10.3 Humanities Faculty Committee and Technical Advisory Committee for Bamboo Technology Project ................................................................................................................. 61
11. Reporting and Evaluation .................................................................................................................. 62
12. Budget Narrative ..................................................................................................................................... 64
13. Appendix 1: Detailed Work Plan ......................................................................................................... 65
Proposal Summary

A consortium of ten universities, led by the University of California, Berkeley, will carry out an 18-month technology project to develop and design applications and shared infrastructure for humanities scholars and projects. We request $1.2 million dollars from the Andrew W. Mellon Foundation and will provide an equal amount of institutional cost-share. This project builds on the major planning effort carried out by Project Bamboo with the generous support of the Mellon Foundation.

This 18-month project is the first phase of a proposed three-year project. Over a three-year period, we aim to create two sets of applications that will directly support the research and teaching of large number of scholars in the humanities. “Bamboo Work Spaces” and “Bamboo Corpora Space” will meet needs that were identified during the planning process as core needs for scholarship in the humanities. The first will provide basic and easy-to-use tools for managing and analyzing content; it will serve a wide range of individual researchers and groups of scholars. The second will allow individuals or groups to work on dispersed digital corpora using a range of sophisticated curatorial, analytic, and visualization tools and services.

These applications will be supported by two shared infrastructure projects that will help technologists who work with researchers in the humanities. The first infrastructure project will be the development of an evolving set of scholarly web services on a services platform that will help technologists create and support a wide range of tools for scholarship. The second will be the adoption and dissemination of standards and services for interoperability. Together, these infrastructure projects will enable IT organizations, libraries, and digital-humanities projects to leverage one another’s accomplishments while ensuring that the world’s ever-increasing store of digital resources will be easy for researchers in the humanities to access and analyze.

This project will begin in October 2010 and run through March 2012.

The partner institutions in this project are Australian National University; Indiana University; Northwestern University; Tufts University; University of Chicago; University of Illinois, Urbana-Champaign; University of Maryland; University of Oxford; and University of Wisconsin, Madison.
1. Introduction and Background

The fundamental questions that launch inquiry in the humanities focus upon issues of interpretation, understanding, meaning, and values. Many of these questions have been with us for centuries—even millennia—yet all of them are posed in places, times, and circumstances that inflect what it is we are asking. The ways we can answer our questions also change over time. Trade routes, printing presses, the rise and fall of nations: all of these have made a difference to inquiry in the humanities.

During the past two decades, the development of technology has created a wealth of new possibilities for research, scholarly communications, and teaching within the broad academic realm that includes the humanities, arts, and humanistic social sciences. Most faculty now routinely stay in touch with colleagues in their fields by e-mail, visit the websites of professional societies, and search library holdings online. Many use digitized archives that they can search electronically or use digital tools to streamline basic tasks like generating bibliographies and footnotes.

As technology continues to evolve, scholars are realizing that it can advance their research in ways that far outstrip this kind of basic repertoire. For example, the realm of digitized books and images is expanding exponentially; faculty are eager to take advantage of this remarkable new resource with technology that would assist in such basic scholarly activities as annotating digital materials or carrying out a search across many digital archives. The limitations of email become readily apparent when two dozen scholars wish to comment on one another’s drafts of essays for an edited volume. A variety of research activities would be enhanced by even more specialized technology; for example, scholars are now able to reconstruct ancient social networks using probability functions on information from digitized corpora of artifacts.

In trying to create technology to meet these research needs, scholars have often invested time and money in creating software that will work well for their project but may be difficult for anyone else to use. These one-off projects yield limited returns on the time and funding invested in them. Although librarians and technology specialists are capable of helping, they are often uncertain where their investment of time and resources would do the most to help humanities scholars, and so the scholars often spend more time on solving technology problems than on addressing research questions. Meanwhile, these one-off applications are difficult to sustain and often prove to have short lives.

This is an ideal time at which to confront these challenges because the past decade has seen technological innovation that can now be harnessed to meet them. For example, cutting-edge work in the information sciences allows software developers to create reusable services that can support individual humanities projects across
many institutions and link hundreds of scattered digital collections. And technologists are finding new ways to operate these services in a web cloud, allowing a small consortium of institutions to meet the needs of many thousands of users around the world. The time is right to harness these new technologies to meet the specific needs of the humanities.

Recognizing that no single institution can carry out this mission by itself, the Andrew W. Mellon Foundation provided a major planning grant to fund Project Bamboo, a multi-institutional collaborative initiative to guide the creation of shared, interoperable tools, services, and content to meet the real needs of humanities researchers. The present proposal emerges from an 18-month planning process that drew together over 600 humanists, technologists, and librarians from 115 institutions to address the question, “How can we enhance research in the humanities through the development of shared technology services?”

From a series of eight focused workshops, a consortium of partner institutions has emerged. These institutions are ready to invest in technology to advance inquiry in the humanities. The Bamboo consortium is committed to several basic principles:

- Bamboo will address needs of humanities scholars.
- It will help to create shareable, reusable, and sustainable tools and services.
- It will foster the interoperability of technology and content.
- It will capitalize on existing technology and content, introducing new ways to connect, mix, and share.
- Its goals will be met collaboratively by a range of universities, colleges, and other organizations.
- It will create and sustain an upward spiral of communication among humanities scholars, technologists, and librarians.

In what follows, we will briefly sketch the context of our proposal and describe the project and its rationale. We then provide more detail about our technical approach, our project work plan, the roles for partner institution, a summary of our longer-term efforts to ensure a sustainable project, and our budget.

All of us involved in the Bamboo project have been fascinated to make a “meta” discovery: we cannot take it for granted that we have a shared language with which to articulate our plans. We are mindful that in places we are introducing terms that are unfamiliar to at least some of our constituencies, whether humanists, technologists, or librarians. Perhaps this is inevitable when a project of this character and scale is being launched; in any case, throughout the proposal, we have aimed to provide both general context and more detailed explanation in order to communicate our plans as clearly as possible to all of our participants.
2. Consortium and Partner Institutions

Before describing the project, we would like to say a few words about the consortial model to clarify the place of the project within a broader endeavor.

The University of California, Berkeley, will serve as managing partner for the project, and it will work to coordinate the contributions of the following partner institutions:

- Australian National University
- Indiana University
- Northwestern University
- Tufts University
- University of Chicago
- University of Illinois, Urbana-Champaign
- University of Maryland
- University of Oxford
- University of Wisconsin, Madison

The consortium of partners includes public and private universities, small and large universities, and universities in several countries. Each institution offers excellence in its humanities faculty, expertise in its library and technology organizations, and a readiness to make significant contributions of both cash and work. Information about the specific contributions of each institution to the work plan and budget may be found below. In addition to the partner institutions, approximately 15 institutions will participate in Bamboo as members. The member institutions will be especially helpful as we test the technology we develop, and we will rely upon their help as we support wider adoption of the technology across the realm of the humanities.

The Bamboo partners and members will participate in a range of activities that extends beyond the technology project proposed here. We describe the broader Bamboo Program in section 10 below.

3. Rationale and Benefits of the Project

The present proposal is for an 18-month project, but we are presenting it within the context of an integrated 36-month plan. We will distinguish clearly between the first 18-month phase and the second, but we believe that the rationale for each phase gains strength from its relationship to the other. We assume that success in the first phase of work would be a necessary condition for securing resources for the second phase of work, but even if the first phase should succeed, we do not assume that either the Mellon Foundation or the current partner institutions would agree to fund the second phase.
Over a three-year period, we aim to create two sets of applications that will directly support the research and teaching of large number of scholars in the humanities. “Bamboo Work Spaces” and “Bamboo Corpora Space” will meet needs that were identified during the planning process as core needs for scholarship in the humanities.

The planning process began with a set of workshops focused on scholarly practices and future needs repeated in four locations: Berkeley, Chicago, Princeton, and Paris. These workshops attracted more than three hundred participants from ninety institutions, and we recorded the strands of discussion that emerged and then grouped them into themes of scholarly practice.¹ We found that participants repeatedly expressed their desire for technology that would allow them to annotate, to collaborate, to gather materials, to organize information, to share materials, to store and preserve materials, and to use social media. Part of the analytical work we aimed to accomplish was to define each of these themes more precisely. For example, with the participants’ help, we described “gathering materials” as "the process of assembling a (yet-to-be-organized) corpus of research materials for subsequent analysis and manipulation, ... creating a collection or an archive or a rendering of something which one has discovered, for subsequent scholarly use, either by the creator or by others who might benefit from such a resource."

Building upon these discussions, we then identified a more general set of categories to describe central scholarly practices in the humanities that could benefit from technology. Prominent among them were the practices of aggregation, annotation, consideration, engagement, and interaction.² Again, one challenge for us was to define these practices as carefully as possible. For example, many participants focused on aggregation, and we summarized this as encompassing "the set of activities scholars perform to acquire and organize artifacts that an inquiry (study, exploration) will address. 'Aggregate' has a more temporary or provisional character than 'Preserve,' and perhaps a more local or individual character as well; but many of the same automation services are likely to apply."

The analyses of these discussions were refined in subsequent Bamboo planning workshops, and those later workshops in turn have led us to propose building two sets of applications. Work Spaces will provide basic and easy-to-use tools for scholarly collaboration and for managing and analyzing content, serving a large variety of individual researchers and groups of scholars. Corpora Space will allow individuals or groups to work on dispersed digital corpora using a range of sophisticated and specialized curatorial, semantic, and visualization tools and services. While Bamboo Work Spaces may evolve to include Corpora Space functions, we also envision other important Corpora-centered applications that would be realized outside of Work Spaces. In sections 4.1, 4.2, and 5.1 below we

¹ Themes of Scholarly Practice data can be found at https://wiki.projectbamboo.org/x/9oAZ
² Theme Groups can be found at https://wiki.projectbamboo.org/x/wYQ8
provide more detailed explanation about these applications and their relationship to Bamboo’s shared infrastructure and to other technologies in the humanities.

All of these applications will be supported by two shared infrastructure projects that will help technologists who work with researchers in the humanities. The first infrastructure project will be the development of an evolving set of scholarly web services on a services platform that will help technologists create and support a wide range of tools. The second will be the adoption and dissemination of standards and services for enabling collections to operate smoothly with one another. Together, these infrastructure projects will enable IT organizations, libraries, and digital-humanities projects to leverage one another’s accomplishments while ensuring that the world’s ever-increasing store of digital resources will be easy for researchers in the humanities to access and analyze.

In proposing these projects, we aim to enhance scholarship in the humanities in several important ways. First, we aim to make research far more efficient than it is now, helping scholars to avoid unnecessary expenditure of time on tasks that technology can make quick and easy. This will allow scholars to dedicate more of their time and energy to meeting the intellectual challenges that their inquiries generate.

Second, we aim to lower the logistical barriers that prevent or discourage scholars from taking full advantage of the varied digital resources that are becoming available in increasing quantity. Without technology tailored to scholars’ needs, this vast resource is for the humanities just so many zeroes and ones. The primary focus of the first phase of the project is on textual materials for two reasons: textual materials are what the majority of humanities scholars work on, and a focus on textual materials is technologically a good way to launch an initial project. Still, we wish to stress that these applications will allow scholars to explore and analyze many visual and other materials in greater depth as the applications evolve. Above all, we aim to give a large number of scholars ready access to tools that are both useful and easy to use.

Third, we aim to provide a services platform that will make many new techniques of analysis and investigation more readily available to scholars—techniques that are impossible to deploy, or nearly so, without the help of technology. These include, for example, computational methods of identifying patterns in very large corpora. More broadly, by providing a shared, stable platform, we aim to encourage software developers at many institutions to collaborate in developing the next generation of applications for research in the humanities.

Finally, we aim more broadly to do everything we can to help inaugurate a new era of technology support for the humanities, one in which services, applications, and tools are shareable; digital archives can all talk to one another and to the contents of each scholar’s technology tool-kit; and investment in a shared infrastructure can lift humanities scholarship at institutions around the world.
4. Project Overview

In service of these goals, our first 18-month phase of work will focus upon creating Bamboo Work Spaces and building the services platform for future humanities applications. During the second 18-month phase of work, we will enhance Work Spaces, add more scholarly and interoperability services, and deploy the services platform to a larger set of institutions.

In phase 2, we will also develop applications that focus on analyzing corpora. In phase 1, we will carry out a design project to determine the precise Corpora Space applications to be developed in phase 2. All Corpora Space applications will use Bamboo web services to support scholarship engaged with corpora of one kind or another, but these applications may be built in a number of different ways. We expect that we will be able to house some of these applications within Work Spaces; others could be situated directly on the Bamboo Services Platform; and others could be built using non-Bamboo technologies as their foundation (see section 5.1 for a discussion of Bamboo’s relationship to the larger ecosystem of humanities technologies). We are using the term “Corpora Space” to cover corpora-centered applications that use Bamboo web services situated in all of these ways. The table below summarizes the time-line for building the applications and technology infrastructure. We expect that by months 12-15 of phase 1, we will have made decisions about which Corpora Space applications to build in phase 2. (The percentages are estimates of the total project resources that would be devoted to each project in each of the two phases of work. Please note that the phase 2 percentages are rough estimates and that we have included a range of efforts for Work Spaces and Corpora Space.)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Applications for Humanities Scholars</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1. Work Spaces</strong></td>
<td>Tailor existing collaborative environments for the humanities; connect WorkSpaces and services platform (30%)</td>
<td>Add further functions of value to humanities and greater interoperation (20-40%)</td>
</tr>
<tr>
<td><strong>2. Corpora Space</strong></td>
<td>Design Corpora Space for phase 2 development (10%)</td>
<td>Build Corpora Space on top of Bamboo services platform (20-40%)</td>
</tr>
<tr>
<td><strong>Shared Infrastructure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3. Scholarly web services on Bamboo services platform</strong></td>
<td>Build services platform, deploy first set of scholarly web services (analytic) on platform, disseminate developer toolkit (45%)</td>
<td>Add further scholarly web services for curatorial, semantic, and visualization functions. Refine platform deployment options for campuses. (30%)</td>
</tr>
<tr>
<td><strong>4. Collections interoperability</strong></td>
<td>Adopt standards and build interoperability services into platform. (15%)</td>
<td>Work with core collections technologies to enhance interoperability with scholarly web services. (10%)</td>
</tr>
</tbody>
</table>

In what follows, we will describe each element of these projects and their work-phases in more detail.

### 4.1 Bamboo Work Spaces

We will provide Work Spaces for a broad range of humanities scholars pursuing research projects individually or in groups. These Work Spaces will include basic capabilities for collaboration, content, and scholarly analysis tailored for the humanities. We will create Bamboo Work Spaces by enhancing several existing collaborative and content-management environments.

The Work Spaces will require only simple assembly by scholars at their home institutions. In their early form, the Work Spaces will make it easy for individuals or self-defined groups of scholars to share and comment on digitized materials; to deploy useful gadgets such as calendars, maps, and RSS feeds; and to carry out various forms of mark-up and discussion related to different documents. The Work Spaces will also enable scholars to share and acquire information about tools and services that colleagues at other campuses are using and evaluating.

At colleges and universities, the Work Spaces will probably be run by an institution’s library or IT organization, because the application will require technical expertise to operate at scale. At many institutions, the campus humanities center will partner
with the technology organization to ensure close engagement with faculty and students in the humanities.

We will develop Bamboo Work Spaces by using and modifying several open source enterprise platforms and related environments. In this way, we can leverage work that has already been done on HubZero, Alfresco, Sakai, and JISC’s Virtual Research Environments. In each case we will be adding user interfaces and greater functionality to produce Work Spaces tailored to the needs of the humanities. In addition, we will add back-end integration technologies\(^3\) so that web services can be accessed between Work Spaces and the Bamboo services platform and so that content can be more easily shared between Work Spaces.

By taking this approach, we hope to make core research activities simpler for a broad range of scholars, achieving this goal through as much reuse and loosely coupled integration as possible. The Work Spaces project is intended to provide large numbers of humanities scholars with a great variety of research interests with at least a basic digital work environment, and to find ways to connect different Work Space at different campuses to each other. It will do this by lightly tailoring existing technologies and stitching them together to make a good fit with the humanities.

In phase 1, we will carry out two Work Space projects using two platforms: one will extend the HubZero platform to the humanities, and one will develop work spaces using a number of front-end technologies (including Sakai 3 compatible tools) supported by an Enterprise Content Management platform. We provide technical and functional detail about these projects in section 5.3 and a description of their project teams in section 6.2 below.

### 4.2 Bamboo Corpora Space

By the end of the first 18 months, we will have prepared the design groundwork to launch the second project: applications that will allow scholars to work on dispersed digital corpora using a broad range of powerful research tools and services. We use the term “Bamboo Corpora Space” here to refer to a set of applications that have in common two fundamental characteristics. First, each application will provide powerful services designed to help scholars curate, interpret, depict, and discuss corpora. This is a key category across many disciplines in the humanities because corpora (textual and non-textual) are a central object of study. Second, each application will use an array of scholarly and other web services provided by the Bamboo services platform. This is a key step on the technology side, because it will permit the development of a generation of applications that re-use tools and software components from a common source.

