth: The final report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities & Social Sciences* (December 2006) <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>. It is interesting to contrast this analysis with, for example, the National Science Foundation's cyberinfrastructure vision and Nature's Science 2020 Roadmap.

text, we believe it is also important to explore the range of models and appropriate place of mash-up services and tap into cloud computing tools provided by companies such as Google and Amazon as well as other universities and research organizations.

Most importantly, drawing upon the perspectives noted above, the time is right to identify how to deliver a core and common set of easy-to-use digital services for humanities scholarship across disciplines and within natural clusters of scholarship, and engage in community discussion around models for staffing, intra-campus collaboration, inter-institutional partnerships, and financial support required to sustain cyberinfrastructure for the arts and humanities.

IV WHAT DOES SUCCESS LOOK LIKE?

To explain our goals and approach for the planning process it is important to distinguish, again, between the future implementation phase, Bamboo implementation project (which would optimistically begin after the completion of the 18 month planning process) and the Bamboo planning project. Our long-term goal for Bamboo implementation project is to collaboratively develop shared technology services *to enhance arts and humanities scholarship now and as the scholarship evolves in the future*. This planning project is a means towards accomplishing that end.

Our goal for Bamboo planning project is to identify, through a series of facilitated workshops carried out over 18 months, a blueprint and core set of institutions to build and deliver shared services for the arts, humanities, and interpretive social sciences. To accomplish this we have defined six objectives for the planning process, which follow and support each other in a progressive sequence of activities:

1. Carry out a community planning and design process in which the community of participants determines answers to the key questions.

Fundamental to the success of both the planning project and the potential longer term Bamboo consortium is our ability to bring together at least five critical communities in ways that add value to each community's work, to our individual campuses, and for higher education in general. We will employ a community design process organized around a series of facilitated workshops that build upon each other. The participants will progressively reach decisions regarding the key questions we have posed for the project and shape the direction of the effort. After each workshop, participants will decide if they want to commit to the next round of planning and participation; if so, they help to prepare analysis, ideas, and materials to be taken up in the following workshop. With each round of workshop participation both the opportunity to shape Bamboo implementation project and the responsibility for participation incrementally increase.

Our community planning process is modeled after the transformative methods that the Quali project, and the University of British Columbia in particular, has demonstrated in its successful efforts to build a balanced, multi-institutional collaboration to develop shared software services for the student experience in higher education. In the next section of the proposal, we provide more detail on how this approach applies to Bamboo.

2. Develop a deep and structured understanding of humanities scholarship and artistic practice which can be used to guide technology services design and organizational models for delivery.

From the start, we emphasize that the planning process will be fundamentally driven by constructing a rich understanding of the practices, cultures, needs, and commonalities of arts and humanities scholars across the full spectrum of literary studies, art and artistic practice, musicology and composition, theater, dance and performance, historical analysis, language study, cultural interpretation, new media, philosophical endeavor, and the like. Our planning process is first and foremost an effort to engage and understand who and what humanities scholarship is and where the central strands of research and creative practice are evolving. This will be the focus of sustained and iterative work over and in between the first two workshops. We will use a number of self-reflective, modeling, and ethnographic methods to engage scholars, technologists, and others together to define scholarly workflows and goals/visions within different and across humanities

disciplines as a whole. With this information we propose to derive commonalities and unique requirements related to practices, functions, barriers, needs, and existing and potential transformations at the disciplinary level. From this work we will publish a document tentatively titled, *Scholarly Practice in the Arts and Humanities: Foundations for Service Development*.

In other words, we seek to start first with the question of “who are we?” rather than skipping prematurely ahead to the question of “what do we need?”

3. Develop a community-endorsed technology services roadmap for scholarship

Using our understanding of scholarly practice in the humanities as a foundation, we will collaboratively develop a roadmap for shared technology service development. Here we propose to address a number of key questions:

- *Are there a core and common set of digital services that could benefit all arts and humanities scholars?*
- *What are services that will bring high value to “clusters” of faculty who carry out related practices and functions?*
- *How can we make digital services of both tools and data easier and simpler for scholars to use?*
- *Where are the best opportunities for innovative collaborations between arts and humanities scholars and computational science researchers, and how might these experimental partnerships fit into a service sharing model so results can be distributed to the larger community?*
- *How can we make use of the services and work coming from national consortia focused on scholarly technologies, community sourcing, and services architecture, such as Fedora, Sakai, Fluid, DSpace, Quali and others?*
- *How can we (and when should we) use tools and information services coming from cloud computing providers from industry, education, research and the like?*
- *How can we develop and use web services frameworks so that we can make it possible to bring together tools and data sources from communities locally to internationally?*

Underlying these discussions is the fundamental premise that we are looking for ways to share, use, and blend together the best tools and information sources from across many arts and humanities software projects, national consortia, digital libraries, and providers, and that there is high value in shared services that can be easily coupled together.

