

**NEW FRONTIERS IN THE DIGITAL LIBRARY: SOCIAL AND
ECOLOGICAL DIVERSITY OF THE AMERICAN WEST**

A PROPOSAL RESPECTFULLY SUBMITTED TO THE WILLIAM AND
FLORA HEWLETT FOUNDATION

BY

THE CALIFORNIA DIGITAL LIBRARY

DANIEL GREENSTEIN, PRINCIPAL INVESTIGATOR
AUGUST 15, 2003

1. Abstract	3
2. Background	3
3. Problems	4
4. Outcomes	7
5. Activities and outputs	11
6. Inputs.....	20
7. Evaluation.....	22

1. Abstract

During the past ten years, universities, libraries, museums, and archives have made a large and growing component of society's cultural, scholarly, and scientific record available openly online to anyone who has access to the Internet. Yet a number of very substantial impediments threaten to slow our further progress. Work described in this proposal seeks to address three of them, notably, the absence of robust mechanisms that enable users to:

- ❑ build their own collections from existing online materials — that is, to select and configure online information to suit their own needs;
- ❑ integrate online materials with the instructional technologies they prefer to use;
- ❑ contribute the digital products of their own work (such as online research publications, teaching materials, research data, personal text or image collections) to digital libraries where they can be managed persistently and made accessible to others.

These problems will be addressed practically. Through extensive consultation with university researchers and faculty, with educators K-20, and with public librarians, this project will design, assemble, and evaluate the use of a large virtual collection of digital materials bearing on the social and ecological diversity of the American West. The collection will be developed and presented with a range of tools that support its extensive re-configuration, its integration with online learning environments, and its continued growth through the addition of relevant research and teaching materials that are produced in the course of its use.

2. Background

The University of California's eleven university libraries provide access to the world's scholarship and knowledge for the University of California and the communities it serves.ⁱ The libraries have worked in close collaboration for over a quarter century. In doing so, they have built the greatest research library and the largest academic digital library available anywhere.ⁱⁱ The university libraries have also adopted an innovative service model that allows them to leverage the library investments that are made by the ten UC campuses, the campus libraries acting together (e.g. on consortial content acquisitions and service developments), and centrally (e.g. on central facilities such as the California Digital Library and on high-density storage), to their collective and individual advantage.

Leveraging this service model, the libraries have taken an internationally recognized leadership role in dealing with the economic, cultural, organizational, and technical challenges that are fundamentally altering research, learning, and scholarly communication.ⁱⁱⁱ Particular attention has been paid to the libraries':

- ❑ development of print and digital collections that the libraries govern collectively;^{iv}
- ❑ proactive investment through the CDL in a digital preservation program;^v
- ❑ investment through the CDL in mechanisms that support faculty, the UC Press, and others who wish to explore ways to disseminate scholarly publications without imposing debilitating access costs on libraries and/or their users;^{vi}
- ❑ technical leadership and innovation.^{vii}

In 2002/03, the UC libraries launched three new initiatives that will help them confront current challenges while shaping their future look, feel, and significance. They will also contribute extensively to the work that is proposed here. Through the CDL, the UC libraries have:

- ❑ launched a public Web site that assembles in one place nearly 500 open-access online collections produced by UC's libraries, museums, archives, academic departments, and research centers;^{viii}

- ❑ started to investigate how to enable campus libraries and faculty to configure online collections so they meet specific institutional and individual needs;
- ❑ in partnership with the Interactive University Project at Berkeley and a consortium at UCLA involving the library, the Office of Instructional Development, and the Office of Instructional technology, begun to explore how to integrate digital library information with the instructional technologies that are used in support of online instruction and learning.^{ix}

The work described in this proposal is an outgrowth of the UC libraries' long history of successful collaboration, innovative, and service to the university and the people of California. It leverages the university's substantial investments in research and teaching, and in its libraries. It builds directly upon and promises to accelerate the new directions that are described briefly above in ways that will ensure they have an impact well beyond the University of California.

3. Problems

In an extensively networked digital age, access to information does not require physical proximity to it. The democratic impulses are as compelling as they are profound. On the back of the Internet, barriers erode and permit broader civic access to the world's leading cultural memory organizations (such as libraries, museums, archives, and historical societies) and to the depth of knowledge that is available at institutes of higher learning. The scholarly process is itself transformed. By exploiting network technologies, researchers worldwide acquire access to primary evidence (as represented in previously less accessible archives and libraries, and in the research data that may be locked away on individual and departmental computers), irrespective of where it is stored and by whom it is managed. They are able in turn to communicate the results of their research to worldwide audiences and participate actively in instruction instantaneously via the Internet.