Corpora Space applications could be realized in many ways. A Corpora Space application might be situated within Bamboo Work Spaces; specialized corpora

\(^3\) Please see section 5.3 for more detail on how Work Spaces will call web services.
management services can be embedded in a general collaborative and content management environment designed for the humanities. But that is not the only way in which Corpora Space applications could be realized. For example, technologists might draw upon the Bamboo services platform in order to build a Corpora Space application that would work as a “Pipes”-based environment in which scholars can create their own mini-applications by mashing-up data and tool services. A Corpora Space application might be realized as a lightweight IPad application, or it might take the form of refactoring part of an existing corpora management tool using Bamboo services. Finally, we might envision the development of a dedicated Corpora Space platform, which, like the CollectionSpace\(^5\) platform for the museum community, could be readily customized and extended to build and host a wide range of complex corpora applications. (We provide more information about these possible Corpora Space application options in section 5.1).

We are not yet certain which of these options it would be best to take up in phase 2. Thus in phase 1, our goal for Corpora Space is to carry out a structured design process with the participation of the project partners, as well as other institutions, to determine which application or applications to develop in phase 2. In making this determination, this broad range of institutions will need to consider both what is technologically feasible and what is of greatest importance for humanities scholars. The key will be to enable scholars to carry out complex research tasks using a common set of tools and services on dispersed corpora. We expect to learn a great deal from the work carried out by a number of more focused corpora environments in the humanities, for example, in French and English literature, classical scholarship, and linguistics. Our design process, which is described more fully in section 5.4, will include the solicitation, discussion, and review of a series of Corpora Space white papers written by scholars and technologists.

### 4.3 Scholarly Web Services on a Bamboo Services Platform

To meet our goals, we must be able to provide shared infrastructure for the arts and humanities across many institutions. Our approach is thus to provide an evolving set of web services on a Bamboo services platform.

This major infrastructure project is made up of three component projects, and we will have occasion in what follows to identify them separately. (1) We will develop and host an initial set of web services that support distinctive needs of scholars in the humanities. (2) We will make these scholarly web services available on a stable and secure Bamboo services platform so that many applications and repositories can use them well into the future. (3) We will create and disseminate a software tool kit and shared services lifecycle process so that new and existing web services can be readily deployed to meet the research needs of scholars in the humanities.

---


The Bamboo planning project has revealed that there are important scholarly methods and functions that are currently locked away in existing digital humanities applications. We believe these scholarly services fall into such key categories as curatorial, analytic, semantic, and visualization services. We propose to model and define these methods and functions as modular and reusable web services on a Bamboo services platform. In that way, they will be able to contribute to many other applications and tools, benefiting many other scholars across disciplines, now and well into the future.

In our first 18-month phase of work, we will derive scholarly web services from six to eight existing applications. These services will include those that provide analytic functionality related to texts and other corpora. In our second phase of work, we will add additional services. (Please see section 5.5.1 for more detailed description of these services and the applications from which we will derive them.)

We aim to make scholarly web services available to as many users as possible in as secure and stable a means as possible. For that reason, we will build a services platform for research in the humanities. The most basic level of the platform will be assembled mainly from established and robust existing technology. We will build on widely adopted open-source infrastructure, harnessing the products of industry-wide software development efforts without binding ourselves to commercial vendors. This will greatly enhance the adoption, sustainability, and reusability of the platform and of the tools and services that will use the platform resources. (Please see sections 5.5.2 and 5.7 for more detailed description of the Bamboo Services Platform.)

In building the platform, we will not be reinventing the wheel. Several of the partner institutions will bring highly relevant experience to the platform work-area, and we will also be leveraging CollectionSpace and related initiatives. (See section 6.7 for more details.)

To deploy the platform, a small number of institutions will host the platform for use by many other universities and colleges. In phase 1, we will test this deployment strategy. In phase 2, we will then craft long-term operational agreements for support of the platform. (Please see section 5.8 for more information on the deployment strategy for the Platform.)

So that many institutions can use and contribute to the platform, we will create a shared-services process and toolkit for use by software developers and campus library and IT organizations that support humanities scholars. This will be valuable in itself, and it will set the stage for future Bamboo development efforts. (Please see section 5.5.3 for an explanation of the Shared Services Lifecycle and Software Developer Kit.)

---

6 We define these categories in section 5.5.1
Finally, we note that in phase 1, we propose to carry out a small number of simple adoption projects in which partners will use and test scholarly services on the Bamboo platform with existing applications or to build simple prototype applications. We envision these as early pilot projects (not production software), which will both inform platform development and Corpora Space design. We have identified two initial adoption projects and describe these in section 6.4.

4.4 Collections Interoperability

To support a broad range of scholarship in the humanities, scholars desperately need for distinct digital collections of research materials to become interoperable. Interoperability must extend beyond support simply for resource discovery; scholars must be able to deploy tools and services across widely distributed collections without needing to be expert in every digital format used and every brand and version of repository software extent in academia today. Digital object descriptions must be rich enough and precise enough to support scholarly reference and allow the implementation of transformation and remediation tools and services that can facilitate digital information resource reuse and recombination, while simultaneously maintaining resource provenance adequate for scholarship.

Despite a variety of technical and policy challenges, we believe that Bamboo can make a significant contribution in this area, by defining standard methods for making digital content available to web services. Where existing protocols, practices or ontologies can be leveraged, we will do so, extending and profiling current community standards as required to meet the rigorous requirements of scholars. Simultaneously we will identify gaps in existing standards, and define new technical approaches, application profiles, and best practices as necessary. We will thus develop, adopt, and publish a set of guidelines, protocols, and specifications that will help content providers enhance interoperability by taking advantage of the Bamboo platform. We will also develop services that will gather usage data from collections. In this way, we can use the platform to track scholarly activity and the ways in which scholars use content collections. This in turn will allow us to understand where the efforts of Bamboo should be focused after the initial three-year building projects.

---

7 In workshop 1 of the Bamboo Planning Project, at which participants were asked about current and future needs in the digital humanities, a significant number of participants raised the issue of collections interoperability and the related theme of content and tool interoperability. Because of the importance of this issue, the Bamboo Planning Project team created a strategic focus area on “Content Interoperability Partnerships” as one of the 11 major elements of Bamboo’s 7-10 year program document, which was presented at workshop 4 of the Bamboo Planning Project. See: https://wiki.projectbamboo.org/x/jQGK. At workshop 4, approximately 30 institutional teams formally voted on which elements of this Program document Bamboo should focus on in the short term and which elements institutions were will to lead. In both rounds of voting “Content Interoperability Partnerships” was ranked in the top 2-3 categories. See https://wiki.projectbamboo.org/x/xYCR.
We will work with libraries, museums, and special collections, and we will work closely with such initiatives as Fedora / DuraSpace, campus enterprise content management activities, Hathi Trust, and CollectionSpace. Many of the Bamboo partners have extensive experience in meeting challenges in this arena, and we believe they will be able to integrate content collections with various content-management technologies in new and valuable ways. Further details about the ways in which Bamboo will fit into a larger ecosystem of humanities technology may be found in section 5.1 below.

4.5 Summary of Project Deliverables

We end this section with a summary of the phase 1 project deliverables. We then explain more of the technical details of each of these areas in the next section, Section 5, Technical Approach.

<table>
<thead>
<tr>
<th>Area of Work</th>
<th>Deliverables and Outcomes</th>
</tr>
</thead>
</table>
| Work Spaces          | 1. Two Bamboo Work Spaces and documentation for the operation of these work spaces by other institutions.  
                       |   a. HUBzero Work Space                                                                   |
|                      |   b. ECM Platform Work Space                                                             |
|                      | 2. Tools and Services Registry, including integration with each of the above Work Spaces and with the Bamboo Services Platform. APIs and documentation so that other environments and applications can share data with the registry. |
|                      | 3. Work Spaces Phase 1 Assessment and Roadmap for Phase 2 (this will address the relationship of Work Spaces and Corpora Space) |
| Corpora Space Design | 1. Three to six white papers for Corpora Space Application Design                        |
|                      | 2. One to two Corpora Space Applications proposals selected by the Bamboo Partners, including detailed design specifications |
| Scholarly Services   | 1. Development of Scholarly Services deployed on the Services Platform                   |
|                      |   a. PhiloLogic Services                                                                 |
|                      |   b. Ancient & Classical Studies / Perseus Services                                      |
|                      |   c. Proxied SEASR Analytics Services                                                    |
|                      |   d. Prosopography Services                                                              |
|                      | 2. Documentation of service descriptions, APIs and service contracts of listed scholarly services |
|                      | 3. Carry out several adoption pilot projects in which existing applications use scholarly services and / or new applications are built with these services. |
|                      | 4. Assessment of Phase 1 Scholarly Service development and plan for Phase 2 Scholarly Services development |
In each of the four areas of work we will carry out assessment activities as part of project evaluation. We summarize these evaluation activities in section 11, "Reporting and Evaluation."

5. Technical Approach

In this section we provide additional technical detail about the four areas of work Bamboo will carry out in phase 1: Work Spaces, design of the Corpora Space applications, scholarly web services on a Bamboo services platform, and collections interoperability. We begin first by situating Bamboo’s work in the context of a larger ecosystem of technology in the humanities and in higher education. This will help to explain why these areas of work are significant and how they are related to one another.

5.1 Technology Ecosystem and Strategy

Project Bamboo will develop technology as part of a much larger ecosystem of existing and emerging digital tools, content sources, methods, and algorithms in the humanities and higher education as a whole. We aim for Project Bamboo to make its own distinctive contributions to the health of this ecosystem, by interconnecting
tools and content sources, opening up scholarly methods and algorithms that are locked away in specific applications, and fostering far greater re-use and efficiency across institutions in the development and operation of scholarly applications.

To make these contributions, we will be working within several key areas at the same time. Work Spaces will provide virtual work environments suitable for large numbers of scholars, and we will build these by leveraging substantial existing technologies. The Bamboo Services Platform will create a foundation that many technologists and scholars can leverage as they use its scholarly and other web services in creating new applications. By working on collections interoperability, we will be reaching out to the digital library community to help scholars work smoothly and productively across many distributed collections that currently do not “play together” well. In the Corpora Space design process, we will draw upon the expertise of a broad community of technologists and scholars to determine how best to leverage our accomplishments going forward and build upon Work Spaces, the services platform, and our interoperability contributions to create sophisticated applications of high value. In each of our areas of work, we will borrow from and re-use as much existing technology as we can, and we will foster further integration and interoperation throughout the technology ecosystem.

The simplified ecosystem diagram below shows three major areas of technology that Bamboo will build, adapt, or interoperate with.

*Corpora Space applications could be built in each of the asterisked areas*
In this diagram, humanities scholars are at the top because it is their work that the technology infrastructure is intended to support. They access tools through the application user interface layer, represented in blue. These interfaces are supported by various underlying technologies that carry out workflow, store data, and interoperate with other applications and content sources. Please note that in this simplified representation some applications access services from the Bamboo Platform but are principally supported by, and part of, a separate technology stack (the two outside columns), while other possible applications are directly supported by the Bamboo Services Platform (the middle column).

**The column labeled** “Other Humanities Applications” represents existing and future tools, content sources, or platforms that can function better by making use of scholarly services deployed on the Bamboo Services Platform. These technologies exist independently of Bamboo, implemented atop technologies chosen by their creators. But Bamboo makes it simple for these applications to interoperate with Bamboo services: they need only provide interfaces for the web services exposed by Bamboo. By piggy-backing on Bamboo services in this way, the developers of these applications can focus their energies on creating new tools and services to support particular areas of scholarship.

**The column labeled** “Shared Bamboo Infrastructure and Scholarly Applications” shows the Bamboo services platform along with the Bamboo services and applications that are built on it: scholarly services and applications, along with services that enhance interoperability among different collections. In phase 1, Bamboo will develop the services platform, selected scholarly services, and collections-interoperability services (that is, we will build up to the blue user application layer in phase 1). Although phase 1 would not see the development of major new scholarly applications directly supported by the Bamboo platform, it is vital to the entire project that the infrastructure for applications be put into place early. That way, both Bamboo partners and other technology developers will be able to develop applications with more efficiency, drawing upon the essential scholarly, interoperability, core, and utility services hosted at lower layers of the service-platform architecture. (More information is provided in sections 5.5 and 5.7 below.)

**The third column represents the** Bamboo Work Spaces, which are the environments in which scholars will be able to collect, annotate, and share content. We will build the Bamboo Work Spaces by using existing technology, including open-source Enterprise Content Management (ECM) technologies (e.g., Alfresco) and virtual research environments (e.g., HUBzero).

With the help of this ecosystem diagram, we can more fully describe the possibilities for phase 2 Corpora Space applications and their relationship to Work Spaces and the Bamboo shared infrastructure projects.
Again, by "Corpora Space" we mean a set of corpora-centered applications that use the services provided by the Bamboo platform to provide powerful functions to scholars as they curate, interpret, depict, and discuss dispersed corpora. Here are some examples of possible Corpora Space applications:

1. **Corpora services within Bamboo Work Spaces.** Work Spaces would be developed so that they offer increasingly powerful and specialized corpora management functions. These functions could be provided via the services on the Bamboo platform and built directly into the Work Spaces environments. The corpora services would be exposed via APIs to the Work Spaces.

2. **“Pipes” model.** Technologists could develop Pipes-based corpora applications on top of the Bamboo Services platform. These applications would be among the “scholarly applications” in the middle column of the diagram above.

3. **Refactoring of an existing humanities corpora application.** The refactored application could be built on top of the Bamboo platform or as a stand-alone application. In the former case, the application would be among the “scholarly applications” in the middle column above; in the latter, it would be in the “other humanities applications” column.

4. **iPad application.** With the help of the Bamboo platform’s web services, technologists could build an iPad application as a stand-alone application. This would be among the “other humanities applications.”

5. **Corpora Space equivalent of the CollectionSpace Platform.** It would be possible for technologists to build such an application platform on top of Bamboo Services platform. In this case, the application would be among the “scholarly applications” in the middle column above.

We wish to emphasize two points here. One is that there is a range of technology options open to us for development of scholarly applications after phase 1, and that these applications can be developed both by the Bamboo partners and, in a number of cases, by others institutions as well; the other point is that the services provided by the Bamboo services platform, along with the interoperability projects, are essential to each option, and will need to be enhanced and sustained to support future applications.

### 5.2 Definitions and Scope of “Content” and “Corpora”

For the purposes of this project we mean the following when we speak of “content” and “corpora.” By “content” we mean text, images, video, audio, and associated metadata. By “corpora” we mean structured sets of these materials. Structured texts may themselves include associated images, video, and/or audio. The corpora may be stored in one location or made up of the aggregation of distributed materials.
across digital collections held in libraries, research centers, museums, and/or archives within and outside of universities.

In phase 1, Work Spaces will be able to ingest and store all content types as digital binaries. This capability – equivalent to storing a file on a hard disk drive, without regard to whether or how that file can be manipulated, analyzed, or transformed – is an essential precursor to extended functionality that is useful to a scholar. Examples of extended capabilities include the ability to transform stored content from one format to another (e.g., Word documents to PDF, TIFF images to JPEG, or unstructured text to an indexed object-relational structure à la PhiloLogic); to generate concordances of textual materials; to generate histograms that represent tonal distribution in a digital image; or to collate multiple drafts or editions of a digitized text. In phase 1, Work Spaces will enable annotation, transformation, discussion, and sharing of documents whose principal content is text. Please see section 5.3 immediately below for a more detailed list of these functions. Bamboo will not develop extended functionality for all content types in phase 1.

In phase 1 Bamboo will define the functions for Corpora Space applications to be implemented in phase 2. We expect to focus initially on the needs of scholars who work with text-centered corpora. It is possible — and may be highly desirable — to address some critical and common needs for the curation, analysis, visualization, and presentation of other media types, such as audio and video. We note, though, that handling of content as varied as music recordings and video clips presents complex and challenging issues. Thus we will not make strong claims about functions that can be applied to an extended range of content types prior to performing the phase 1 review and design work for Corpora Space.

The scholarly services to be developed in phase 1 will operate on digitized texts. As we evolve our work in phase 1, we intend to consider services that are of value to the analysis of other digital content types. Certainly a number of our partners bring both expertise and interest in focusing on services that operate on audio and video media. The Corpora Space design process and the capacities and interests of the partners in this project will help to define which additional scholarly web services we explore in phase 1.

### 5.3 Work Spaces

Project Bamboo Work Spaces will be implemented atop each of two open-source technology platforms, HUBzero and Alfresco Enterprise Content Management System. In the current phase of work, strong native capabilities of these platforms will be augmented to deliver functionality useful to broad groups of humanist scholars in three aspects: as a collaboration tool; as a means of accessing digital objects residing in collections repositories; and as a means of invoking analytical and other tools to operate on these objects. A core set of capabilities, listed below, will be available in all Work Spaces, independent of underlying technology platform.
Work Spaces will also deliver additional capabilities native to the underlying technology, HUBzero or Alfresco.