Answering these questions will occur over several workshops, in particular workshops two, three and four. In these workshops we will enlist the assistance of professional facilitators and educators in the services domain who can help us to understand what it means to identify and design a services based approach that builds from our models of scholarly practice. We also recognize that the questions posed above are complex, ambitious, and difficult. We’re not seeking perfect answers, but rather reasonable strategies and decision-making criteria that can help us move ahead. From this work we will publish *An Arts and Humanities Digital Services Roadmap*.

4. Identify organizational, staffing, and partnership models to support the on-going provision of these services.

After focusing scholarship and shared services, the next step will be to surface and discuss organizational, staffing, incentive, and funding issues essential to encouraging and improving sustainable and scalable digital scholarship locally, regionally, nationally, and internationally. We propose to address these questions more systematically in the later stages of our workshop series (specifically workshops three, four, and five) so that we can identify a range of practical models for organization and service delivery that could benefit different types of institutions, environments, and partnerships. We propose to take a “light touch” to these questions, namely identifying key issues and developing several approaches without seeking a single or definitive set of answers, attempting to solve problems that are specific to a particular institution, or premature for a partnership-based approach. From these discussions we produce a brief tentatively entitled, *Organizational Issues and Options for Sustainable Digital Services in the Arts and Humanities*.

5. Identify a community of collaborators who want to work toward the vision of Project Bamboo

By the end of the workshops, a diverse community of collaborators from across a spectrum of institutions should be ready and able to commit to the community-defined requirements for Bamboo implementation project. These partners will have demonstrated that they have the both the desire and leadership from arts and humanities scholars, technologists, and libraries at their home institutions. We expect that in order to move ahead with Bamboo, there will need to significant resource commitments made by partners who decide to participate in the Bamboo consortium. However, and we stress this, Bamboo is not meant to be exclusionary (for example, open only to well off institutions), and as such, these commitments will be proportional to the resources and capabilities of the participating institutions and the community will define and agree upon these requirements as part of the planning process.

6. Produce a proposal to build and share the Bamboo implementation project

Our final objective for the planning project will be to produce a detailed argument and blueprint for the development of Bamboo in the form of a follow-up proposal to the Andrew W. Mellon Foundation. Building from the reports outlined above, this proposal will describe what will be needed, detail requirements and plans, and define the value of an investment in a shared, services-based architecture to support of humanities scholarship. We expect that such a proposal will cover an approximately three year time span and require substantial cost-share from the institutions that move ahead with the Bamboo implementation project after the completion of the planning effort. Those institutions that wish to move on to the implementation project will need to ensure that key arts and humanities leaders on their respective campuses endorse the strategy and commitment outlined in the implementation proposal.

V SCHEDULE AND SUMMARY OF DELIVERABLES

As we prepared this planning proposal we were strongly encouraged by the Mellon Foundation to develop a planning process that is driven by arts and humanities scholars and leaders; that is representative of different types of higher education institutions; and that builds from the lessons learned from other multi-campus software collaborations.

This planning process starts with the perspective and ends with the ratification of arts and humanities scholars. In an effort to be representative, our workshop process has been designed to be multi-disciplinary, multi-institutional, and international. The workshops must form a collaborative community inclusive of arts and humanities scholars, technologists, and, as needed, facilitators who can translate the semantics and frameworks of the various parties. As the process evolves and deepens, dialog will need to reach a critical point where arts and humanities scholars begin to shape technology options by questioning impacts of potential technological choices, clarifying misinterpreted goals and ultimately co-determining a roadmap of goals to pursue, tools to provide, platforms on which to run, and architecture to use.