Since the introduction of the public Internet and the World Wide Web, libraries, scholars, and teachers have contributed immeasurably to achieving this vision. An already large and growing component of society's cultural, scholarly, and scientific record is available openly online to anyone with access to the Internet. Yet a number of very substantial impediments threaten our further progress. Work described in this proposal seeks to address three of these problems, each of which is described in greater detail below.

3.1. Building personalized collections. Readily accessible, easy-to-use, and low-cost mechanisms that enable scholars, teachers, and librarians to selectively assemble digital collections that meet their very specific needs from the vast array of online material are absent in today's Internet environment. This undermines our ability to exploit the Internet's democratic promise.

Libraries and other organizations that guard our cultural memory have developed a countless number of digital collections based on materials selected from their local holdings and digitally reformatted so they are accessible online. Typically, the collections are organized to address specific cultural, historical, or other themes, or to highlight the unique holdings of the host institution. Such collections are at once wonderful, idiosyncratic, and rich; each of them a marvelous cabinet of curiosities. Yet in their current form, they bind the digital objects they contain (the digital images, encoded texts, sound clips, etc.), to a single interpretation, exhibition, or display. To realize their full potential, the objects contained within these collections need to be made available in a way that permits their use in different contexts, surrounded by other materials, and in support of different interpretations. After all, a digital image surrogate for Frederick Remington's *Victory Dance* that may appear in an online collection based on the artist's work does not know whether it belongs to that collection, or to another that illuminates

changing representations of aboriginal cultures, or still another that bears on the significance of watercolors in 19th century American genre painting.^x

Recognizing the immense value that is potentially achieved from the flexible aggregation and re-presentation of their collective holdings, academic and research libraries have pursued the means of integrating access to their distributed digital collections. The National Science Foundation's National Science Digital Library, and AmericanSouth.org are two good examples of this pursuit.^{xi} Using very different strategies, but a common range of technologies, they integrate online access to information about collections that are distributed across participating libraries, archives, museums, and other organizations. The Scholar's Portal, a project of the Association of Research Libraries, and the Distributed National Electronic Resource that is being developed by the Joint Information Systems Committee of the United Kingdom's Higher Education Funding Councils, are similarly pursuing the means of integrating access to distributed, deeply heterogeneous digital collections.^{xii} In part, owing to the success of at least some of these initiatives, libraries and educators are beginning to ask questions that will define the next generation of work in this area, notably about how to achieve:

- ❑ a greater degree of selectivity than is currently possible: "I want to assemble from these online collections only those items — bibliographic references, online journal articles, digital image surrogates — that bear directly on the life and times of Frederick Remington";
- ❑ a deeper level of integration: "I want to include materials drawn from my own list of favorite Web sites, my own digitized image surrogates for Remington's works, and two articles that that I have published on my Web site";
- ❑ a higher degree of configurability: "I want to configure how results are displayed when I query my virtual collection, and to be alerted automatically by email when new items are added to the underlying information resources."

Work described in this proposal will address these very challenges. After extensive consultation with researchers, with educators at all levels, and with academic and public librarians about how they want to build personalized collections, it will investigate technologies that promise to support configurability, and apply selected technologies in the construction of a large virtual collection bearing on the social and ecological diversity of the American West. The collection will be assembled from existing materials that are made openly accessible online by some of the world's leading digital libraries. To date, we have been offered access to some of the world's largest and most important digital collections at the University of California (where many of the 500 online collections included in our Web site contain relevant materials), at the university libraries of Indiana, Michigan, and Virginia, at the Library of Congress, and at the Colorado Digitization Program. The collection on the American West will be presented along with the tools we develop to support its further configurability by institutions, such as libraries, schools, and academic departments, and even individual users.

3.2 Integrating online collections with instructional learning technologies. Digital library collections are not readily integrated in the online environments that are developed in support of teaching and learning. These collections typically allow users to search and display digital objects such as digital images, documents, sound recordings and film clips. They lack the interpretive content that knits the objects together to create learning materials that can be brought meaningfully to bear in the classroom on particular topics, courses, or lessons. The waste is enormous.^{xiii} Even at those institutions that have constructed the most sophisticated and pervasive online learning environments, the wealth of information that is available through the library is underutilized because it is not made accessible in a way that supports instruction.