5.3.1 Core Work Spaces Capabilities

The following capabilities will be available in all Bamboo Work Spaces:

- store and organize sets of any digital content type, including text, images, video, and audio;
- record core metadata about the digitally represented object (e.g., author, title, pertinent copyright information, provenance);
- record technical metadata pertaining to the formats in which objects may be stored (e.g., standard image, video, and audio metadata);
- define custom metadata schemas appropriate to the collected materials and research methods, and record information structured according to schema definitions;
- annotate stored content (e.g., tags, unstructured notes);
- create and use elements that facilitate collaboration (e.g., calendars, RSS feeds, discussion threads, surveys);
- create and use elements that facilitate analysis and consideration (e.g., concordances, tag clouds, maps generated from geolocation metadata, ratings);
- call on remotely deployed web-services to operate over an object or objects (e.g., scholarly services deployed on the Bamboo Services Platform; in phase 1, Bamboo-hosted scholarly services will be applicable only to textual content);
- record information on methods, procedures, and tools used in scholarly research, and report to one or more central aggregation stores. (For more information, see Section 5.3.5, Tools and Services Registry.)

The penultimate bullet, above, will benefit from an example. Two of the Scholarly Services listed in Section 5.5.1 are Document Mapping and Concordance. A scholar managing a collection of texts in a Work Space could select a text from her collection, and call the Document Mapping service on the Bamboo Services Platform to prepare it for further analysis; then call the Concordance service to operate on the mapped (indexed) text to generate a concordance built using PhiloLogic\(^8\) algorithms. Such a concordance could itself be saved in the Work Space as an outcome of scholarly inquiry, and considered, analyzed further, shared, or published.

5.3.2 Work Spaces Technology Selection Priorities

We have selected two platform technologies appropriate for enterprise deployment and hosting (e.g., in an institutional data center or in a vendor cloud) as bases from

\(^8\) PhiloLogic: http://philologic.uchicago.edu/
which Work Spaces functionality will be adopted, adapted, or extended. Selection was guided by these priorities:

- significant overlap between native functionality (not adapted or extended by Bamboo) and Bamboo Work Spaces functional and technical requirements;
- exposure of an interface to a platform technology's repository that conforms to the emerging OASIS Content Management Interoperability Services (CMIS)\(^9\), to better insulate scholars from changes in back-end technology;
- flexibility to revise user interface substantively to conform to scholar-specific workflows;
- flexibility to brand user interface with the identity of an institution, humanities center, discipline, or project that is hosting and/or utilizing Work Spaces.

### 5.3.3 HUBzero Work Spaces

Indiana University (IU) and the University of Wisconsin-Madison will jointly implement and configure the scientific collaboration platform HUBzero\(^10\) for use by humanists involved in collaborative research projects.

The many instances of HUBzero in current operation demonstrate the utility of the platform’s base functionality. The HUBzero platform is an evolution of the nanoHUB.org web site, which was developed by the Network for Computational Nanotechnology, supported by the NSF as an infrastructure and research network since September 2002. This environment has been adopted and enhanced at 25 operating research HUBs, with more in the planning stages. Over 360,000 scientists currently use these HUBs. HUBzero includes capability to utilize national computational grids such as the TeraGrid and the Open Science Grid. Its job submission and monitoring system has been used effectively to submit tasks to local as well as national computing and storage systems.

Project Bamboo chose HUBzero as a Work Spaces platform in significant part because its collaboration component is built on the open source content management system Joomla!.\(^11\) HUBzero's base configuration of Joomla! allows researchers to form groups easily, to create and share resources, to collaborate on ideas through wiki and forum tools, and to comment on the work of colleagues. The platform also enables collaborators to notify each other as pieces of work are completed, to make collective decisions, and to share their work with larger communities.

Groups of scholars will be able to customize their local Bamboo/HUBzero Work Spaces to suit particular needs by adding extensions available from the Joomla!

---

\(^9\) CMIS: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cmis

\(^10\) HUBZero: http://hubzero.org

\(^11\) Joomla!: http://www.joomla.org/
community. It is not possible to fully capture the collaboration requirements of digital humanists, or to predict future possibilities. With Joomla!, scholars will be able to add to the Bamboo/HUBzero tools adopted or built during the grant project by tapping into the creative energy and the development talent of a large, diverse, and active open source community formed around Joomla!. This community has implemented more than 5,000 extensions\(^2\) to the base Joomla! product. If researchers in the Bamboo community find that they need a new collaboration widget, it is likely that the widget will have already been built as an extension by someone else in the Joomla! community.

Several other open source projects will be tapped to provide additional functionality to scholars. Because a number of Bamboo partners use Fedora\(^3\) as the backbone of their institutional repositories, we plan to use Fedora as a local object repository for Bamboo’s HUBzero Work Spaces. Staff at Northwestern University are currently developing a JCR\(^4\) connector for Fedora that we will leverage to implement a Collections Interoperability API layer between HUBzero and Fedora. The JCR connector will enable local collections to behave like any other collection, allowing exposure of local collection, via the API, to other Bamboo tools and workspace instances.

We will look to the Open Annotation Collaboration\(^5\) for guidance on developing an interoperable means of annotating objects and activities. Shibboleth/InCommon\(^6\) will support cross institutional authorization and authentication, a crucial component of brokering access to objects under copyright. HUBzero provides for remote web service calls that operate on Work Space-managed content (e.g., to services deployed on the Bamboo Services Platform) to be presented to the user as ‘native’ HUBzero tools. From a Work Space user perspective, these tools operate as if they were deployed locally; from a programmer’s perspective, these web-service-calling tools may be implemented in any of several programming languages (Java, Python, Perl). Web service calls that present information to users (e.g., information gleaned from the Tools and Services registry, described below) will be implemented in HUBzero’s web layer as Joomla! extensions.

The HUBzero implementation will seek to leverage the work done in related projects at our institutions whenever practical, allowing us to extend the base functionality described in this proposal. For example, Indiana University is planning to integrate the VIVO\(^7\) semantic web engine within HUBzero under a

---

\(^{12}\) Joomla! extensions: http://extensions.joomla.org/extensions

\(^{13}\) Fedora Commons: http://www.fedora-commons.org/

\(^{14}\) Content Repository for Java technology API (JCR): http://jcp.org/en/jsr/detail?id=170

\(^{15}\) Open Annotation Collaboration: http://www.openannotation.org

\(^{16}\) InCommon: http://www.incommonfederation.org/

\(^{17}\) VIVO: VIVO Web describes itself as follows: "The National Network enables the discovery of researchers across institutions. Participants in the network include institutions with local installations of VIVO or those with research discovery and profiling applications that can provide semantic web-compliant data. The information accessible through VIVO’s search and browse capability will reside and
separate NSF grant. VIVO could allow us to provide a rich discovery tool for scholarly profiles maintained outside the HUBzero Work Spaces. Should the NSF decline to fund the separate grant, the capability and value of HUBzero described elsewhere in this proposal will not be affected.

The University of Wisconsin-Madison will host a production instance of the Bamboo/HUBzero workspace, and expect to provide this hosting beyond the grant period by securing funding and support as part Bamboo sustainability planning. While other institutions will be able to deploy HUBzero Work Spaces on their own campuses, UW-Madison’s long term hosting will also leverage HUBzero’s group functionality to allow distinct groups of scholars to share a HUBzero instance without interfering with each other’s work. In this way, HUBzero Work Spaces will support groups of scholars whose funding and access to technical support is modest, as is typical in the humanities.

Indiana University will host a development instance of the Bamboo/HUBzero Work Spaces through the duration of the grant, and intends to secure ongoing funding and support for this Work Spaces instance once Work Spaces are a proven, viable component of the Project Bamboo Program. IU’s instance will be available as a hosted development environment to all members of Project Bamboo for use in creating and testing modules and toolsets for the HUBzero Work Spaces. IU’s Work Spaces will also mirror the UW-Madison instance, to assure reliability of the Bamboo HUBzero Work Spaces.

During the initial phase of our work beginning in October 2010, we plan to implement two concrete and complimentary use cases.

The Digital Humanities Working Group, led by professors Robin Valenza and Michael Witmore in the University of Wisconsin-Madison English department and Peter Gorman in the university library, will begin using an instance of HUBzero as the working environment for their research group’s expanded analysis of English textual genres from 1500 through 1800. This project includes data sets from three different sources, each of which include thousands of book-length documents that need to be transformed so that they can be used in the same experiments. Their work uses a range of existing software for providing alternative markups and categorizations of documents as well as new tools being developed specifically for their work. The project has reached a level of complexity such that the use of these tools and the activities of the group’s members need to be tracked and coordinated.

\footnote{Barnett, W., Ding, Y., and McDonald, R.H. “SI2-SSI: An Identity Management Infrastructure for HUBzero.” National Science Foundation Proposal No. 1047899 under NSF OCI Program 10-551 Software Infrastructure for Sustained Innovation. This is currently under review by NSF. Note that Bamboo Work Spaces built atop HUBzero will leverage both the identity management and semantic web engine integration funded by this grant, should it be awarded. Project Bamboo has provided a letter of support submitted to NSF with this proposal.}

\textit{be controlled locally, within institutional VIVO's or other semantic web-compliant applications.}
in ways that reach beyond their current methods of collaboration. The research group includes faculty, staff, graduate students, and undergraduates in computer science, English, education, and library science, each of whom has different areas of expertise and contributes different work to the project. For example, Bill Blake in English has taken on the task of comparing the software’s analysis to those produced in more traditional ways by humanists. Michael Correll in Computer Science is designing tools for visualizing the data that results from the experiments (these tools that will initially be deployed in the work space but may ultimately become part of the Bamboo services platform). HUBzero will enable the Digital Humanities Working Group to coordinate their activities in more predictable and reproducible ways, the gold standard for academic research that depends on the large-scale analysis of data. The shared environment will also allow the group to work together on interpreting their results. Jointly-written papers for academic publication will be evidence of successful collaborations. (It is not the aim of the Bamboo/HUBzero workspace implementation to support collaborative authoring or editing tools.)

**Cross Institutional Resource Sharing within the CIC.** IU is a member of the 12 institution Committee on Institutional Cooperation (CIC). The CIC licenses scientific resources in large batches in order to gain leveraged pricing for institutional members. These resources can be used in new and different ways including data-mining across institutional resources. Several Bamboo Partners have a vested interest in connecting with the NCSA/UIUC supported Monk (text-mining) service and in providing a gateway to the Text Creation Partnership (EEBO and ECCO) and ProQuest (Chadwyck-Healey Nineteenth-Century Fiction) specialized full-text dataset that is supported within the current CIC affiliated authentication methodology. Brokering access to content of this kind will be a key component of Bamboo. The planned Work Space authorization tool will resolve many issues for Shibboleth authentication and authorization among Bamboo Partner sites. This type of community sharing platform has also been proposed as a starting point for the forthcoming HathiTrust Research Center (HTRC) that will be supported through Indiana University and the University of Illinois. Approval of HTRC proposal by the HathiTrust Executive Committee is still pending at this point. Also, the proposed HTRC would be a Google Book Research Center and consequently depends on the Google Books settlement. If all goes as we hope with the HTRC, then we will use the opportunity to integrate access to HTRC to further demonstrate the usefulness of the Bamboo/HUBzero Work Space in mediating access to texts via its authorization tools. However, even if the HTRC proposal fails or funding for that

---

19 The text creation partnership produces standardized, digitally-encoded editions of early print books. Primary sources of this partnership’s digital images are Proquest’s Early English Books Online (EEBO) and Gale’s Eighteenth Century Collections Online (ECCO). http://www.lib.umich.edu/tcp/
20 ProQuest: http://www.proquest.com
21 CIC affiliated authentication methodology is shown at: https://monk.library.illinois.edu/cic/public/
22 HathiTrust: http://www.hathitrust.org
effort is less than we hope, the Bamboo/HUBzero Work Space will have demonstrated its value as outlined above.

5.3.4 Enterprise Content Management (ECM / Alfresco) Work Spaces

The University of California at Berkeley, the University of Oxford, and Australia National University will build Work Spaces atop Alfresco, an open-source Enterprise Content Management (ECM) platform to support the work of individual and groups of humanist scholars. We will use a number of front-end user-facing technologies, including the OpenSocial framework (which will enable compatibility and connectivity with Sakai 3 and Google, among other platforms); Drupal content management system; and Alfresco’s Share application. The project team will build upon Berkeley’s ongoing work to develop collaborative and content management environments for academic departments, research centers and museums. The ECM Work Space will also draw upon Oxford’s detailed prototype and specifications for a virtual research environment (VRE) for the humanities. This project will also include the development of a Drupal module so that content management environments using Drupal can exchange information about tools and services of value for the humanities with other Bamboo Work Spaces.

Enterprise Content Management (ECM) platforms are robust, scalable and extensible standards-based systems for providing and developing new content management services. An ECM platform offers a powerful suite of tools to capture, manage, store, preserve and deliver digital content, including applications for document management, collaboration, web content management, records management, and workflow or business process management – all built on top of a robust data repository.

ECM Features include:

- Document Management (access/retrieval, document transformations, metadata management)
- Collaboration Tools (blogs, wikis, groups, project sites)
- Web Content Management (web publishing, content management)
- Workflow Management (custom rules, workflows)
- Scalability (designed to accommodate thousands of users, millions of assets)
- Security (for sensitive data)
- Integration with other content platforms (SharePoint, Drupal, Joomla, etc.)
- Configurability, Extensibility (designed for developing complex web applications)

---

24 Drupal: http://drupal.org/ is an open-source content management system, to which an active development community has contributed nearly 8,000 modules (extensions to core functionality) compatible with the current release (cf. http://drupal.org/node/206666#DRUPAL-6).
• Interoperability (open APIs, RESTful services\textsuperscript{25}, common data interoperability standards)

Alfresco is a leading open source enterprise content management platform. A Java-based platform, Alfresco supports open standards\textsuperscript{26} Alfresco provides a suite of powerful and easy-to-use web-based applications for web publishing, document management and collaboration (Alfresco Share); versioning and transformation of documents; ability to add workflows, rules and aspects to objects; and customizable content models (to support, for example, multiple metadata schemas). Alfresco is also one of the leaders in defining and delivering CMIS\textsuperscript{27} functionality, making it a powerful platform for application development. Alfresco Share is an open source collaboration and document management application that sits atop the Alfresco ECM repository. Share allows users to easily create collaborative workspaces that include a document library, wiki, blog, calendar, and discussion forum, among many other features.

This project will leverage Drupal, a popular open source content management system, via an existing CMIS module to connect collaborative Work Spaces and research collections with existing Drupal-based projects.

OpenSocial\textsuperscript{28} provides a common API for social applications across multiple websites, including popular commercial web services offered by MySpace, Yahoo, LinkedIn, and Google. An open source implementation of the OpenSocial APIs is available from the Apache Shindig\textsuperscript{29} project, and support is integrated into the next generation of the Sakai VLE software. The Oxford team will lead implementation of OpenSocial-compatible “gadgets,” which can be embedded by simple user configuration in many different customizable web environments. For example, a scholar with access to a tool, developed as a gadget and served from the University of Oxford, would be able to embed the gadget within the Oxford VRE, in their own institution’s VRE, in a future Virtual Learning Environment supporting OpenSocial (such as Sakai), or even in external services such as their iGoogle homepage. Oxford has already created web-based tools to assist in the study of ancient documents as part of the Virtual Research Environments for the Study of Documents and Manuscripts project.\textsuperscript{30} These include tools for selecting, viewing and manipulating images; for creating annotations on images; and for taking part in collaborative discussions. Since their creation for the VRE project, these components have been

\textsuperscript{25} A practical distillation of RESTful service design principals is provided by IBM’s Alex Rodriguez, at http://www.ibm.com/developerworks/webservices/library/ws-restful

\textsuperscript{26} Open standards supported by Alfresco include REST, RSS, Atom publishing, JSON, OpenSearch, OpenSocial, OpenID, Web Services, JSR 168, JSR 170 Level 2, MyFaces, CIFS, FTP, WebDAV, SQL and ODF

\textsuperscript{27} CMIS: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cmis

\textsuperscript{28} OpenSocial: http://code.google.com/apis/opensocial/

\textsuperscript{29} Apache Shindig: http://shindig.apache.org/

\textsuperscript{30} Virtual Research Environments for the Study of Documents and Manuscripts: http://bvre.humanities.ox.ac.uk/VRE-SDM
ported to run as a collection of gadgets in an OpenSocial environment and will be contributed to the Project Bamboo ECM Work Spaces. The CMIS standard will be used to integrate OpenSocial gadgets with the Alfresco platform.

Berkeley is developing campus content management capability (Media Hub\(^{31}\)) using the Alfresco Share application. The Media Hub is the product of several years of collaboration between campus departments, research centers, and museums and Berkeley’s Information Services and Technology (central IT) department, the UC Berkeley Library, and the California Digital Library. Berkeley’s selection of Alfresco is based on a detailed needs analysis and platform evaluation process using functional, business and technical requirements gathered over a one year pilot program.