In an effort to learn from past efforts, we spent substantial time speaking with the senior leadership of the Quali Student Project at the University of British Columbia, from which we gained significant insight. We learned of both the effort it took to bring all the critical partners together as well as the powerful results they achieved when their partners learned to communicate with a common vocabulary and set of concepts. The Quali Student leadership strongly encouraged us to utilize a professional outside facilitator with experience in leading a services based approach. We have taken their advice and will ask professional consultants to facilitate significant portions of the workshops.

The initial workshop will be open to all interested parties. Participation and workload requirements will incrementally increase throughout the planning process. We expect these increases in commitment to cause some institutions to limit participation in later workshops, naturally selecting those institutions capable of committing to the cost and/or resources required to lead the future Bamboo Implementation Project.

At the end of each workshop, small inter-institutional teams will be formed to complete specific assignments based on the workshop training. These assignments will be used in the written reports as well as for shaping the exercises and discussions of subsequent workshops.

Between workshops, pilot projects will be designed and implemented to demonstrate potential uses of services-facilitated technology in the arts and humanities domain. We have found that faculty are very effective at critiquing a prototype and such critique has served as a low cost means of clarifying issues as well as exposing misunderstood goals and objectives before moving into a full-scale development mode.

CAVEAT

It should be noted that the Bamboo Planning Project is designed to be community driven. As such, the current workshop agendas and assignments should be seen as initial plans. As this planning process unfolds, workshop agendas and assignments may be altered to address community concern.

V.1 CHICAGO AND BERKELEY BAMBOO PLANNING PROJECT LEADERSHIP MEETING

PRELIMINARY DATE

April 2008

SUMMARY

In early April, we will conduct a one-day leadership meeting with senior institutional stakeholders from the University of California, Berkeley and the University of Chicago. This meeting will formally kickoff the community planning effort as well as clarify any institutional criteria or requirements that need to be accounted for in the planning process.

V.2 WORKSHOP ONE: PLANNING PROCESS AND UNDERSTANDING ARTS AND HUMANITIES SCHOLARSHIP

PRELIMINARY DATES

Late April 2008 at the University of California, Berkeley; early May 2008 at the University of Chicago; and mid-to-late May 2008 at the University of Chicago Center in Paris.

SUMMARY

We will conduct three identical three-day workshops in Berkeley, Chicago and Paris where arts and humanities scholars will enter a facilitated dialog with technologists to help each other understand current arts and humanities scholarship practices as well as the future directions and visions of various disciplines. We will use this new understanding to explore and identify commonalities across various disciplines.

A substantial portion of this workshop will be used for training participants and completing hands-on exercises to identify and define common scholarly activities and the tasks required to complete these activities. In business analysis parlance, activities are commonly referred to as *workflows*. In our context, we will be assessing scholarly workflows.

For example, *searching for literature* might be considered a scholarly workflow that is comprised of many tasks; these tasks may include selecting words on which to search, selecting databases and repositories to search, assessing the quality of a returned search, saving and documenting specific search results, etc.

From the training and exercises, we hope to identify a common set of scholarly workflows as well as knowledge of outlier workflows. With this understanding of current scholarship practice, future directions of scholarship, and identified workflows, we will produce a draft report on *Scholarly Practice in the Arts and Humanities*.

After the Workshop One series, a small set of pilot projects will be initiated by the program directors, services architects, and developers to demonstrate uses of technology in the arts and humanities domain. These pilot projects will be presented at subsequent workshops.

V.3 WORKSHOP TWO: IDENTIFYING SERVICES

PRELIMINARY DATE

July 2008

SUMMARY

Workshop Two will be a three-day workshop that builds off of the training and assignments of Workshop One. We will analyze the results of the scholarly workflows assignments. This discussion may demonstrate multiple ways to model similar workflows as well as highlight the differences between various workflow models.

By evaluating these similarities and differences, we should be able to identify which tasks are core to each workflow (i.e. they appear in every model), which tasks are specialized (i.e. they may only appear in one discipline), and which tasks are hidden or assumed and consequently need to be made explicit. Once these tasks are identified, participants will be trained to leverage their models so that they can be used for more than one scholar and/or discipline.

Abstracting these tasks and defining their input and output options will allow us to identify *reusable services*. Services are mechanisms that use a defined set of input, follow a defined sequence of tasks, and produce a defined set of results. Technology designed to fully leverage services is referred to as *Services-Oriented Architecture (SOA)*.