The Massachusetts Institute of Technology (MIT) — an acknowledged leader in the development and use of online learning — is exemplary in this regard. MIT has made a very substantial number of its courses openly available online.^{xiv} Yet a brief review of the materials demonstrates the current incapacity that afflicts us all. Where materials direct students to essential readings, they do so with static texts. Where bibliographic citations are concerned, the references do not link automatically to the local OPAC (where a student could discover instantaneously the item's location and availability), or to the local bookstore or to Amazon.com, where the item may be purchased. Where electronic journal articles or items held digitally by the library in its e-reserves are concerned, the inefficiency and waste is even more maddening. Despite the fact that the requisite readings are, in theory, only a click away, they remain inaccessible from the syllabus.

The fault is not MIT's. The technical environments in which Web-based course materials are produced don't integrate with those through which libraries disclose resources online. And if local holdings remain at a distance, those available through the numerous and proliferating open access collections produced by libraries, archives, museums, and academic centers worldwide are that much further removed. MIT course materials are littered with URLs that refer their users to third-party online collections, but these are not and cannot be assembled in a way that selectively references the relevant items that these collections contain. The problem is not only pressing, it has also become something of a *cause celebre*. For example, the Coalition for Networked Information and the IMS Global Learning Consortium have joined together to raise awareness in the academic, library, and instructional technology communities as a means of stimulating investment in viable solutions for bridging a yawning technology gap.^{xv}

Educators need to draw selectively from information that is available in a countless number of, differently formatted online resources. They need to include items they select from these resources in the course and learning materials they are developing, and to surround them with an interpretive web that turns information into knowledge and knowledge into pedagogical practice. And they must be able to do this in a way that permits them to use whatever software they have to hand locally, whether it is WebCT, Blackboard, PowerPoint, or something else. Work on this project leverages two partnerships that the CDL has established: one with UC Berkeley's Interactive University Project and one with a consortium at UCLA that involves the library, the Office of Instructional Development, and the Office of Instructional technology. Working with UCLA we are exploring the means of presenting digital library resources so they integrate directly with learning software that conform to a digital object standard known as SCORM that is gaining ascendance in the instructional technology community. The American West collection will provide an important testbed for the furtherance of this work.

Since standards convergence is a slow process whose ultimate course is difficult to predict, our work with Berkeley's Interactive University Project focuses on an interim solution. It will develop to an operational specification a tool that exists already as a prototype. The Scholar's Box is being developed by the Interactive University Project at UC Berkeley in partnership with the CDL and others. It enables users to select individual information objects in a small but growing number of different formats, include them in a personal workspace where they can be organized, arranged, and annotated, and then output in formats recognizable to a variety of commonly used learning management and desktop software. The Scholars' Box is being developed as an open source tool that can be integrated as a feature, or toolset, in any virtual collection. Our aim is to make it openly available with our virtual collection on the American West to assist users in the collection's educational and instructional exploitation. Until such time as the digital library and instructional technology communities converge around a common set of interoperable standards (which is unlikely in the short term), the Scholar's Box promises to provide an important bridge between them.

3.3. Capturing and encouraging re-use of the digital products of research and teaching.

Presently, we have limited ability to capture, persistently manage, and encourage educational re-use of the online research, learning, and course materials that scholars and teachers produce as a natural by-product of their work. Accordingly, we are deprived of some of the very best that scholarship and pedagogical practice have to offer. We have access to the published record (the journal articles monographs and course texts), but this represents only the tip of a vast iceberg of knowledge that can and should be made more readily available to the benefit of all.

In the United States, the problem is pressing in schools and community colleges, particularly those that provide education to underserved communities. It is acute at educational institutions in the developing world. Rapid advances in telecommunications furnish network connections that promise seamless access to the scholarly knowledge that exists at leading universities.^{xvii} Yet the availability of educational content does not keep pace with the network's extension; its provision has become a cause for public education networks and non-governmental organizations.^{xviii} In the meantime, where scholarly knowledge and deep cultural resources are made available publicly via the Internet, it is heavily exploited. MIT makes its courses openly available via the Internet through its OpenCourseWare initiative, which experiences very heavy national and international use. The Library of Congress's American Memory collections — with some 8 million document and other images available online — is also massively exploited. At UC, the eScholarship Repository, where faculty publish openly to the Internet, attracts some 1,500 full paper downloads per week, 98 percent of them from outside UC.

Work described in this proposal will approach the problem from the perspective of scholars and educators who are producers and consumers of online research and teaching materials, and from the perspective of the information manager who is tasked with persistently managing those materials. In close consultation with faculty, our work will explore and contrast three approaches that are currently used to capture the online information that results from research and teaching, and integrate the best of these with the American West collection. Accordingly, scholars, teachers, and curators who create online materials based on their work with the collection will be able to submit some of their materials to it, thereby enriching the collection and enhancing its usefulness.