The Bamboo ECM Work Spaces team will augment, customize, and extend the Alfresco Share application to meet the needs of arts and humanities scholars. With little or no customization, the Share platform can meet the first seven of nine functional requirements listed in Section 5.3.1, Core Work Spaces Capabilities. The last two requirements will require somewhat more significant customization:

- call on remotely deployed web-services to operate over an object or objects;
- record information on methods, procedures, and tools used in scholarly research, and report to one or more central aggregation stores

Customizations to call web services will be implemented using the Alfresco Web Script Framework.\(^{32}\) Content usage metrics will be generated by Alfresco’s embedded rules engine, accessed by RESTful APIs built into the platform, and contributed to aggregated registries on the Bamboo Services Platform and/or the Tools and Services registry using a set of customized web-services calls.

At Berkeley, pilot participants for the ECM Workspaces will be drawn from arts and humanities scholars who are current or prospective users of the Media Hub service, and from projects associated with the Berkeley Humanities and Arts Research Technologies (HART) initiative.\(^{33}\) Among the initial users of the environment within a Virtual Research Environment at the University of Oxford will be members of the Centre for the Study of Ancient Documents\(^{34}\) working collaboratively on epigraphic and papyrological material.

### 5.3.5 Shared Tools and Services Information Registry

Participants in the Project Bamboo planning workshops called for an accessible and transparent venue where scholars, librarians, technologists, and campus resource

\(^{31}\) MediaHub: http://mvp.berkeley.edu/wiki/About_MediaHUB


\(^{33}\) HART: http://inews.berkeley.edu/articles/Spring2008/1186.html

\(^{34}\) Centre for the Study of Ancient Documents: http://www.csad.ox.ac.uk/
managers can learn about existing tools for the humanities, identify areas of
technology need, specify requirements, and evaluate solutions as they evolve.
Traditional “waterfall” methodology\(^{35}\) – in which prospective users (scholars) hand off elaborate specifications to technologists and wait to see what comes back – was broadly rejected as inadequate to addressing evolving scholarship. Participants identified need for iterative and collaborative assessment and revision processes that build upon modern agile and user-centric approaches to software design and implementation.\(^{36}\)

In response to this identified need, we will implement a Tools and Services registry to share information and assessment about scholarly engagement with technology. This registry maps to two elements of work considered and evaluated by participants in Bamboo Planning Workshop Four, held in April 2009: a "Tools and Content Guide" and a "Services Atlas." These proposed elements of work were judged to be of High or Medium importance by 86% and 93%, respectively, of workshop participants who rated them; the “Service Atlas” garnered the greatest number of High and Medium votes among eleven elements of work considered.\(^{37}\)

This registry will aggregate, in a common data store, description of technology and content contributed by users of any Bamboo-supported or Bamboo-enabled Work Spaces, subject to user-granted permission. The registry will also aggregate service registry\(^{38}\) and usage data from the Bamboo Services Platform. These data will include:

1. descriptive information about services deployed on the Bamboo Services Platform, both in human-friendly format, and in technically formal formats (e.g., WADL or WSDL service description standards\(^{39}\))
2. information about tools, services, and/or digital content contributed by scholars who use Work Spaces to describe, discuss, and assess technology applied to their research and teaching
3. references (e.g., URLs) to described tools, services, and/or digital content in cases where it is downloadable or integratable, and references to information about how scholars, librarians, and technologists can use them

---


\(^{36}\) Worthy Martin, Associate Professor of Computer Science and Acting Director of the University of Virginia’s Institute for Advanced Technology in the Humanities, participated in the full series of Planning workshops, and summarized the history and trajectory of these discussions in his “Community Design” blog post on the Bamboo Planning Wiki, at https://wiki.projectbamboo.org/display/BPUB/2009/06/13/Community+Design

\(^{37}\) Final poll results, Bamboo Workshop Four: https://wiki.projectbamboo.org/display/BPUB/W4+-+Polls#W4-Polls-17AprilDiscussionofSection3andSection4Poll%2332

\(^{38}\) The Bamboo Services Platform registry will describe services deployed on the platform itself; the Tools and Services registry will record a superset of this information.

\(^{39}\) WADL describes RESTful services: https://wadl.dev.java.net/. WSDL describes services that conform to SOAP web services standards defined by W3C: cf. http://www.w3.org/TR/wSDL and http://www.w3.org/TR/soap/
The Tools and Services registry will associate a set of comments, ratings, and reviews (assessments) with these descriptions, contributed by users of Bamboo technology (including Work Spaces and Scholarly Services), and any others who wish to participate in assessment activities. This set of registry features will address the call for collaborative and iterative assessment expressed by Bamboo planning workshop participants.40

A clarifying analogy might be to consider that the Tools and Services registry will function similarly to Amazon.com, insofar as Amazon provides a variety of book descriptions, as well as ratings and free-text reviews by anyone who wishes to contribute them. Instead of the wide universe of books that Amazon.com and its consumer community describes and rates, Bamboo’s Tools and Services registry will focus on technology and content of interest to humanist scholars. Analogous to Amazon’s and similar forums’ affordances for identifying trusted or expert reviewers, profile services on the Bamboo Platform will collect and expose a limited, user-approved set of information that will enable consumers of information from the Tools and Services registry to filter assessment based on evolving trust networks among Bamboo users (e.g., assessment by scholars in similar disciplines, by reputable librarians, etc.).

Bamboo will enable contribution to and harvest/display of Tools and Services registry information from a variety of web sites, applications, Work Spaces, and environments:

- Bamboo will build a web-browser interface for general access.
- Bamboo will build a RESTful web-services API to permit integration of data contribution and access functionality with other environments, including Work Spaces built on both the HUBzero and Alfresco platforms.
- Bamboo will build an OpenSocial-compliant API to permit access and contribution via “gadgets” embedded into social networking and/or virtual research environments.

Because the aggregated data store will be accessible across institutions, disciplines, and Work Spaces, assessment activity can include broad ranges of dialogue and perspective. The Tools and Services registry will enable evolution of rich description and critique as institutions affiliated with and interested in the Project Bamboo Program evaluate and adopt the evolving products of technology implementation projects described in this proposal.

---

40 Worthy Martin, ibid.
Work Spaces deployed centrally or at local campuses contribute and consume data to/from the Tools and Services registry.

The diagram above illustrates an ecosystem in which:

- Work Spaces utilize Scholarly Services on the Bamboo Services Platform and contribute data about use of those services to the Tools and Services registry;
- Work Spaces enable end-users to record information and assessment about scholarly engagement with technology, and to share that information more broadly via the Tools and Services registry;
- a browser-based user interface enables access to information from the Tools and Services registry for anyone who can use a web browser;
- RESTful and OpenSocial interoperability standards provide the ‘glue’ between these loosely-coupled elements, as well as between the Tool and Services registry and social networking or virtual research environments for which integrations may be developed beyond the scope of this proposal.

During the grant period (through March 2012), UC Berkeley, with participation of other Work Space institutions, will deploy and maintain the Tools and Services registry. As noted above, initial contributors to the registry will be users of Bamboo technology, including users of Scholarly Services and Work Spaces. Management (curation) of the data contributed to the registry will be provided by individuals and institutions affiliated with the Project Bamboo Consortium. Hosting of the Tools and Services registry will transition to the Project Bamboo Consortium as an element of the sustainability plan for technology implemented during the grant period covered by this proposal (see section 10 below).
5.4 Corpsa Space: Phase 1 design

In Phase 1, Corpsa Space design work will explore the needs, possibilities, and challenges required for implementing powerful web-based applications for research across multiple and dispersed corpora. This work will require close coordination among all areas of work in this project. The primary deliverable for the Corpsa Space design phase will be detailed roadmaps for building approximately two Corpsa Space applications.

The Corpsa Space design effort will focus first on people, next on exploration, and last on decision-making in preparation for the Corpsa Space implementation phase.

The Corpsa Space design effort will begin by recruiting a larger team to participate in the design effort. The current (planning phase) Corpsa Space participants (University of Maryland (project lead), Tufts University, University of Oxford, Northwestern University, University of Wisconsin at Madison, University of Chicago, and UC Berkeley) will identify a wide group of individuals and institutions in the humanities who are interested and able to consult about this effort, thus expanding the design team. The team will also consult with consortia (e.g. CLARIN, DARIAH and CHAIN) that can provide insight, guidance and experienced counsel for Corpsa Space design.

Once the full design team is assembled, its first order of business will be to leverage what already exists. The team will identify potential corpora, collections, projects and tools for consideration as part of Corpsa Space design. The team will develop an appropriate Corpsa technology evaluation framework to help establish a common language for assessing existing and future applications. Design activity will include assessment of current corpora-based applications in the humanities and their future development and support plans; exploration and discussion with faculty across disciplines about the use of shared corpora applications; lessons learned from other projects that are attempting to build a common platform for multiple application instances across disciplines and data models (e.g. CollectionSpace in the museum domain); identification and evaluation of corpora that may be ready for Corpsa Space (e.g. the University of Virginia and University of Michigan’s digital library collections); and technical consultation with other related services environments (e.g., those of CLARIN and SEASR). Examples of ongoing scholarship and scholarly technologies we will consult with include Nines, Perseus, ARTFL, TAPoR, and Oxford’s JISC-supported VRE-SDM. We will also consider more recent projects, such as Berkeley Prosopography Services and its relationship to the Cuneiform Digital Library.

Once the overall parameters and the most promising opportunities have been identified, the design team will recruit 3-6 groups to develop white papers about possible Corpsa Space applications. We will request that these white papers take into account opportunities and lessons learned Bamboo’s work in Work Spaces,
Bamboo scholarly services and services platform, and collections interoperability. The Corpora design team will review the white papers and, based on technical fit, corpora readiness, broader benefit to the humanities community, and other criteria, recommend one or two applications for implementation during Phase 2. The Bamboo Project Executive Group and Steering Council will then make the final determination regarding which candidates to implement. During this process we will actively share results with the Mellon Foundation and solicit guidance on possible areas of interest.

To complete the design phase, thorough technical and project development roadmaps will be created for the chosen projects. Roadmaps will identify deliverables, technology, corpora, consortial partnerships, resource requirements and a development timeframe. Roadmaps will also include early design elements such as wireframes and data models.

5.5 Scholarly Services on Bamboo Services Platform

This area of work includes three activities: (1) developing scholarly web services that model and support scholarly research methods; (2) assembling, developing, and deploying the Bamboo services platform; and (3) defining processes for contributing services to the Bamboo platform and releasing a software development toolkit. These activities are treated separately in the description below, and in the work plan and time line in section 8 and appendix 1.

5.5.1 Scholarly Services

Scholarly web services will support research and teaching in the humanities. These services will be of use across a wide range of disciplines, and by being deployed on the Bamboo platform with SOAP or RESTful web services interfaces, they can be integrated with diverse external tools, applications, content repositories, and platforms, facilitating further, broad uptake of valued technology. The services will be engineered to offer robust and scalable capabilities to support this integration.

Scholarly web services built and deployed in phase 1 will be derived from existing, proven projects in the humanities, including ARTFL, Perseus, Berkeley Prosopography Services, and SEASR.\(^{41}\) We will “factor out” broadly applicable capabilities and algorithms from these original contexts and repackage them for deployment as scholarly web services on the Bamboo Services Platform. Utility of the selected capabilities has been and continues to be demonstrated by ongoing scholarship that relies on projects from which these services will be refactored.

---

\(^{41}\) ARTFL (http://artfl-project.uchicago.edu/); Perseus (http://www.perseus.tufts.edu); Berkeley Prosopography Services (http://wikihub.berkeley.edu/x/2YYAAQ); SEASR (http://seasr.org/)
Services will be coded as OSGi bundles of Java classes that can be deployed in the Bamboo Service Platform’s OSGi container. Each service will be built by a cohesive team affiliated with the institution where the existing humanities projects is based, and integrated into the Bamboo Services Platform hosted at UC Berkeley on behalf of Bamboo (cf. Section 5.8, Deployment Strategies for the Bamboo Services Platform).

The organizations supporting the original tool, application, etc., may then choose to refactor their technology to consume a Bamboo-hosted capability. However, the broader value of scholarly web services on the Bamboo services platform will be to expose algorithms and capabilities to new environments, tools, and applications – and the scholars who use them -- without requiring that functionality be reinvented.

Scholarly services to be delivered in Phase 1 include the following, to be implemented by teams based at the indicated Partner institutions (further detail in Section 7, Staffing Plan):

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Scholarly Services Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Mapping</td>
<td>map features in documents to an object-relational indexing model</td>
<td>Chicago/ARTFL</td>
</tr>
<tr>
<td>Concordance</td>
<td>generate a concordance for one or multiple texts, where each element of the concordance is associated with contextual material drawn from its occurrence(s) in the analyzed text(s)</td>
<td>Chicago/ARTFL</td>
</tr>
<tr>
<td>Collocation/Cloud</td>
<td>return word-counts of matches occurring in a text or texts to a specific query term or terms</td>
<td>Chicago/ARTFL</td>
</tr>
<tr>
<td>Frequency</td>
<td>return word-occurrence frequency in a text or texts, grouped by metadata elements (e.g., author) and as a statistic normalized to a specified quantity of text (e.g., 10,000 words)</td>
<td>Chicago/ARTFL</td>
</tr>
</tbody>
</table>

42 OSGi (formerly the “Open Services Gateway Initiative,” but the name is no longer associated with any acronym) is defined in a nutshell on the OSGi Alliance FAQ: “The OSGi specifications define an in-VM Service Oriented Architecture (SOA) for networked systems. An OSGi Service Platform provides a standardized, component-oriented computing environment for cooperating networked services. This architecture significantly reduces the overall complexity of building, maintaining and deploying applications.” Cf. http://www.osgi.org/About/FAQ#q6

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Description</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological Analysis</td>
<td>generate a morphological analysis of an inflected word supplied by the service consumer</td>
<td>Tufts/Perseus</td>
</tr>
<tr>
<td>Syntactic Analysis</td>
<td>generate a syntactic analysis of words in a sentence or set of sentences</td>
<td>Tufts/Perseus</td>
</tr>
<tr>
<td>Named Entity Identification</td>
<td>generate semantic classification of named entities in a text or set of texts (i.e., associate a name in a text with a particular entity in the real world)</td>
<td>Tufts/Perseus</td>
</tr>
<tr>
<td>Proxied SEASR Analytics</td>
<td>configure and initiate analytical services deployed alongside major content repositories, such as HathiTrust or JSTOR, where co-location of computational power permits analysis of very large data sets or data whose intellectual property constraints prohibit a researcher's direct possession of content</td>
<td>UIUC/SEASR</td>
</tr>
</tbody>
</table>

As indicated in this proposal’s overview (Section 4), participants in the Bamboo Planning workshops identified categories of scholarly need that map to curatorial, analytic, semantic, and visualization services. In phase 1, Bamboo will deliver analytic services that operate on digitized textual materials, focusing in this category because such services:

- are readily realizable by Project Bamboo partner institutions within the phase 1 timeframe;
- will provide value to text-focused scholars by expanding accessibility of the selected, already-proven algorithms;
- will prove the value and workability of the component-services model for building technology that this proposal represents;
- will set the stage for future Scholarly Services implementations that address additional corpora and capabilities.

It is noteworthy that each of the service categories – curatorial, analytic, semantic, and visualization – can be mapped to thematically grouped areas of scholarly practice surfaced during the Bamboo Planning workshops. These mappings suggest rich opportunities for supporting humanist inquiry with future Scholarly Service implementations.