We will provide several days of basic training on SOA facilitated by a professional consultant. Throughout the training, participants will complete hands-on group assignments around services relevant to arts and humanities, and discuss how the application of SOA might best enhance humanities scholarship.

After the workshop, program leaders, services architects and developers will build pilot projects to demonstrate uses of technology in the arts and humanities domain. These pilot projects will be presented at subsequent workshops.

V.4 WORKSHOP THREE: SELECTING SERVICES

PRELIMINARY DATE

October 2008

SUMMARY

Workshop Three will be a three-day workshop. Building off the training and assignments of Workshop Two, we will analyze the results of the Services Identification assignments. This discussion should demonstrate a variety of ways to model similar services. It should also highlight the differences between various services design choices.

A professional facilitator will provide several days of additional training. Drawing from the various models, we will select the best descriptive terms, input types, and output types to define the most useful and reusable service. Once defined, these services can then be organized and ordered so as to make them most effective.

After the workshop, program leaders, services architects and developers will build pilot projects to demonstrate uses of technology in the Arts and Humanities domain. These pilot projects will be presented at subsequent workshops.

V.5 WORKSHOP FOUR: ORCHESTRATING SERVICES AND DEFINING PARTNERSHIPS

PRELIMINARY DATE

February 2009

SUMMARY

Workshop Four will be a three-day workshop. Building off the training and assignments of Workshop Three, we will analyze the results of the Services Evaluation assignments. A professional facilitator will provide additional training on how to organize and order these services to make them most effective.

The *organization* and *ordering* of services is called *orchestration*. Throughout the training, participants will complete hands-on group assignments around services orchestration relevant to Arts and Humanities.

Finally, we will review the latest draft of the *Services Roadmap* to discuss the initial resources necessary to implement it. We will draft an initial set of criteria required to become a foundational partner to implement the services roadmap.

V.6 WORKSHOP FIVE: FORMING A CONSORTIUM

PRELIMINARY DATE

April/May 2009

SUMMARY

A three-day workshop will be held for key stakeholders from each interested institution desirous to commit to the criteria defined in the *Bamboo Partner Criteria*. This workshop will formalize foundational partners for implementing the next steps of the *Arts and Humanities Services Roadmap* report.

This workshop will walk through the methodology of our workshop process and also our findings as reflected in our reports: the *Scholarly Practice in the Arts and Humanities* report, the *Arts and Humanities Services Roadmap*, and the *Bamboo Partner Criteria*.

After providing final feedback to the community-proposed roadmap, participants will need to evaluate the roadmap in light of organizational and resources costs. We will discuss local and national organizational models currently supporting humanities scholars. We will then assess what personnel and/or roles are needed to implement this roadmap at individual institutions and consortium-wide. We will design an initial organizational model(s) that would support these defined personnel: their roles, function, and communication channels. We will then provide finalize specific institutional contributions, effort and resource requirements to serve as a founding member of this consortium. At this point, each institution will decide its level of involvement.

Institutions committed to be founding partners, will prepare an outline of a Bamboo Implementation Proposal to be submitted to the Andrew W. Mellon Foundation, including required institutional documents. We will assign writing tasks, deadlines and follow-up meetings to complete this proposal.

VI LEADERSHIP COUNCIL, RELATED INITIATIVES AND PROPOSAL CONTRIBUTORS

The success of Bamboo and the planning process largely depends on the contributions of its partners. Recognizing the wide range of experience partners can bring to Bamboo, we are seeking to engage groups, institutions, and individuals from a broad range of backgrounds, starting with the two sponsoring institutions and their work on related initiatives at each campus.

VI.1 LEADERSHIP COUNCIL

Given the wide range of disciplines, areas research, and domains of activity that will be explored during the 18-month period of the planning program, we felt it necessary to establish the Leadership Council that will help guide activities and keep the program focused on its objectives. The Council will meet formally at the beginning of the planning program and then on an as needed basis, either physically or virtually, over the duration of the project.

The Council is made up of the principal investigators as well as key representatives from both the University of California, Berkeley and the University of Chicago. For arts and humanities perspectives beyond that of the co-PI and Dean of the Arts and Humanities at Berkeley, Janet Broughton, we have turned to the center/institute for the humanities on each campus for leadership, and included the directors on the Council: Anthony Cascardi of the Townsend Center (UCB) and James Chandler of the Franke Institute (Chicago). Aside from bringing their personal experiences as scholars and researchers, they will assist with understanding the breadth of singular and interdisciplinary issues facing the arts and humanities today and into the future.