4. Outcomes

4.1. Intermediate outcomes. The project will assemble from some of the world's leading academic and library organizations (at the University of California, the libraries of the University of Indiana, Michigan, and Virginia, from the Library of Congress, and from the Colorado Digitization Program) an openly accessible online collection that bears on the social and ecological diversity of the American West. The collection will be made available along with the tools we develop as we seek the means of helping users to:

- ❑ configure the collection in a manner that meets their own curatorial, research, or teaching requirements, integrating additional, third-party resources where appropriate;
- ❑ selectively integrate items drawn from the collection on the American West into the learning management systems and other software they use to create, manage, and disseminate online course and learning materials;
- ❑ contribute to the American West collections some of the research papers, teaching materials, or other digitized content they create based on their work with it.

4.2. Ultimate outcomes. The project will:

- ❑ build a public educational and cultural resource demonstrating its value through use in K-20 instruction and in public libraries;

- ❑ build for scale and for generalizability to ensure that the techniques and strategies we apply can be applied by others working in different contexts, in all sectors, and at all educational levels. Thus, we will indirectly support the assembly and exploitation of digital collections in other thematic areas, at all educational levels, and in cultural and educational organizations (e.g. in public as well as academic libraries, museums, etc.);
- ❑ put in place mechanisms that will help the University of California sustainably disclose some of the vastness of its online resources openly and to the world. These resources take on many forms (such as scholarly publications, course and other learning materials, research data, digital library, and online museum collections) and exist in a variety of places (including within the university's libraries, museums, archives, research centers, and academic departments);
- ❑ demonstrate in practical ways how a leading research university can, by serving its own research and teaching mission, seamlessly provide essential support for teaching and learning in schools, community colleges, and in other centers of higher learning;
- ❑ demonstrate in practical ways the benefits that accrue to leading research libraries through their meaningful collaboration and deep resource sharing.

4.3. Key obstacles. Key obstacles are technical and organizational/cultural in nature and are discussed below.

4.3.1. Obstacles to building configurable virtual collections.

- ❑ Metadata consistency. Metadata is information about a digital object and may contain references to its contents, provenance, structure, location, and terms of use. At least some metadata is available for all digital collections, for example, their URL and any descriptive information that may be gleaned from the Web page to which the URL links.^{xix} Some digital collections make extensive metadata available for the individual items they contain. For example, in David Rumsey's collection of American maps, the collection is known by its location and title.^{xx} Items in the collection (digitally reformatted surrogates for historic American maps) are also described with information about their title, creator, date of creation, etc. Metadata varies enormously in terms of quality (how much information is accurately supplied) and quantity (how much information is supplied about a collection or item). Yet the ability to search effectively across a corpus of distributed online materials relies, in large measure, on the consistency of metadata. Although it is relatively straightforward to "dumb down" richly descriptive metadata records by stripping them back to some minimal representation of creator, title, and location, we have not done as much as we might in exploring the means of "smartening up" records that are less descriptive.
- ❑ Selection. We have few robust mechanisms for searching selectively across distributed collections, and developing views of them as virtual collections that bear on certain themes. It is relatively straightforward to search simultaneously across the University of Michigan's Making of America, the Library of Congress' American Memory, the holdings of the University of Virginia's electronic text center, and the University of California libraries' Online Archive of California. Developing a subset of these collections that bears on the ecology of the Pacific Northwest, or that represents a corpus of the great works of western American fiction is less straightforward. Collection-level records are insufficiently descriptive while item-level ones are too varied to return reliable results. To build virtual collections selectively, we may need to acquire and subsequently enrich available metadata, possibly exploiting controlled vocabularies, subject and other thesauruses, or query trails that can be made available from large databases (even online library union catalogs) to help us automatically intuit and ultimately reveal the information objects that are actually required.