**Curatorial services** enable scholars to track and organize digital materials they wish to archive and preserve. Digital materials includes (digitized) primary sources, but also records of scholarly methods applied to materials in the course of an
inquiry, records of a performance, records of relationships between objects that change over time, geo-spatial data, and visualizations. The constellation of activities involved in tracking and organizing digital materials, as described by participants in Bamboo Planning workshops, is documented on the project wiki in Theme Groups “Consider” and “Preserve.” These Theme Groups map back to the following constituent Themes: “Organizing Information,” “Store, Archive, and Preserve,” and “Changing Technologies and Unsustainable Tools.” Theme pages on the Bamboo Planning wiki quote and cite relevant papers and conference proceedings, such as Cathy Marshall’s “From Writing and Analysis to the Repository: Taking the Scholars’ Perspective on Scholarly Archiving” and the final report from the 2005 Summit on Digital Tools for the Humanities, held at the University of Virginia. Excerpts from cited materials and participant statements, transcribed during table-discussions and/or recorded on flip charts, are also included. A sampling of these follows:

- “Finding a secure, persistent place for storing resources. Should be like Library of Congress Minerva project.”
- “Archives of conference papers, electronic working papers like the sciences do. Registry of research activities [...] pre-print, awareness of area of work and particular argument, approach, methodology. A 10-page searchable "visiting card". ... Store and retrieve personal research, being able to find what you’ve already digested somewhat and make sense of it. [...] Has both a mechanical and archival quality as well as a summary and reflective side”
- “Dating materials (texts, artifacts, etc., events - chronological); assign/create metadata ... Provenancing materials - Primary and secondary materials; Being able to explain where the materials came from, and what processes were used to create”

Analytic services enable scholars to perform algorithmic inquiries against corpora of digitized materials. While the majority of analytic software applicable to humanist research enables discovery of pattern and structure in text, this category of technology support for the humanities also includes algorithmic analysis of image, video, and audio material. Scholarly activity in this area, as described by participants in Bamboo Planning workshops, is documented in the "Consider" and "Discover" Theme Groups, and their constituent Themes, "Analyze," "Contextualize," and "Filter and Synthesize" on the project wiki. Theme pages on the wiki quote and cite applicable papers and conference proceedings, such as the final report from the 2005 Summit on Digital Tools for the Humanities. Excerpts from cited materials and participant statements, transcribed during table-discussions and/or recorded on flip charts, are also included. A sampling of these follows:

---

44 Theme Groups on the Bamboo Planning wiki: https://wiki.projectbamboo.org/display/BPUB/Theme+Groups
45 Themes on the Bamboo Planning wiki: https://wiki.projectbamboo.org/display/BPUB/Identify+Themes
• "Critical analysis of 9 volumes of French edition, 6 volumes in Dutch, 1 volume in German edition. Want to use technology to do literary analysis beyond what's possible for 1 human. Analysis and understanding of voluminous text. Compare & contrast maps is akin to comparing text"
• "Find patterns. Discover relationships. Analyze data (linguistics: morphological analysis). Calculate similarity metrics. ... Identify causal relationships. Identify stylistic patterns. Aggregate critical editions (analyze secondary, identifying "State of the Art")."
• "Identifying patterns, relationships; figure out how texts fit together; contextualize the object"
• "Digital representations of material can be searched, analyzed, and altered at electronic speed. More dramatically, they lead to orderly cooperation by many, perhaps hundreds or thousands, of individuals. And, all of these collaborators can access and edit the same representation of data from geographically distant sites." ("Summit on Digital Tools for the Humanities," p. 15)
• "Assessing relevance. This is the whole issue of sifting the wheat from the chaff"

Semantic services -- which can be viewed as a specialized category of analytic services -- enable scholars to algorithmically identify meaning in digital materials, and relationships between materials derived from algorithmically- or manually-associated meaning. Semantic services are categorized distinctly from analytic services in this proposal because technology that derives meaning from digital objects is currently less sophisticated (and more provisional) than analytic technology used to discover structural patterns in digitized corpora. While the VIVO semantic web engine will be integrated into HUBzero Work Spaces (pending NSF funding), as described in section 5.3, a roadmap has not been developed at this stage of semantic technology's evolution for how Bamboo might implement generalized semantic services of demonstrable value across broad areas of humanist inquiry. Though semantic services are not in scope for phase 1 implementation, participants in Bamboo Planning workshops have identified scholarly interest in this area. This interest is documented in the "Consider" and "Discover" Theme Groups, and their constituent Themes, "Conceptualize" and "Filter and Synthesize" on the project wiki. Theme pages on the wiki quote and cite applicable papers and conference proceedings, such as the final report from the 2005 Summit on Digital Tools for the Humanities. Excerpts from cited materials and participant statements, transcribed during table-discussions and/or recorded on flip charts, are also included. A brief sampling of these follows:

• "Grand Narrative of Humanities Scholarship: Condensing meaning from vapor of nuance"
• "Tokenize, segment the resource (automatically or manually); Name and rename parts; Align annotation with parts (including time-based material); vary or match the notation of the original content; Sort and rearrange the resource (perhaps in something as formal as a semantic concordance, perhaps just in some unspecified relationship); Identify and analyze patterns that arise out of relationships; Code relationships, perhaps in a way that encourages the
emergence of an ontology of relationships (Allow formalizations to emerge, or to be brought to bear from the outset, or to be absent)" ("Summit on Digital Tools for the Humanities," p 7)

**Visualization services** enable scholars to model and present large or complex sets of information in graphical forms that reveal patterns at a strategic distance, without first requiring a viewer to digest and comprehend each element of the set. This category of scholarly activity, as described by participants in Bamboo Planning workshops, is documented in the "Consider" Theme Group, and its constituent Themes, "Model and Visualize" and "Analyze" on the project wiki. Theme pages on the wiki quote and cite applicable papers and conference proceedings, such as Willard McCarty’s "Modeling: A Study in Words and Meanings" and the final report from the 2005 Summit on Digital Tools for the Humanities. Excerpts from cited materials and participant statements, transcribed during table-discussions and/or recorded on flip charts, are also included. A sampling of these follows:

- "Making the invisible visible. Show & tell (other than through text) […] Changeable modalities of representation […] Virtual unification of physically disparate materials/objects […] making virtual whole from parts"
- "Visualization techniques are up and coming practices. To compare content that otherwise had to be described. Also used for ethnography, creating websites that map out terrains."
- "Discussion of […] various approaches to this cluster of related issues led us to think of the solution to the problem [of Visualization of Time, Space, and Uncertainty] as entailing not so much a new software tool as a software 'machine,' i.e., an integrated suite of tools. This machine would allow us to collect data, evaluate them as to their reliability/probability, set them into a time frame, define their temporal relationships with other features of interest in our study, and, finally, represent their degree of (un)certainty by visual conventions of stylized rendering and by the mathematical expressions stated in fuzzy logic or some alternative representation." ("Summit on Digital Tools for the Humanities," p. 21)

**Planning for Further Scholarly Services:** Partner institutions in this proposal will identify and track opportunities for additional service development within these and emerging categories. The Scholarly Services and Bamboo Platform team (see section 6.4) will be responsible for coordinating these activities. They will not only draw upon the expertise of partners in the project, but work with the Humanities

---

Faculty Committee and Technical Advisory Committee\textsuperscript{49} to solicit expertise and advice about work and discussion occurring outside Project Bamboo's participant institutions, projects, and disciplinary expertise. The Project Executive Group, described in Section 6.1 of this proposal, will prioritize identified scholarly service development and integration opportunities, considering the value to scholarship, technical feasibility, and prospects for broad adoption in deciding which additional services to integrate during phase 2.

5.5.2 Bamboo Services Platform

The Bamboo Services Platform (BSP) includes the technology stack on which software services will be deployed; as well as services that address general functionality that underlies more specialized Scholarly Web Services.

The Deployment Stack includes language, framework, and libraries (e.g., service container, service implementation libraries, authentication framework, logging framework, message mediation). Bamboo will rely on an OSGi\textsuperscript{50} standards-compliant service container as a cornerstone of the platform's service deployment infrastructure. OSGi allows services based on diverse software dependencies to be co-deployed with minimal effect on implementation code. This platform quality is essential to a model in which multiple software development groups contribute services to be deployed on the Bamboo Services Platform. We will use the Progress FUSE\textsuperscript{51} stack as major elements of technology at this level of the Bamboo Services Platform. FUSE technologies are controlled releases of Apache Foundation Service Oriented Architecture (SOA) infrastructure projects, such as the ServiceMix enterprise service bus (ESB).\textsuperscript{52}

Services that meet general functional needs can be classified as core and utility services. Core services include general functionality such as modeling and maintaining indexes of digital content and metadata modeled around persistent URIs, persons, organizations, and groups in a Bamboo context. Utility services handle basic, cross-cutting functionality required by higher-level services, such as authentication, storage, generation of persistent URIs, and usage tracking.

More details and a diagram of the Bamboo Platform architecture can be found in section 5.7 below.

\textsuperscript{49} See sections 6.1 and 10.3 for more information about these advisory committees.

\textsuperscript{50} OSGi (formerly the "Open Services Gateway Initiative," but the name is no longer associated with any acronym) is defined in a nutshell on the OSGi Alliance FAQ: "The OSGi specifications define an in-VM Service Oriented Architecture (SOA) for networked systems. An OSGi Service Platform provides a standardized, component-oriented computing environment for cooperating networked services. This architecture significantly reduces the overall complexity of building, maintaining and deploying applications." Cf. http://www.osgi.org/About/FAQ#q6

\textsuperscript{51} Progress Fuse: http://fusesource.com/

\textsuperscript{52} Apache ServiceMix: http://servicemix.apache.org
5.5.3 Shared Services Lifecycle and Software Developer Kit

To run on the Bamboo Services Platform, software must meet certain criteria. A Shared Services Lifecycle will define the qualities and processes through which software must evolve as it acquires the “services architecture” needed for operation on the Bamboo Services Platform. It will include criteria for identifying candidates for inclusion in a family of scholarly web services: value to scholarship, conformity with appropriate standards (software or data structure), and technical feasibility.

Evolution through the Shared Services Lifecycle will broaden and deepen the value of software to scholars and improve the long-term viability of the technology. Currently, much technology created for worthwhile purposes is narrowly tailored and entails idiosyncratic technical requirements, non-standard data formats, or atypical deployment. By helping to establish the best software engineering practices in increasingly rigorous phases, Bamboo will help to transform the development of new technology for humanities scholars and encourage refactoring in partner projects. The stages of services evolution can be described in this way:

1. Early interest in, adoption by, or development by the Bamboo community;
2. Software refinement to addresses concerns like services architecture, compliance with standards, reliability, and conformity to deployment requirements of the Bamboo Services Platform;
3. Deployment on the Bamboo Service Platform and reliable availability there.

We will assemble and document a Software Development Toolkit (SDK) suitable to building services for deployment on the Bamboo Services Platform from best-of-breed open-source candidates (e.g., Maven, Eclipse). This kit will be used and refined by service developers engaged in the work defined by this proposal, and it will be released as a package for future use by software developers, and the campus library and IT organizations that work with them, to create services tailored to the needs of humanities scholars and projects. The SDK will include key process documentation, along with training materials and software development tools. It will be useful to individuals and projects in the near term; over time, it will help to build a community of technologists who share training in the best way to understand scholarly needs in the humanities and ensure that those needs are met. We will select and deploy a set of fundamental collaborative software development tools (e.g., a development wiki, issue tracker, and version control system for source code) to create a basic infrastructure for Bamboo service development, and we will establish a model for service development in support of the humanities going forward.

5.6 Collections Interoperability Services

Collections interoperability services deployed on the Bamboo Services Platform will deliver capabilities that are likely to fall in the following categories:
• metadata interchange using metadata interoperability and complex content description standards, such as RDF,53 OAI-PMH,54 OAI-ORE,55 ATOM Publishing Protocol,56 OAC,57 etc.
• facilities that exploit and support use of persistent resource URIs at appropriate levels of granularity and adhering to the principles and emerging best practices of the W3C Linked Data Initiative58
• proxied access to, aggregation of, or aggregated indexing of content stored in repositories that expose a CMIS-compliant 59 web services interface
• facilities to support updating and/or augmenting, in place or virtually, existing collections and/or collection metadata as transformations and new scholarly derivatives of content are generated
• transformation between broadly-used content formats, where need exists and opportunity permits, especially with respect to transformations that are prerequisite to consuming scholarly services deployed on the Bamboo Services Platform

The first three categories in this list address content metadata – the 'lower hanging fruit’ in Collections Interoperability. Standards in these areas include those mentioned above as well as metadata semantic schemas, such as DCMI (“Dublin Core”)60, MODS (“Metadata Object Description Standard”),61 and LIDO (“Lightweight Information Describing Objects”),62 and a spectrum of approaches to handling persistent references to digital content.63 While these standards and related practices are reasonably well-understood by humanists, librarians, and information technologists, experience suggests that profiling and more detailed guidelines and best practices will be required to support the level of content exchange and interactivity envisioned for Bamboo.

The fourth category in the list above addresses the increasingly dynamic nature of digital scholarly content. As scholars interact with distributed digital resources through Bamboo, they will create new relationships between content objects, new transformations of content and new derivatives of content. Bamboo will develop services that will gather usage data from these interactions and will help provide the basic information required to promulgate updates to resource metadata and the

53 RDF (Resource Description Framework): http://www.w3.org/RDF/
54 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting): http://www.openarchives.org/pmh/
55 OAI-ORE (Open Archives Initiative Object Reuse & Exchange): http://www.openarchives.org/ore/
57 OAC (Open Annotation Collaboration): http://www.openannotation.org/
58 http://www.w3.org/standards/semanticweb/data
59 CMIS (Content Management Interoperability Services): http://docs.oasis-open.org/cmis/CMIS/v1.0/os/cmis-spec-v1.0.html
60 DCMI (Dublin Core Metadata Initiative): http://dublincore.org/
61 http://www.loc.gov/standards/mods/
62 http://www.lido-schema.org/
63 Cf. “Persistent Identifiers, Considering the Options,” (Emma Tonkin, Ariadne, Issue #56, July 2008); http://www.ariadne.ac.uk/issue56/tonkin/
potential deposit of new resource derivatives in ways that maintain provenance and intellectual cohesion.

In approaching problems that require interoperability of content at a structural and semantic level, the fruit grows higher on the tree. TEI, for example, addresses structure of textual documents in a manner that enables some algorithmic operation across textual collections or corpora. Yet TEI’s diverse variants, which have in many cases evolved to address concerns of real import to adopting scholars, gives rise to much frustration among textual researchers in the arts and humanities – and more so when complicated by uneven mapping of semantic meaning to structural markup. Interoperability hampered by differences in point-of-view is even less tractable (is the Royal Pavillion in Brighton, England situated on Le Terrain Crétacé or in the County of Sussex?)

Proposed Collections Interoperability standards selection and adoption work will address these areas. To support selection, adoption, profiling, and development of digital object identifier and metadata interoperability standards and practices, some limited experimentation and interaction with both the Work Spaces and Scholarly Services tasks will be necessary to inform this process. We start with somewhat less of a foundation in regard to our understanding of the specific ways that user and service interactions with content and collections can extend and augment content metadata and sometimes the resources involved directly. We anticipate that substantive work on this issue will extend into the second 18 months of the project. The final category in the list above is a nod to initial service development progress possible in the near term on the more difficult set of interoperability problems, charting an achievable, high-value course between utopian goals (“everything will be interoperable”) and solutions that apply only to a single collection or small groups of corpora.

5.7 Architecture of the Bamboo Services Platform

Software services (including web services) in a "Service Oriented Architecture" (SOA) implement functionality that has been decomposed then logically grouped to flexibly facilitate interoperability with other software and/or digital content; and/or to be flexibly combined with other services in support of multiple tasks or workflows. Depending on its generality and the contexts in which it is to be utilized, a service will be associated with a particular architectural layer of services that have a similar generality or specificity. In the broad Bamboo context, in which capabilities are organized around digital content of interest to scholars of arts and

---

humanities, we anticipate that most web service interfaces will be modeled as operations on digital resources, and will conform to "RESTful" design principles.65

---

**Bamboo Architecture Layers**

**User Interfaces** - *Scholarly Application Interfaces* will be browser-based interfaces for scholarly applications created in phase 2 of the Bamboo technology project. *Services Registry, Usage Reporting, and Platform Services Testing Interfaces* will be a set of simple, browser-based interfaces by which service metadata can be accessed and service functionality can be verified; these are not expected to be the only, or even the primary means of accessing Bamboo services, but provide access for simple consumption and for evaluation conducted by potential service consumers. In each of these areas, sets of functionality and information may also be exposed through social networking platforms via plug-ins (a.k.a. "widgets" or "gadgets").

**Scholarly Applications** - *Scholarly Applications on the Bamboo Platform* add functionality to compositions of scholarly, interoperability, core, and utility services that are specific to particular scholarly uses. Applications realized as service

65 RESTful (REST = REpresentational State Transfer) principles for web service design are usefully discussed here: http://www.ibm.com/developerworks/webservi

compositions will be deployed on the Bamboo Services Platform infrastructure, and they may include proxy services to functionality that is provided elsewhere. Corpora Space application(s) planned for phase 2 of Bamboo's initial 3-year development may be realized at this level of the architecture.

**Scholarly and Collection Interoperability Services** - Scholarly services offer functionality of direct application to objects of scholarly interest, such as textual and prosopographic (social network) analyses, and they may include proxy services to functionality provided elsewhere. Collection interoperability services connect externally-hosted content to the Platform and its services via interfaces that conform to adopted interoperability standards. Scholarly and interoperability services will be deployed on the Bamboo Services Platform infrastructure.

**Core and Utility Services** - Core services include general functionality such as modeling and maintaining indexes of digital content and metadata modeled around persistent URIs, persons, organizations, and groups in a Bamboo context. Utility services handle basic, cross-cutting functionality required by higher level services, such as authentication, storage, generation of persistent URIs, and usage tracking.

**Deployment Stack and Infrastructure** - Deployment Stack includes language, framework, and libraries (e.g., service container, service implementation libraries, authentication framework, logging framework, message mediation). The Progress FUSE stack\(^66\) (composed from open-source, Apache-centric SOA infrastructure distributions) will form the core of this level of the Bamboo Services Platform. Infrastructure includes hardware/virtual-machine, operating system, storage platforms (ECM, relational database), SAN, and network.

### 5.8 Deployment Strategies for the Bamboo Services Platform

In phase 1, we will deploy the initial production release of the Bamboo Services Platform in a single instance hosted by UC Berkeley; partner institutions involved in Scholarly Service development will deploy development and integration instances. The single production instance is expected to include clustered and load-balanced servers at appropriate layers of the deployment infrastructure (e.g., Storage Area Network\(^67\)).