To reflect the computational, computer, and information science perspectives, we have solicited the support and input of respected leaders at each campus and within their fields, Ian Foster (Chicago) and Stuart Russell (UCB). Each brings considerable research experience in computer and computational science including Grid and cloud computing, high-performance computing, service-oriented architectures, technology-enabled collaboration, and web services.

Since libraries play a central role in arts and humanities scholarship, it seemed natural to include library representatives from each campus. Charles Faulhaber of the Bancroft Library (UCB) and Judith Nadler of the University of Chicago Library will bring critical perspectives to the Bamboo conversation, that of digital collections, open repositories, and archiving along with the evolving and rapidly changing role of academic research libraries in higher education.

Finally, the perspective of campus information technology organizations and the role enterprise IT can and does play in supporting the academic missions of institutions shall be provided by co-PI, Gregory A. Jackson (Chicago), and Shelton Waggener (UCB). Both are the chief information officers (CIOs) of their institutions and respected IT leaders in higher education.

The program directors, David Greenbaum (UCB) and Chad Kainz (Chicago), will be members of the Council in an *ex officio* capacity.

VI.2 RELATED INITIATIVES

UNIVERSITY OF CALIFORNIA, BERKELEY

Lead Partners include the Arts and Humanities Division, Townsend Center for the Humanities, Department of Computer Science, Bancroft Library, and Information Services & Technology.

The University of California, Berkeley's participation in Bamboo will be steered and coordinated by the new Humanities and Arts Research Technologies (HART) Initiative, with technical leadership and support coming from the Office of the Chief Information Officer (CIO) and the Information Services and Technology (IST) Division.

HUMANITIES & ARTS RESEARCH TECHNOLOGY INITIATIVE (HART)

In Spring 2007, the campus community formed HART. HART is led by Janet Broughton, Dean of the Arts and Humanities (Professor of Philosophy); Sue Schweik, Associate Dean, Arts and Humanities (Professor of English); Anthony Cascardi, Director of the Townsend Center for the Humanities (Professor of Comparative Literature); Charles Faulhaber, Director of the Bancroft Library (Professor of Spanish); Stuart Russell, Chair of Computer Science (Professor of Computer Science); Merrill Shanks, Associate CIO for Arts, Humanities, and the Social Science (Professor of Political Science); and David Greenbaum, Director of IST-Data Services Department and Director of the Open Knowledge and the Public Interest (OKAPI) Project.

Members of the HART Steering Committee are soliciting interest and feedback around Bamboo from a wide range of arts and humanities' faculty, computer science faculty, and other campus technologists, as well as assessing how to integrate the work of other campus Mellon-funded initiatives into Bamboo. HART's core goals are:

- 1) Provide substantial new support and leadership for digital technologies for the arts and humanities at UC Berkeley.
- 2) Help arts and humanities scholars to understand technology opportunities and technologists to understand and support common and unique technology needs of humanists. Build a community that enables cross-disciplinary collaborations.
- 3) Develop a rich set of shared technology services for the arts and humanities so that each scholar/group does not have to create their own siloed solution. Draw on the best technologies from on and off the Berkeley campus.
- 4) Create new collaborative organizational models between Arts and Humanities Division, Centers, IST, the Library, and other central providers to deliver and sustain these services.

BERKELEY TECHNOLOGY FOR SCHOLARSHIP

With the reformulation of the campus's central IST Division and CIO's Office in 2006, CIO/IST has identified as one of its four campus priorities a renewed focus on coordinated *technology to enhance scholarship*. The new IST-Data Services (IST-DS) department is helping Berkeley to coordinate this campus wide focus on cyberinfrastructure, especially those shared services that will facilitate information management/media vault and semantic services, data creation and sensor technologies, visualization and analytic tools, museum informatics, and a broad range of collaborative technologies and social software. Campus partners that IST supports and/or works with in these efforts, and who are likely be partners in HART/Bamboo, include such departments as Art History, Archeology/Anthropology, Architecture, Berkeley Language Center, Biology, and

the Berkeley Natural History Museums, Educational Technology Services (Sakai), Berkeley Libraries, and the Quali/Student Systems 2012 effort.

UNIVERSITY OF CHICAGO

Lead partners include the Division of the Humanities, University of Chicago Library including the Digital Library Development Center, Franke Institute for the Humanities, the Computation Institute, and Networking Services & Information Technologies (NSIT), the central IT organization of the University of Chicago.

In detail, NSIT Academic Technologies will act as the partner organization for Bamboo and liaise with the various campus IT organizations, units, scholars, projects and activities including the *Digital Humanities & Computer Science Colloquium* (see below). At present, campus partners include those listed above as well as Biological Sciences Information Services and units within NSIT including Administrative Systems, General Services, Data Center Services and Scholarly Technology & Research Computing.

DIGITAL HUMANITIES & COMPUTER SCIENCE COLLOQUIUM

In 2006, the Division of the Humanities, Department of Computer Science, Computation Institute and the University of Chicago Library co-organized a colloquium with Northwestern University and the Illinois Institute of Technology (IIT) with a goal to:

*...bring together researchers and scholars in the Humanities and Computer Sciences to examine the current state of Digital Humanities as a field of intellectual inquiry and to identify and explore new directions and perspectives for future research.*²⁵

This jointly organized event that resulted from a shared interest from within humanities, computer science, and libraries attracted papers and presentations from around the world and from a broad range of organizations and institutions including: University of Illinois at Urbana-Champaign; IIT; Michigan State University; University of Chicago; Loyola University Chicago; University of Victoria; Tufts University; University of Illinois at Chicago; Johann Wolfgang Goethe-Universität, Frankfurt am Main; University of Virginia; MIT; Yale; University of Wisconsin-Madison; Centre d'Études Supérieures de la Renaissance; Northern Illinois University Libraries; George Mason University; NEH; CourseForge; Vassar College; Marist College; Los Alamos National Laboratory; Pacific Northwest National Laboratory; Oak Ridge National Laboratory; University of Michigan; Distributed Proofreaders Foundation; University of Southern California; University of California, Santa Barbara; University of Toronto; University of Maryland; The Metropolitan Museum of Art; and the New Zealand Text Centre.

The success of the first colloquium and ongoing commitment to support this regional endeavor²⁶, illustrates the combined interest and interdisciplinary investment at the University of Chicago to further explore the connections among scholars, researchers, computer scientists, librarians and

²⁵ From the *Chicago Colloquium on Digital Humanities and Computer Science: What to do with a Million Books?* November 5-6, 2006. http://dhcs2006.uchicago.edu/call_for_papers

²⁶ The second event, *Chicago Colloquium on Digital Humanities and Computer Science: Querying Text and Image Archives for Collaborative Scholarship*, centered around a theme of “exploring the scholarly query potential of high quality text and image archives in a collaborative environment.” The event was held at Northwestern University on October 21-22, 2007. Once again, this effort was supported by the University of Chicago, the Illinois Institute of Technology and Northwestern University.

information technologists both on our campus and across higher education. Bamboo will benefit greatly from this existing activity and further reinforce importance of exploring and moving forward with next-generation digital humanities projects both on campus and beyond.

SCHOLARLY PROJECTS INFRASTRUCTURE INITIATIVE

In 2003, NSIT Academic Technologies was asked by the Provost, Deputy Provost for Research, Board of Computing Activities and Services (faculty board), CIO, and other senior administrators to explore the growth of computation in research, and the related need for skilled research application support. Rather than focus on the obvious area of high-performance computing, emphasis was specifically directed toward research topics and programs that have not traditionally sought or required high-end, large-scale computational capabilities. Coincidentally, an informal internal assessment of infrastructure needs, service requirements, and future directions was underway to chart the direction for the next half-decade of Academic Technologies. The confluence of these seemingly disparate activities along with a shift in learning technology needs led to a number of actions, one of which was the launch of the Scholarly Projects Infrastructure Initiative (SPII).

The SPII aimed to provide much needed infrastructure for piloting projects, hosting emerging applications, and sharing data in support of research and education at the University of Chicago. Unlike previous central IT efforts to develop a campus research resource tuned to high-performance computational problems (massively parallel computation center, high-performance computing cluster, etc.), SPII focused on creating a shared, distributed, and layered infrastructure driven by accessibility and utility that could be applied to any academic problem by any member of the campus community whether student, faculty or staff.