The rationale for a single production deployment is to defer resolution of data synchronization and other distributed architectural issues until open-source infrastructure conforming to Enterprise OSGi standards published in March 2010 has a chance to mature.\(^68\) In addition, during phase 1 we prefer to focus resources

---

\(^{66}\) Progress FUSE: FUSE: http://fusesource.com/

\(^{67}\) Storage Area Network (SAN): http://en.wikipedia.org/wiki/Storage_area_network

on providing a strong set of scholarly deliverables rather than on maximizing technical robustness.

Late in phase 1, we will commence exploration of hosting strategies and initiation of service design in support of distributed deployments of the Platform as steps toward fully distributed, synchronized or syndicated deployments to be delivered in phase 2. Exploration of hosting strategies will include consideration of cloud-hosting models treated by Indiana and Berkeley CIOs Brad Wheeler and Shelton Waggener in their November 2009 article in Educause Review, “Above-Campus Services: Shaping the Promise of Cloud Computing for Higher Education.”

6. Project Organization and Partner Roles

We have designed the organizational structure of this project to leverage the expertise and resources of its ten university partners so that we can deliver high quality results for each of the project’s major work areas and overall project goals. We recognize the challenges and risks of carrying out this project across multiple institutions in a number of countries. We believe that we have not only addressed these risks but that this particular project requires, and its strength comes from, a partnership model in which colleagues from the humanities, libraries, and IT across institutions work closely together to deliver results for phase 1 and prepare for the long-term sustainability of Project Bamboo’s efforts. We begin with a statement of key points about why the project’s overall structure will enable us to deliver high quality results, then provide details about each component of the project’s structure.

First, it is helpful to view the technical deliverables of this project as falling into two major categories: software development and software integration. Software development requires teams that are often tightly integrated and managed. Software integration requires technical staff from different projects who can work together at the right moments, and who can agree upon, implement, and test integration protocols and standards within and across projects. Integration projects are more loosely coupled than development projects, but require the right kind of community, incentives, and common language, often drawing deeply upon the skills of architects and projects managers in addition to those of software developers. Unlike many other community-source technology projects, this project is not just building one large application or tool, but rather developing, enhancing, and, critically, integrating several foundational platforms and applications. Our project structure is a reflection of these technical goals.

Complex software development described in this proposal takes place within an area of work, either Work Spaces or Scholarly Services and Bamboo Platform, but generally not between areas. That is, there is not major software development

within Work Spaces that relies on software builds within the Bamboo Services Platform on a daily or weekly basis, or vice-a-versa.

There is, however, important software integration between these major areas of work. Work Spaces will need to be able to access services from Bamboo Services Platform, for example. To accommodate the project’s substantial software integration needs, we have not only built in staff time for this within each relevant area of work, but we have created the Collections Interoperability work team to provide additional specialized expertise and consulting related to those forms of integration in which software services and collections come together.

Second, for each of the project’s four major areas of work we have designed teams appropriate to realize the deliverables of that area. The Work Spaces and Scholarly Services on the Bamboo Services Platform areas of work are the two largest and most technically complex. We have made the most substantial commitment of resources here and identified lead institutions and lead staff for each area of work and sub-area. We have determined the technical and project management resources required to complete deliverables within each area and to integrate deliverables across areas.

Third, we note that this project benefits from the extensive planning and community building that the Bamboo Planning Project, generously supported by the Mellon Foundation, has afforded us. The partners in this project know each other and, in a number of cases, have built strong relationships. We have established an important foundation of trust, common language, and common commitment. These are essential to being able to start the work of this complex project quickly and to carry out it out across different teams and time zones.

Fourth, many of the institutional project teams carrying out the work of this project bring together staff not just from one area of their institution, but from humanities centers, libraries, IT organizations, e-research centers, and the like. This allows the project to draw on expertise and guidance from multiple parts of a campus. In the area of Collections Interoperability, for example, where the expertise and historical context of the library community are essential, we can turn to leaders with deep experience and connections in this domain. Project Bamboo will similarly benefit from the engagement of humanities centers and campus technology organizations.

Fifth, throughout the project model we have made a major investment in technical project management and architecture. Individuals filling these roles are essential to provide day-to-day guidance and “glue,” both within an area of work and, especially, between areas of work.

Sixth, we are making a major investment in coordinating and communication processes across the project as a whole and within each area of work. We will be using a number of technical project management processes and tools to facilitate project planning, tracking, and assessment, and to foster harmonious interaction
within teams. We will make extensive use of the project wiki, web sites, mailing lists, collaborative writing and review practices, so that all and any participants (as well as other interested parties) can see and contribute to project progress.

Seventh, this project's use of technical project management and architecture resources, as well as coordination and communication processes, will build on lessons learned in higher education about collaborative, community-source development from projects such as Kuali, Sakai, CollectionSpace, and OpenCast.

Eighth, we believe that we have brought together in this partnership talented and motivated individuals from each institution. This group is committed to working across institutional lines to realize the project's goals at a very high quality, and to build a sustainable partnership model for support the long-term vision of the Bamboo Strategic Program.

### Bamboo Technology Project - Project Structure

<table>
<thead>
<tr>
<th>Work Spaces</th>
<th>CorporaSpace</th>
<th>Services on the Platform</th>
<th>Collections Interoperability</th>
<th>Tools and Services Registry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area Coordinator</td>
<td>Area Coordinator</td>
<td>Area Coordinator</td>
<td>Area Coordinator</td>
<td>Noah Wittman* (Berkeley)</td>
</tr>
<tr>
<td>Bruce Barton* (Wisconsin)</td>
<td>Neil Fraistat* (Maryland)</td>
<td>Steve Masover* (Berkeley)</td>
<td>Tim Coe* (UIUC)</td>
<td>Bruce Barton* (Wisconsin)</td>
</tr>
<tr>
<td>Debjani Ganguly* (ANU)</td>
<td></td>
<td>Kaylea Champion* (Chicago)</td>
<td></td>
<td>Bruce Barton* (Wisconsin)</td>
</tr>
<tr>
<td><strong>HULzero Work Space</strong></td>
<td><strong>Contributing Institutions</strong></td>
<td><strong>Scholarly Services</strong></td>
<td><strong>Philologic Services</strong></td>
<td><strong>Ancient Studies Services</strong></td>
</tr>
<tr>
<td>Bruce Barton* (Wisconsin)</td>
<td>Berkeley</td>
<td>Services Coordinator</td>
<td>Bob Griffith (Chicago)</td>
<td>Greg Crane* (Tufts)</td>
</tr>
<tr>
<td>Robert McDonald (Indiana)</td>
<td>Chicago</td>
<td>Kaylea Champion* (Chicago)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The organization chart above shows the major project structure. In what follows in section 6 we describe the leadership structure of the project, the organization of the four areas of work, and then note a number of additional cross-cutting coordinating bodies that will help to integrate project activities. In section 7, we list all of the staff that will be part of these leadership and project teams.

6.1 Project Leadership

UC Berkeley will act as the managing partner to coordinate all project activity, manage the project funds, and act as the lead institution. Partner institutions will contribute within specific areas of work as either leaders or contributors, coordinating their activities with Berkeley in order to fulfill the goals of Project Bamboo. The Berkeley management team will include individuals in these roles: Principal Investigator, Executive Director, Project Manager, Administrative Support, and Financial Staff.

A Steering Council will guide the strategic directions of the project as a whole and ensure its continued alignment with the long-term strategic vision of Project Bamboo. The Steering Council will also be responsible for maintaining strong relationships with senior levels of institutional leadership, including leadership in technology, in libraries, and in the humanities. The Steering Council will be made up of senior humanities, library, and/or technical leaders from each partner institution, as well a number of external representatives. This group will meet as a minimum in quarters one, three, and five of the first 18-month project.

A Project Executive Group, made up of Project Leads from each major area of work and chaired by the Executive Director, will meet monthly to ensure that work in each area is on track and that work in different areas is coordinated.

We note as well that the Project Bamboo Consortium will convene a Humanities Faculty Committee and a Technical Advisory Committee to support Project Bamboo’s efforts as a whole. (Please see Section 10 on Long-Term Sustainability for an overview of the Project Bamboo Consortium and section 10.3 for a preliminary list of possible committee members.) These two committees will assist this project by communicating with a broad community of scholars, technologists, projects, organizations, and institutions worldwide. These committees will also help with the vital work of accelerating adoption of both the Bamboo services and the Bamboo philosophy of promoting sharable, reusable, open-source, interoperable technology.

6.2 Work Spaces

Work Spaces will be made up of two projects and project teams. Each project team will adapt existing collaborative and content management technologies to create Work Spaces for humanities scholars that can be efficiently run by campus organizations that provide technology support for these scholars. The project teams will do this by enhancing the user interface and adding to back-end technologies so
the Work Spaces can access web services, share information, and share content. Each project team will include, at a minimum, a project lead, technical project manager, and a software developer.

The two Work Space projects and the institutions leading them are:

1. **HUBzero. University of Wisconsin, Madison and Indiana University.** This project will build upon the substantial accomplishments the HUBzero project (http://hubzero.org/) has achieved for the sciences and evolve a collaborative hub for the humanities. UW Madison and Indiana Universities are consortium partners for the HubZero project and will lead this project team.

2. **Enterprise Content Management Platform. UC Berkeley, University of Oxford, and Australian National University (ANU).** This work space will be built on top of the open-source Alfresco enterprise content management (ECM) platform. We will use a number of front-end, user-facing tools connected to the back-end Alfresco ECM platform. These will include the OpenSocial framework (which will enable compatibility and connectivity with Sakai 3 and Google, as important examples), Drupal, and Alfresco's Share application. This project team will build upon work UC Berkeley is carrying out to develop research-driven collaborative and content management environments for multiple disciplines. It will draw upon Oxford’s detailed prototype and specification for a virtual research environment (VRE) for the humanities. (See http://bvreh.humanities.ox.ac.uk/VRE-SDM; this project was sponsored by United Kingdom’s JISC VRE program.) This project will also include the development of a Drupal plug-in so that content management environments using Drupal can exchange information about tools of value for the humanities with other Bamboo Work Spaces.

In addition to these Work Space projects, we will carry out work to develop the integrating technology required so that users of each Work Space can access a common list of tools and scholarly web services and share information about these across any Bamboo supported or enabled Work Space:

**Tools and Services Registry. UC Berkeley and University of Wisconsin, Madison.** This project team will include representatives from each of the Work Space projects above, as well developer support dedicated to the registry. The team will develop a common data model, simple interfaces, and a back-end data store for a Tools and Services Registry. In phase 1, we plan to keep the elements of this work as simple as possible so that we can deliver integrating technologies quickly and test their value across the different work space projects and their user communities. The Technical Project Managers of the Work Space area will manage the work.

**Coordination of the Work Space Projects:** Technical Project Managers from the University of Wisconsin, Madison and UC Berkeley will ensure coordinated
management across all Work Space teams and the work on the Tools and Services Registry.

6.3 Corpora Space

As described in section 5.4, during Phase 1 a project team made up of faculty and technical representatives from a number of partner institutions will carry out the design of the Corpora Space scholarly applications that we propose to build in Phase 2. We will organize these planning activities by consulting scholars and technologists within and beyond the Bamboo community to help us write between 3-6 high-level proposals for applications, from which 2 will be selected for detailed development.

This project team be led by the University of Maryland and will include faculty and technologists from the following partners: Tufts University, University of Oxford, Northwestern University, University of Wisconsin at Madison, University of Chicago, and UC Berkeley. A 100% time project manager will support this design project.

At the beginning of the design phase the Corpora Space design team will develop a governance model that clearly identifies specific roles and responsibilities, project communications tools and protocols, and decision-making guidelines, including how to manage risks and course corrections.

As project lead, the University of Maryland will ensure that the team develops and maintains close working relationships with the other Project Bamboo area leads so that the Corpora Space design effort and ultimate implementation will leverage the research and learning among these groups.

Each of the supporting institutions contributing to the Corpora Space design effort will work in close collaboration with one another and will ensure their recommendations and decisions are in alignment with Bamboo Project overall goals. Given the exploratory and collaborative nature of this area of work, the Corpora Space design team will check in frequently with project leadership teams, technical and humanities advisory groups, and the broader humanities community to ensure that their work continues in alignment with project needs and goals.

6.4 Scholarly Services on a Bamboo Services Platform

This area includes three interrelated strands of work: (1) developing scholarly web services by modeling scholarly research methods; (2) developing and deploying the Bamboo services platform; and (3) defining how services can be contributed to the Bamboo platform and releasing a software development tool kit explaining this.

This area of work will require close coordination of software development and integration activities and careful technical coordination across multiple partners throughout phase 1. We will carry out this work as an integrated project made up
of two sub-teams, with overall leadership provided by UC Berkeley. One team will focus on engaging partners and members to develop scholarly web services from existing and new applications and, in the latter part of phase 1, to coordinate projects that will encourage the adoption of the scholarly services. The other team will develop and deploy the Bamboo Services Platform on which these scholarly services will run and will work on the shared services lifecycle and software development toolkit.

**Scholarly Services.** Team Coordinator: *University of Chicago*. **Institutions Contributing Scholarly Services:** *University of Chicago* (PhiloLogic); *Tufts University* (Perseus and other Classical and Ancient Studies); *University of Oxford* (CLARIN); *UC Berkeley* (Prosopography), and *University of Illinois at Urbana-Champaign* (SEASR proxy integration). **Institutions Carrying Out Adoption Projects Using Scholarly Services** (preliminary list): *University of Wisconsin, Madison* (PhiloLogic) and *University of Chicago* (“ARTFLmobile” iPad application). This project team will include technical project management, software developers who will perform service coding, developers to help facilitate and structure adoption projects, and documentation staff.

**Platform and Shared Services Lifecycle.** Team Coordinator: *UC Berkeley*. This team will include staff for technical project management, architecture, platform and framework development, system administration and tools management, test and release management, and software development kit development. The same set of people will carry out these two areas of work. Architecture here will overlap with architecture on Scholarly Services.

Coordination of Scholarly Services on the Services Platform: UC Berkeley will provide architectural leadership across all three strands of work in this area to ensure overall technical coordination and to integrate work with the Work Space and Collections Interoperability teams.

### 6.5 Collections Interoperability

Work here will include the identification and adoption of interoperability standards for the interchange of metadata and content and the application of scholarly services to collections. We will have one team made up of Library and Digital Collections architects and technologists from *Northwestern University, University of Illinois Urbana-Champaign, University of Wisconsin – Madison, Indiana University,* and *University of Chicago*. Project coordination and leadership will be provided by the University of Illinois-Urbana-Champaign and Northwestern University.

### 6.6 Other Coordinating Practices and Groups

**Work Area Coordination:** In addition to the Project Steering Council and Project Executive Group described above, the following groups of project staff will meet and communicate regularly:
• Architects Working Group: It will be critical that the architects and certain
developers are in regular communication, especially given Bamboo’s
emphasis on software integration and collections interoperability. UC Berkeley will establish and facilitate regular communications within this
group.

• Technical Project Managers: It will also be important for technical project
managers to be in regular communication, both within certain work areas
(e.g., Work Spaces) and between work areas. We will convene regular
project manager meetings, reports, and check-ins.

**Coordinating Principles:** Bamboo will adhere to these basic principles:

- The project is a collaborative and integrated software development and
deployment initiative. We will leverage skills and knowledge across institutions. We will invest in developing the collective capacity of our teams and institutions.

- The nature of integration will be appropriate to the project. Some areas of the
project will require tightly integrated project teams, developers, and
management, for example, platform development. Some areas of the project can
be more loosely coupled, for example, the two Work Spaces development teams.

- We will expect institutional teams to have at a minimum a local project leader,
software developer, and technical project management staff.

- We will build an active culture of wiki-based collaboration, and we will
encourage communication and transparency. Project participants will regularly
communicate with each other and will share documentation and reflection on
work practices. This is particularly important in a distributed technology
endeavor.

- Project partners and staff will engage with, learn from, and share information
with humanities scholars, librarians, software developers, IT staff, and
technology consortia outside of Project Bamboo.

### 6.7 Related Projects and Initiatives

Bamboo will work closely with such consortia as the Coalition of Humanities and
Arts Infrastructure and Networks (CHAIN), CenterNet, CLARIN, DARIAH, and the
Consortium of Humanities Centers and Institutes (CHCI). In a number of cases,
leaders of these initiatives are part of the Bamboo Technology Project.

Bamboo will also seek to work closely with other projects in the humanities that are
focusing on the development of scholarly web services and/or the interoperability
of digitized corpora in the classics, medieval, and/or modern studies. For example,
we will build connections to the Stanford University medieval manuscript project,
which includes the participation of representatives from John Hopkins University, Fribourg University, UCLA, and Cambridge University.

In addition, this project will learn from and coordinate planning, standards, and/or technical development with existing and developing community-source projects such as CollectionSpace, OpenAnnotation, ArchiveSpace, ResearchSpace, OLE, OpenCast, and Kuali.