More than just a one application or single computational model, the SPII infrastructure, based on both open-source and commercially-available technologies, was viewed from the start to be usable beyond the boundaries of one project, discipline, or campus. As a utility, virtualization, and services framework, SPII was envisioned to be capable of being deployed by other institutions interested in providing generalized computing resources to scholars.

The overarching goal of SPII was to create an affordable, flexible, and scalable academic utility computing architecture that could be adopted by others. Once adopted, it could then be combined across institutions to enable application sharing, encourage collaborative development and use of tools across disciplines, and promote inter-institutional provisioning of computational resources to help address the ever-expanding demand for IT resources in support of scholarship – hence the connection with Bamboo.

VI.3 CONTRIBUTORS

This proposal represents significant input, support, effort and guidance from a number of faculty, staff, scholars, technologists, and researchers from across higher education and industry as well as organizations such as Quali and EDUCAUSE. In addition, we would like to gratefully acknowledge the following people who are among the myriad individuals that have shared thoughts and ideas over the last 18 months or contributed directly to this effort:

Sophia Accord (UCB)	Emily Baker (Chicago)	Janet Broughton (UCB)
Maneesh Agrawala (UCB)	Tom Barton (Chicago)	Ed Campion (UCB)
Danielle Allen (Chicago)	Robert Bartlett (Chicago)	Cathy Carson (UCB)
Greg Anderson (Chicago)	Charles Blair (Chicago)	Anthony Cascardi (UCB)
Michael Ashley (UCB)	Arno Bosse (Chicago)	Kaylea Champion (Chicago)
Bob Badal (Jamestown College)	Rich Breen (Chicago)	James Chandler (Chicago)
Ruzena Bajcsy (UCB)	Lois Brooks (Stanford)	Diane Conaster (Chicago State)

Meg Conkey (UCB)
Julie Cox (UCB)
Ian Crew (UCB)
Chas DiFatta (Carnegie-Mellon)
Helma Dik (Chicago)
Quinn Dombrowski (Chicago)
Shirley Dugdale (DEGW)
Jim Farmer (Georgetown)
Charles Faulhaber (UCB)
Ian Foster (Chicago)
Martin Foys (Hood College)
Michael Fary (Chicago)
Andrew Garrett (UCB)
Duffy Gillman (Arizona)
Sharon Goetz (UCB)
Deniz Gokturk (UCB)
Ken Goldberg (UCB)
Serge Goldstein (Princeton)
David Greenbaum (UCB)
Gary Handman (UCB)
Andrew Harrison (DEGW)
Chris Hoffman (UCB)
Gregory Jackson (Chicago)
Sally Jackson (UIUC)
Shannon Jackson (UCB)

Michael Jinks (Chicago)
Rosemary Joyce (UCB)
Tim Kachel (Jamestown College)
Chad Kainz (Chicago)
Mark Kaiser (UCB)
Rick Kern (UCB)
Kurt Keutzer (UCB)
Dan Klein (UCB)
Celeste Langan (UCB)
Ray Larson (UCB)
Greg Levine (UCB)
Elisabeth Long (Chicago)
Roberto Marques (Chicago)
Donald Mastronarde (UCB)
Patrick McGrath (UCB)
Martha Merritt (Chicago)
Rich Meyer (UCB)
Jonathan Miller (Chicago)
Margaret Mitchell (Chicago)
W.J.T. Mitchell (Chicago)
John Moses (Chicago)
Judith Nadler (Chicago)
Barbara Nelms (Chicago)
Mark Olsen (Chicago)
Tod Olson (Chicago)

Kathryn Powell (Manchester)
Martha Roth (Chicago)
Stuart Russell (UCB)
Ken Sadowski (Chicago)
Patrick Schmitz (UCB)
Alice Schreyer (Chicago)
Merrill Shanks (UCB)
Jeffrey Skoller (UCB)
David Trevvett (Chicago)
Ruth Tringham (UCB)
Niek Veldhuis (UCB)
Frederique Van Till (JISC)
Alexander Von Rospatt (UCB)
Shelton Waggener (UCB)
Sara Ware (Chicago)
John Wawrzynek (UCB)
David Wessel (UCB)
Brad Wheeler (Indiana)
Erik Wilde (UCB)
Lisa Wymore (UCB)
Adrian Freed (UCB)
Steve Masover (UCB)
Sue Schweik (UCB)