The connection with CollectionSpace (CSpace) will be especially important for the development of scholarly services and the Bamboo Services Platform. We will leverage UC Berkeley’s technical leadership in CSpace and Bamboo as follows:

• Bamboo will build on the methodology refined by CollectionSpace to model and design services and service messages.

• Co-location and co-leadership of the CollectionSpace services team and the Bamboo Services Platform team will leverage technical skills and service design appropriate for adoption by both projects. For example, it is likely that basic services such as Vocabulary, ID, Person, Organization, and Relation already developed by CSpace will be adopted on the Bamboo Services Platform.

• Bamboo’s deployment stack – an OSGi-compliant service container – is a long-term target for CollectionSpace. CSpace deployment on this stack will be facilitated by experience adapting individual CSpace services to the Bamboo environment; and by OSGi skills shared between the two projects at the appropriate stage of CSpace evolution.

• Bamboo and CollectionSpace will share a host of developer tools, from collaborative environments (Confluence wiki, Jira issue-tracker, Subversion code repository) to IDE, test, and build environments (Eclipse, Maven, Ant, Hudson). These tools, most of which are also adopted by Sakai, Kuali, Opencast and other consortial efforts, are becoming a stable IT development ecosystem for intra- and inter-campus development effort, and can thus be provisioned and maintained more efficiently by project consortia and/or institutions.

7. Staffing Plan

This staffing plan follows the structure defined in “Project Organization and Partner Roles” (section 6 above). The detail has been removed from this document, but is included in the full proposal to the Andrew W. Mellon Foundation.
8. Work Plan

We have provided two high-level work plans below; a more detailed plan is provided in Appendix 1. The first view below summarizes major deliverables for each quarter by area of work. The second highlights a smaller number of milestones over the 18-months of the project.

<table>
<thead>
<tr>
<th>Work Spaces</th>
<th>Corpora Space</th>
<th>Scholarly Services</th>
<th>Services Platform</th>
<th>Shared Services Lifecycle</th>
<th>Collections Interoperability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Model &amp; design adaptations to base technology</td>
<td>Services design; articulate platform support requirements</td>
<td>“Hello World” services; core &amp; utility (support) service design</td>
<td>Evolve toolkit and Lifecycle documents; engage expert review</td>
<td>Refine functional scope &amp; standards adoption criteria</td>
</tr>
<tr>
<td>Q2</td>
<td>Implement user interfaces and workflow</td>
<td>Develop functional descriptions for 3-6 proposed Scholarly Applications</td>
<td>Implement first iteration of Scholarly Services</td>
<td>Initial core &amp; utility service implementation</td>
<td>Incorporate lessons &amp; requirements from Scholarly Services &amp; Platform teams</td>
</tr>
<tr>
<td>Q3</td>
<td>Integrate with Services Platform</td>
<td>Select 2-4 candidates for preliminary proposal development</td>
<td>Initial deployment on Platform; functional and technical testing</td>
<td>Initial production BSP running scholarly services; refine core &amp; utility services</td>
<td>Community review and comment; refine toolkit &amp; Lifecycle documents</td>
</tr>
<tr>
<td>Q4</td>
<td>Refine user interface and experience; documentation</td>
<td>Scholar, technical, feasibility review; select 1-2 candidates for refinement</td>
<td>Refine services, establish adoption / engagement teams</td>
<td>Version 1.0 release. Support Scholarly Service &amp; Work Spaces teams</td>
<td>Support adoption; refine toolkit &amp; Lifecycle documents</td>
</tr>
<tr>
<td>Q5</td>
<td>Version 1.0 release</td>
<td>Further refinement; roadmaps for funding proposal</td>
<td>Version 1.0 release; support; select services for Phase II development</td>
<td>Platform refinement; support integrations; distributed-Platform roadmap</td>
<td>Version 0.7 Lifecycle release (final for Phase I). Initial mapping for Phase II evolution</td>
</tr>
<tr>
<td>Q6</td>
<td>Support pilots</td>
<td>Model and initial design for Phase II implementation</td>
<td>Support adoption / engagement; model Phase II services</td>
<td>Support integrations; design distributed-Platform services</td>
<td>Map and draft/pilot Phase II toolkit &amp; Lifecycle</td>
</tr>
</tbody>
</table>
High Level Timeline

Work Spaces

- Q1: Model & design adaptations to base technology
- Q2: Implement & test Work Spaces
- Q3: Version 1.0 release
- Q4: Version 1.0 release
- Q5: Version 1.0 release
- Q6: Version 1.0 release

Corpora Space

- Q1: Develop functional descriptions
- Q2: Roadmaps, models, initial design for Phase II implementation
- Q3: Roadmaps, models, initial design for Phase II implementation
- Q4: Roadmaps, models, initial design for Phase II implementation
- Q5: Roadmaps, models, initial design for Phase II implementation
- Q6: Roadmaps, models, initial design for Phase II implementation

Scholarly Services and the Bamboo Services Platform

- Q1: Implement core & utility services on Bamboo Services Platform
- Q2: Implement first iteration of Scholarly Services
- Q3: Evolve toolkit and Lifecycle documents
- Q4: Version 1.0 release of Bamboo Services Platform
- Q5: Version 0.7 release of Services Lifecycle
- Q6: Version 1.0 release of Scholarly Services

Collections Interoperability

- Q1: Develop Standards Profile
- Q2: Develop compliance tests
- Q3: Support interoperability profile adoption
- Q4: Support interoperability profile adoption
- Q5: Support interoperability profile adoption
- Q6: Support interoperability profile adoption
9. Intellectual Property

The partners in this proposal will adopt the Educational Community License, Version 2.0, of the Open Source Initiative for all software and services developed by this project. This license is based on the Apache 2.0 license and can be found at: http://www.opensource.org/licenses/ecl2.php.

10. Long-Term Sustainability of Results

Our two 18-month projects will enable us to build core software infrastructure, but Project Bamboo is a broader endeavor. The faculty, researchers, librarians, and technologists who carried out the planning process look to the long term. They envision an expanded repertoire of activities and projects that builds upon early successes, and they recognize the critical importance of long-term sustainability. Of course, sustainability requires the attention of technologists to the Bamboo infrastructure, but it includes much more. It calls for the creation of a stable organization that can encourage broad adoption, promote the sharing of services and resources, and ensure that Bamboo continues to address the needs of arts and humanities researchers.

Support from the Mellon Foundation is critical to launching Project Bamboo, and during the 3-year launch phase, a small group of institutions and organizations will be planning and building the specific projects that realize the software infrastructure of Project Bamboo. But the initiative’s long-term independence and sustainability will depend upon the collective efforts of a broader community of institutions.

Work on creating the consortium will occur outside the scope of this proposal, but we are including information about the consortium here in order to clarify the way in which we expect to sustain the project that is within the scope of the proposal. The next section describes the vision for the consortium followed by the scope and high-level timetable for consortial development.

10.1 Project Bamboo Consortium

The Project Bamboo Consortium will be built upon two key concepts that emerged during the planning process. First, there is a real need for long-term, shared, interoperable, and sustainable technology infrastructures to support research and pedagogy in the arts, humanities and interpretive social sciences. These shared and sustainable technologies must enable open discovery and shared access to content, collections, services, tools, interfaces and systems, and they must stay focused on the needs of scholars in the humanities.
The development of a good shared technology infrastructure is not simply a matter of developing good technology. It requires careful thought about how people use technology: how they come to know that it exists, what leads them to give it a try, and how they learn what they can do with it. Technologists need to be able to be aware of which scholars are using technology for what purposes so that they can modify the technology to make it more visible, more attractive, and more useful. The interplay between use and development is complex, and it is the key to sustaining viable shared technology infrastructures for the arts and humanities. Project Bamboo thus aims to create and cultivate open social and scholarly networks to discover, encourage, demonstrate, and promote the use, reuse, remixing, and repurposing of collections, services, and tools.

The Consortium will inevitably be affected by forces beyond its control. At any given time, the Bamboo ecosystem will be shaped by resource availability, disciplinary interest, the financial climate, member participation, agency investment, and community readiness. For this reason, the organization must be flexible enough to react quickly when circumstances shift, and yet sturdy enough to survive major change. To provide both flexibility and sturdiness, the Consortium will adopt five basic principles. It will

• promote innovation through sharing and collaboration,
• respect the diversity of its members,
• strive for shared understanding of research, pedagogical, and technological issues,
• implement solutions for the common good, and
• evolve as needs and interests change.

These principles will help the Consortium develop an upward spiral of exploration, planning, building, and maintenance.

**Exploration.** Participants observe and analyze the adoption and use of services, tools, and programs. They assess community needs, connect activities, provide context for use, and they recommend ways in which Project Bamboo can evolve and adapt to changing needs. Their recommendations may be community focused, and
they may propose outreach programs or pull together community-wide ideas as recommendations for planning.

**Planning.** Planners must take into account not just the needs of scholars but also basic questions of usability, scale, and sustainability. Planners share guidelines and practices, express formal standards and specifications adopted by Project Bamboo, design infrastructure elements and key projects, and align Bamboo with other consortia, societies and organizations. Planners can filter recommendations that emerge from exploration, and they can set priorities for action. Planners will need to operate in a formal manner, documenting and sharing decisions, recommendations, etc.

**Building.** Builders create the components, infrastructure, and tools deemed essential for Project Bamboo. They implement technology infrastructure and shared services, and over time they update Bamboo’s tools, services and infrastructure. Builders act on input from planners, and they are responsible for respecting the principles of sustainability, interoperability, and openness.

**Maintenance.** Bamboo participants manage the shared infrastructure for the Consortium, developing operating and service-level agreements, overseeing compliance, ensuring that the infrastructure is performing as intended, and identifying ways to improve reliability and performance. Through maintenance work, Bamboo enables new projects to be built upon the Bamboo Service Platform, and it helps new projects to capitalize on existing Bamboo software. It thus circles back to the work of exploration, propelling an upward spiral.

By adopting this approach, participants can best align their institutional and individual strengths with aspects of the consortial vision. For example, a humanities center may best contribute via exploration and planning, while an information technology organization may best align its resources with building and maintenance. This approach also provides a checks-and-balances system, ensuring that no one constituency will dominate the growth and direction of the Consortium.

Once established, the Consortium will also include a governance and advisory structure that will oversee Bamboo’s first investment projects and integrate the work of those efforts within the long-term vision of Project Bamboo. The governing and advisory bodies will also assist with the appointment of an Executive Director who will speak for the Consortium and represent Bamboo’s efforts more broadly.

To create its financial model, the Consortium will work with partner and member institutions to balance financial and in-kind investments and adapt them to the scale and character of each institution. The Consortium will work to minimize its operational overhead so that partner and member investments can contribute as directly as possible toward the sustainability of Project Bamboo. The financial model will reflect the diversity of participants as well as the inevitable fiscal challenges that organizations face.
The underlying Bamboo philosophy entails that the Consortium’s work must be freely available for adoption and reuse: this is essential for the creation of sustainable and shared technologies, services, and activities. The Consortium will adopt the Educational Community License, Version 2.0, of the Open Source Initiative for software and services developed by the Project Bamboo Consortium. This license is based on the Apache 2.0 license and can be found at: http://www.opensource.org/licenses/ecl2.php

10.2 Consortial Development

On 18 June 2010 at the Bamboo Planning Workshop 6, we began forming the community-based consortium that will assume responsibility for building and sustaining Project Bamboo over the long term. Participants in Workshop 6 included representatives from:

- Australian National University
- Brown University
- City University of New York
- Dartmouth College
- Harvard University
- Indiana University
- Northwestern University
- Oxford University
- Pennsylvania State University
- San Jose State University
- Tufts University
- University of Alabama
- University of California Berkeley
- University of Chicago
- University of Illinois Urbana-Champaign
- University of Louisiana Lafayette
- University of North Carolina
- University of Wisconsin Madison
- Washington & Lee University
- Willamette University

Together, we defined a set of working groups to advance the development of the consortium that is aligned with the vision and direction of the Bamboo Technology Project.

At present, there are three working groups that include 1) consortial development and membership, 2) outreach and communications, and 3) adoption projects and Bamboo Labs. Before the close of the workshop and as preparatory work for the

---

Footnote 70: By “adoption projects” we refer to simple, focused, and early pilot efforts within the Bamboo Technology Project to test the use of Bamboo software; by “Bamboo Labs” we mean organized activities at local campuses to help facilitate adoption projects and/or to explore other areas of the digital humanities in the context of the Bamboo Consortium.
forthcoming working groups, we developed an updated model for institutional, commercial, and small group/independent scholar participation in the Project Bamboo Consortium, outlined an initial set of targeted outreach materials, set goals for transitioning the website from the planning project to the consortium, and framed the concept of Bamboo Labs and its relation to adoption projects as described in this proposal. These efforts will be integrated into the working groups as they are established.

In July 2010, the working groups will be formally constituted with each led by members of the current workshop community. Volunteering to lead the working group efforts include representatives from the City University of New York, San Jose State University, University of Alabama, University of Chicago, University of Louisiana Lafayette, and the University of Wisconsin Madison. We view this level of engagement as a positive force for consortial development as only two of the six volunteers are Bamboo Technology Proposal partners and one, the University of Alabama, is new to Project Bamboo as of the sixth workshop. Of the six institutions, the University of Chicago will coordinate the effort in alignment and close coordination with the leadership of the Bamboo Technology Proposal effort led by the University of California, Berkeley.

Funding of the initial effort would come from in-kind contributions from the Project Bamboo membership. By summer 2011, we intend to implement a fully-defined multi-year participatory model based on membership backed by fees and in-kind contributions from institutions, organizations, industry, projects, and others. The financial model was discussed both at Workshops 5 and 6, and from the positive feedback we have received to date, we believe we can implement a model that will be embraced by the Project Bamboo community and that will be balanced in terms of benefit with contribution.

Turning to timelines, the initial work of the consortial development and membership working group is slated to be complete by January 2011. Work during this period will include finalizing the membership and participation approach that will be based on a multi-year commitment and tiered contribution model, defining the consortial leadership approach that will function in harmony with the Bamboo Technology Proposal, and articulating the characteristics of the consortium as an operational and coordinating entity for Project Bamboo. During this period, we intend to reach out to all 114 past and present participating institutions and organizations to invite them to re-engage, if they haven’t done so already, with Project Bamboo. We will also extend an open invitation to others who wish to take part. Once the initial working group activity is complete, our attention will shift toward the establishment of a non-profit foundation that will be built upon the consortial development work of late 2010. The goal is to have the non-profit consortium formed by late summer 2011.

In terms of outreach and communications, our initial efforts (including the transition of the website) shall be largely complete by October 2010 to coincide with the proposed launch of the work described in this proposal. At that time, we will
formalize the outreach and communication working group into a communication team that will be focused on regular outreach to members of the Project Bamboo community, define community programs and events, and create channels of communication with those not engaged with Project Bamboo. Within this domain, we will explore the existing informal liaison relationships with organizations such as CLARIN, DARIAH, NITLE, Internet2, CenterNet, ADHO, and others to establish a firm consortial-level foundation for collaboration. It is essential to Project Bamboo’s success to fit into the efforts of these other activities.

The adoption projects and Bamboo Labs work will begin in October 2010 and continue through the completion of the Bamboo Technology Proposal, with a major milestone being April 2011 when adoption projects begin to ramp up within the technology project. The term “Labs” has resonated with many participants in the Bamboo planning project. It gives local campuses a way to call out, organize, and label a set of activities where scholars, librarians, and/or technologists can explore, demonstrate, and provide training about technologies relevant to the humanities. The efforts of this working group dovetail both with the other two working groups, provides an early-adoption opportunity for participants, and acts a means for constructive feedback during technology development that would otherwise be limited to the subset of partners implementing the Bamboo Technology Proposal. Bamboo Labs also provide outreach opportunities to share tangible benefits and outcomes of technology development with a wider audience. Therefore, we envision a timeline that fully overlaps with the Bamboo Technology Project to ensure the design of the Bamboo Labs will mesh with how technology is shaped during project development. We are looking toward partners such as the University of Wisconsin Madison to help lead this effort to make sure it is synchronized to the ongoing work within the Bamboo Technology Project.

Through these initial efforts, we intend to bring the Project Bamboo Consortium into alignment with the Bamboo Technology Proposal 12-15 months after technology development work begins.

10.3 Humanities Faculty Committee and Technical Advisory Committee for Bamboo Technology Project

We have included preliminary and partial list of possible members for the Humanities Faculty and Technical Advisory Committees below. Please note that the Bamboo Partners will solicit recommendations for additional candidates to consider for these committees, and then will ask the Project’s Steering Council to select the set of members to be invited to serve on the committees. We expect that these committees will start in Fall 2010.

Humanities Faculty Committee (preliminary membership)

Tony Cascardi (Berkeley )
Donald Mastronarde (Berkeley )
11. Reporting and Evaluation

Reporting

We will submit two reports, prepared by Principal Investigator Janet Broughton and Executive Director David Greenbaum, to the Mellon Foundation. We will provide an interim report nine months into the project, by June 30, 2011, and a final report two months after project completion, by May 30, 2012. Reports will include documentation of all project activities, deliverables, and expenditures; and information regarding any variances between this information and the goals, deliverables, and budget information contained in this proposal. Also included will be descriptions of, and links to, downloads and demonstration versions of the software, and a summary discussion of opportunities and challenges facing the project.
Evaluation

Evaluation of this project will include external review of the architecture, service design, and code for major technical deliverables in the project. Project Bamboo’s Technical Advisory Committee will recommend individuals and/or firms to carry out this external review; the Project Executive Group will select the reviewers.71

In addition to this external review, we have built evaluation activity into the ongoing work, as summarized below. These activities are drawn from the summary of phase 1 project deliverables in section 4.5, above; and from the detailed work plan of section 13, below.

<table>
<thead>
<tr>
<th>Area of Work</th>
<th>Evaluation Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Spaces</td>
<td>1. User experience feedback (quarters 2, 3, and 4)</td>
</tr>
<tr>
<td></td>
<td>2. Work Spaces Phase 1 Assessment (quarter 6)</td>
</tr>
<tr>
<td>Corpora Space Design</td>
<td>1. Review and critique of Corpora Space Application Design white papers (quarter 3)</td>
</tr>
<tr>
<td></td>
<td>2. Evaluation and selection of white papers for further development as preliminary proposals (quarter 4)</td>
</tr>
<tr>
<td></td>
<td>3. Review and critique of preliminary proposals (quarter 5)</td>
</tr>
<tr>
<td>Scholarly Services</td>
<td>1. Iterative cycles of design and coding, with review by scholar stakeholders (quarters 2, 3, and 4)</td>
</tr>
<tr>
<td></td>
<td>2. Assessment of Phase 1 Scholarly Services development (quarter 6)</td>
</tr>
<tr>
<td></td>
<td>3. Assessment of applied use of Phase 1 Scholarly Services in/by adoption projects (quarter 6)</td>
</tr>
<tr>
<td>Services Platform</td>
<td>1. Iterative cycles of development, deployment, uptake by Scholarly Services &amp; Work Spaces teams, and review with change requests (quarters 2, 3, and 4)</td>
</tr>
<tr>
<td></td>
<td>2. Analyze platform use and adoption by Scholarly Service developers and adoption projects (quarter 5)</td>
</tr>
<tr>
<td></td>
<td>3. Assessment of Phase 1 Services Platform (quarter 6)</td>
</tr>
<tr>
<td>Services Lifecycle</td>
<td>1. Iterative cycles of drafting, publication, review, and revision of lifecycle documents (quarters 1, 2, 3, and 4)</td>
</tr>
<tr>
<td></td>
<td>2. Assessment of Shared Services Lifecycle (quarters 5 and 6)</td>
</tr>
<tr>
<td>Collections Interoperability</td>
<td>1. Iterative cycles of drafting, publication, review, and revision of Collections Interoperability standards profile (quarters 1, 2, and 3)</td>
</tr>
<tr>
<td></td>
<td>2. Assessment of Collections Interoperability profile adoption and effectiveness (quarters 4, 5, and 6)</td>
</tr>
<tr>
<td></td>
<td>3. Assessment of Collections Interoperability standards profile and service development (quarter 6)</td>
</tr>
</tbody>
</table>

71 This proposal’s Project Executive Group, as well as Project Bamboo’s Technical Advisory Committee, are described in Section 6.1, above.
12. **Budget Narrative**

The budget narrative has been removed from this document, but is included in the full proposal to the Andrew W. Mellon Foundation.
13. Appendix 1: Detailed Work Plan

Work to be Completed in the Planning Phase

We have built a work plan on the accomplishments of the Project Bamboo Planning Phase, supported by a major grant from the Andrew W. Mellon Foundation. Planning and preparation will continue through the completion of the Planning Phase, in September 2010. Planning-phase precursors to the work plan are listed as background and context, below:

- **Work Spaces**: Develop specification and technology-adaptation roadmaps
- **Corpora Space**: Identify scholars for Phase I design activities
- **Scholarly Services**: Model Scholarly Services to be deployed in Phase I
- **Services Platform**: Articulate technical requirements
- **Services Platform**: Assemble, configure, and deploy Platform infrastructure technology (v0.1; cf. Technical Addendum, Appendix 1)
- **Services Platform**: Identify and model candidate Core and Utility services (cf. Technical Addendum, Appendix 1)
- **Services Lifecycle**: Draft and publish scope and elements of a proposed Bamboo Shared Services Lifecycle
- **Services Lifecycle**: Select, deploy, and configure collaborative development tools (e.g., code repository, issue tracker)
- **Collections Interoperability**: Draft and publish scope
- **Collections Interoperability**: Identify candidate standards, protocols, and specifications

Legend:

In order to fit information in a compressed table, the following acronyms are used in the Work Plan:

- **BSP** = “Bamboo Services Platform” (a.k.a. “Platform”)
- **BSSL** = "Bamboo Shared Services Lifecycle" (a.k.a. "Lifecycle")
- **Cl** = “Collections Interoperability”
- **EoQ** = “End of Quarter”
- **Phase I** = first, 18-month phase of implementation described in this Project Bamboo Technology Proposal, Oct 2010 – Mar 2012
- **Phase II** = a second 18-month phase of implementation, continuing Phase I effort, Apr 2012 – Sep 2013
- **SDK** = “Software Development toolKit” – a set of tools to facilitate coding, deployment
- **UI** = “User Interface,” i.e., the screens through which a user accesses the functionality of an application, tool, etc.
- **v#.##** = version numbers; v1.0 signifies a full, production release, while lesser numbers signify fractional progress toward production

<table>
<thead>
<tr>
<th></th>
<th>Oct-Dec 2010 (Q1)</th>
<th>Jan-Mar 2011 (Q2)</th>
<th>Apr-Jun 2011 (Q3)</th>
<th>Jul-Sep 2011 (Q4)</th>
<th>Oct-Dec 2011 (Q5)</th>
<th>Jan-Mar 2012 (Q6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Scholarly Work Spaces</strong></td>
<td><strong>Formalize Work Spaces and Tool &amp; Service Registry development roadmaps. Model and design adaptations to base technology (e.g., to UI, to research task workflows, to consumption of Platform services). Develop Work Spaces project requirements for scholar/student pilots.</strong>&lt;br&gt;EoQ: Adaptation roadmap, models and design. Requirements for pilots.</td>
<td><strong>Implement UI and research task workflows. Engage scholar community for user experience feedback; identify UI and workflow refinements to be implemented in v1.0 Work Spaces. Model and design Tool &amp; Service Registry data store and services. Solicit proposals for Work Spaces pilots.</strong>&lt;br&gt;EoQ: UI and workflow functionality complete to v0.6. Tool &amp; Service Registry design complete.</td>
<td><strong>Implement service integration with Platform; refine UI &amp; research task workflow functionality based on user experience feedback to-date. Engage scholar community for user experience feedback; identify further refinements to be implemented in v1.0 Work Spaces. Implement and test Tool &amp; Service Registry functionality. Review and select Work Spaces pilots, support development of pilot roadmaps.</strong>&lt;br&gt;EoQ: Service integration functionality complete to v0.6; UI and workflow functionality complete to v0.9. Tool &amp; Service Registry implemented to v0.8. Pilots selected.</td>
<td><strong>Complete Phase I WorkSpaces and Tool &amp; Service Registry functionality to release-candidate stage. Engage scholar community for final round of feedback; identify final refinements to be implemented in v1.0. Support evolving pilot roadmaps; document; establish pilot teams.</strong>&lt;br&gt;EoQ: Work Spaces, Tool &amp; Service Registry functionality complete to v1.0-rc.</td>
<td><strong>Complete Phase I WorkSpaces and Tool &amp; Service Registry functionality for v1.0 release. Refine documentation. Support pilots, monitoring teams. Continue to solicit feedback from scholarly community, initiate mapping of Work Spaces evolution roadmap.</strong>&lt;br&gt;EoQ: Release v1.0 Scholarly Work Spaces and Tool &amp; Service Registry.</td>
<td><strong>Support pilots; assess applied use of Bamboo standards, services, and processes. Articulate roadmap for evolution of Scholarly Work Spaces and Tool &amp; Service Registry technology in Phase II.</strong>&lt;br&gt;EoQ: Assessment of applied use of Bamboo standards, services, and processes in this area. Roadmap for evolution of Scholarly Work Spaces technology in Phase II.</td>
</tr>
<tr>
<td>2. Corpora Space</td>
<td>--</td>
<td>Engage scholars in disciplinary clusters to identify broadly useful corpora-centered functionality applicable to humanities scholarship, building on prior efforts. Recruit 3-6 small groups to compose white papers with functional descriptions of 3-6 potential corpora-centered Applications. EoQ: 3-6 teams recruited; broad scope for each white paper defined.</td>
<td>Develop and publish white papers for comment, critique, and technical review. Revise white papers to reflect feedback. EoQ: 3-6 white papers reviewed by scholars and technologists; revised by writing teams.</td>
<td>Leadership groups to select 2 white papers for further development into preliminary proposals for corpora-centered Applications. Publish proposals for comment, critique, and technical review. Revise &amp; refine proposals based on initial round of critique and technical review. Further define and refine as necessary. EoQ: 2 well-defined feasible, technically sound, corpora-centered Applications proposals of prioritized value to humanist scholarship.</td>
<td>Further define and refine proposals (note: descriptions of work to be included in Dec 2011 funding proposal). Publish evolved proposals for comment, critique, and technical review. EoQ: 2 refined proposals for corpora-centered Applications; descriptions included in December funding proposal.</td>
<td>Develop roadmap, model, and initial design for 2 corpora-centered Applications to be implemented in Phase II of Bamboo implementation. EoQ: Roadmap, model, and design for 2 corpora-centered Applications for which development funding will be requested.</td>
</tr>
<tr>
<td>3a. Scholarly Services on Bamboo Services Platform:</td>
<td>3b. Scholarly Services on Bamboo Services Platform (BSP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Platform</strong></td>
<td><strong>Platform</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scholarly Services</td>
<td>Scholarly Services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Services</strong></td>
<td><strong>Services</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design Scholarly Services to be deployed on v1.0 BSP; design Scholarly Services; articulate core/utility service support needs. Develop requirements for adoption/engagement projects.</td>
<td>Implement and deploy “Hello World” services on BSP 0.1 stack. Pilot service development &amp; deployment processes: support Scholarly Services partners in deployments of and calls to services on the BSP. Initiate core and utility service design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: Service model and design near-complete for Scholarly Services slated for BSP 1.0 deployment. Requirements for adoption/engagement projects.</td>
<td>EoQ: BSP v0.1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Develop Scholarly Services for deployment on BSP:</strong> test and refine service interfaces and Collections interoperability compliance (as necessary); draft service and contract documentation. Solicit proposals for adoption/engagement projects.</td>
<td>Refine BSP infrastructure, services, and dev/deployment processes. Respond to issues raised by Scholarly Services partners regarding technologies, tuning, and configuration of BSP. Initiate core and utility service implementation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: Service designs complete; service implementations in-progress (~v0.5 on average).</td>
<td>EoQ: BSP v0.4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deploy initial set of Scholarly Services on the BSP:</strong> test and refine; confirm Scholarly Services meet functional requirements and core/utility services deliver necessary support; revise service and contract documentation. Review and select adoption/engagement projects, support development roadmaps.</td>
<td>Deploy BSP 0.7 at a single location shared across partner institutions. Implement, test, and deploy core and utility services. Document BSP and support Scholarly Service deployment and consumption.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: Services deployed (v0.6), refined (v0.7), tested (v0.8). Service documentation up-to-date for v0.8. Adoption/engagement projects selected.</td>
<td>EoQ: BSP v0.7, initial core &amp; utility service implementations.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Promote Scholarly Services to adoption/engagement project teams; establish co-development teams (Scholarly Service developers &amp; service consumers); assist adoption/engagement projects as necessary; refine services; monitor and track issues; identify final revisions to v1.0 services.</strong></td>
<td>Deploy BSP 1.0 at a single location shared across partner institutions; refine BSP documentation. Operate BSP 1.0 and continue to support Scholarly Service deployment and consumption.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: v0.9 services; adoption/engagement project work plans articulated; final revisions mapped for next (v1.0) release.</td>
<td>EoQ: BSP v1.0.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Review and make initial selections of candidate services for Phase II BSP deployment. Partner with adoption/engagement teams to effect initial integrations, monitoring teams. Implement and deploy v1.0 services.</strong></td>
<td>Support and analyze platform use and adoption of BSP by Scholarly Service developers and adoption projects. Assist early adopters with technical issues. Initiate exploration of issues to resolve re: distributed BSP deployment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: In-progress adoption/engagement (variable progress). v1.0 service releases.</td>
<td>EoQ: Final Phase 1 releases of Scholarly Services. Assessment of applied use of Bamboo standards, services, and processes in this area. Evaluation of candidate Phase II services against BSP capabilities; Phase II service models.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Finalize initial scholarly services (v1.0.x); finalize documentation; revise and release Scholarly Services as needed (v1.1). Support adoption/engagement projects; assess applied use of Bamboo standards, services, and processes; finalize selection and model Phase II Scholarly Services.</strong></td>
<td>Support initial Scholarly Service integrations. Initiate design of services to support distributed BSP deployment. Assess BSP 1.0 adoption and use; make adjustments to BSP.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EoQ: Initial service designs to support distributed BSP deployment. Formal assessment of BSP 1.0.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c. Scholarly Services on Bamboo Services Platform: Collections Lifecycle</td>
<td>Revise initial draft of Bamboo Shared Service Lifecycle (BSSL) documentation; publish second draft; refine candidate Software Developer Kit (SDK); engage experts to review lifecycle vision; identify and draft service contract requirements. Iterate SDK as necessary. EoQ: Lifecycle docs v0.2 published for review. Fully functional SDK.</td>
<td>Incorporate service lifecycle requirements from Scholarly Services, Collections Interoperability, and BSD; revise lifecycle documentation; publish third draft for community review. Stabilize SDK for Phase I use. EoQ: Lifecycle docs v0.3 out for review. Stable SDK.</td>
<td>Revise Lifecycle docs to v0.4; shepherd through community review and comment; cycle docs to v0.5; promote adoption of v0.5 for Bamboo service development, deployment, and maintenance. Refine SDK as necessary, and document. EoQ: Lifecycle docs v0.5. Refined SDK, SDK documentation.</td>
<td>Revise Lifecycle docs to v0.6. Support adoption of BSSL 0.6 across adoption projects during adoption planning phase, establish co-development teams; monitor lifecycle adoption and track issues; suggest revisions to BSSL 0.6. Support adoption ofSDK, monitor issues, enrich documentation. EoQ: Lifecycle docs v0.6 &amp; suggested revisions. Enriched SDK documentation.</td>
<td>Revise Lifecycle v0.6 &amp; SDK adoption and evaluate effectiveness; revise, and release Lifecycle v0.7 docs. Support uptake and monitor issues with Lifecycle v0.7, track changes for Phase II revisions. Update SDK as necessary. EoQ: Lifecycle docs v0.7 (final for Phase I). Updated SDK and documentation.</td>
<td>Continue to support uptake and monitor issues with Lifecycle docs and SDK during adoption work. Track and fully document proposed changes for Phase II Lifecycle and SDK additions and revisions. EoQ: Roadmap for evolution of Lifecycle and SDK during Phase II.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4. Collections Interoperability (CI)</td>
<td>Review initial functional scope against user stories &amp; initial BSD and Scholarly Services roadmap; refine functional scope; develop criteria and protocols for standards adoption; propose testing protocols for service contracts and integrations. Develop requirements for adoption projects. EoQ: Complete identification &amp; selection of key standards; gap analysis.</td>
<td>Initiate development of tests for CI contracts and integrations against stub or initial Scholarly Services; create draft CI profile of standards and specifications for initially scoped degree of collections interoperability; publish draft for review. Solicit proposals for adoption projects. EoQ: CI profile v0.5. Initial compliance levels definitions.</td>
<td>Further testing and refinement of CI; develop content store compliance tests and conformance requirements for the CI profile; revise CI profile and publish for review. Review and select adoption projects, support development of adoption roadmaps. EoQ: CI profile v0.8; adoption projects selected; initial demonstrations of what levels of compliance mean.</td>
<td>Support integration of CI 0.8 in adoption project roadmaps, establish co-development teams (CI partners &amp; adoption project teams); monitor CI profile adoption and track issues; evaluate compliance and conformance requirements against declared adoptions; iterate to CI v0.9 and map revisions to CI v1.0. EoQ: CI profile v0.9; mapped revisions for v1.0 (release); begin work on bi-directional content interoperability.</td>
<td>Assess CI profile adoption and evaluate its effectiveness as a Bamboo interoperability profile. Iterate to CI profile v1.0 compliance and conformance specification; support adoption of CI 1.0 in adoption projects, monitoring teams.</td>
<td>Continue to support uptake and monitor issues with CI in adoption project work. Track and fully document proposed changes for Phase II CI additions and revisions. Support adoption projects; assess applied use of Bamboo standards, services, and processes. EoQ: Assessment of applied use of Bamboo standards, services, and processes in this area. Roadmap for CI Phase II.</td>
</tr>
</tbody>
</table>