- ❑ Access to master files. We are becoming aware that a link to a digital collection (or to an item within it) that is available in a metadata record may not support some of the more interesting functions that we may wish to build into a virtual collection. We might, for example, want to make works in a virtual collection of western American fiction available through local print on demand services, by downloading them to a PDA, through Blackboard-based course packs and, of course, via the Web. To support this kind of application, we may need to acquire the master files that an institution creates for its online digital objects, not just the objects' metadata and HTML versions. At present, institutions that create openly accessible digital collections are loath to give up their master files. This is more of a cultural/organizational problem than a technical one.
- ❑ Format diversity. Even if institutions contributed master files as well as metadata to virtual collections, the diversity of those files would constrain their use. Take, as an example, the scenario described above in which the creator of a virtual collection of works of fiction, wishes to make the works available by a variety of means. Delivery to the screen is not a problem. Most of the file formats users encounter (HTML, XML, PDF, TIFF) can be rendered intelligibly by a Web browser. None would be rendered intelligible by a PDA or a short-run digital printing system without substantial reformatting. The problem is one that needs to be addressed collectively by those who build collections and those who aggregate collection content in order to present it in new and different ways. Content creators may need to review the role that standards play in their work. At present, library and scholarly communities have resolved a number of standard metadata representations (even though they may not be applied with any degree of consistency), and are making some progress (as represented in work on MPEG21 and METS) thinking about consistent mechanisms for representing format diversity. Borrowing concepts from the file sharing and data grid communities (the former made famous by Napster and rendered successful by the proliferation of the MP3 file format), content creators may need to think more about what it means to develop collections that can be used for their own local purposes while supplying “bricks” in a data grid that can be used to build an infinite array of virtual edifices. At the same time, service providers need to think about the development of a library of standard reformatting and data processing routines that are shared and used routinely to deal with commonly encountered cases.
- ❑ Library collaboration with end users. Many of the great library and museum collections in the world today began their meteoric rise in the 19th century. Their collections manifest imperial ambition or civic pride and were born of the close interaction of scholars and curators working in very embryonic, often densely connected, and highly overlapping “professions.” This close interaction continued well into the 20th century before it somehow became unwound. Nowhere is evidence of this decoupling in greater display than in the online collections emanating from our cultural memory organizations. They focus more on local holdings than on specific research and teaching needs, and are built with a view that if the information is made available, then scholarly and instructional use will ultimately follow. This logic has not been proven to be correct, and has given rise to a concern often cited by librarians, scholars, and the federal and philanthropic funding agencies that support their work, that the scholar and the teacher be reintegrated into the curatorial process. Yet nearly 150 years after the birth of the great modern library, the cultural landscape has changed dramatically. Professions that were once inseparable, indeed, indistinguishable, have grown apart in their needs, interests, and practices, even though they continue to share common objectives.^{xxi}

4.3.2. *Obstacles to integration of online collections and instructional technologies.* Here we encounter virtually all of the obstacles in the path of virtual collection development. In order to use existing digital information in their work, teachers have to locate, select, acquire, and reformat it so that it can be rendered intelligibly by the software they use to create and manage learning materials. There are also a number of distinctive challenges:

- ❑ **Simplicity.** Some cultural memory organizations have technologists on their staff. This is not typically true of researchers and teachers outside large university science departments. Accordingly, solutions that may help libraries and even some large or technologically sophisticated academic departments build locally configured virtual collections, will need to be developed, refined, simplified and made part of readily available desktop software in order to have deeper impact.
- ❑ **Generalizability.** The market for online learning systems is immature and offers a great variety of very different products, no one of which predominates. Although the market is beginning to converge around a small set of standards, they are highly descriptive. Accordingly, products can (and do) claim standards compliance without being interoperable. In order to support online learning, our work must integrate with a great diversity of products and approaches to courseware creation and management.
- ❑ **Organizational fragmentation.** The purveyors of digital libraries and instructional technologies, respectively, reside in independent and competing spheres even where they are institutionally co-located, as they are on many university and college campuses. As a consequence, it is difficult for any one of these stakeholders to realistically claim ownership for or direct resources into the integration of digital libraries and instructional technologies. Those few institutions that have achieved some synergy do so through a combination of strong leadership (as at MIT), and strategic planning (as at Indiana).^{xxii} The functional division of labor found on college campuses is replicated in the associations through which instructional technologists and libraries consort. The IMS Global Learning Consortium (a consortium that is developing open specifications of instructional technologies) has only a few library members, and the application of instructional technologies is only just emerging on the agenda of the Digital Library Federation and the Association of Research Libraries.^{xxiii} Institutionally, organizationally, perhaps even culturally, these two communities of practitioners are more disposed to replicate mutually interdependent functions rather than to seek more optimal cross-fertilization.
- ❑ **Motivation and incentive.** Innovative research and teaching is driven in combination by passion, commitment, and reward, and is constrained by time and resources. While the first two drivers are sufficient to stimulate some individual experimentation with new technologies, the reward structure typically joins time and resources as a constraint for the vast majority. For research, reward is based principally on publication in peer-reviewed journals and monographs. It does not account for the time spent creating research data that is consistent and well documented, and as such, readily and immediately useable by others. Teaching, in turn, is rewarded by a combination of student results and throughput. Innovation is “nice but unnecessary,” and may impinge on progress with other key success factors. Furthermore, in-person lectures and seminars are enormously efficient means of delivering instruction. They are also personal, can be interactive, and meet most common expectations. Innovative use of technology is resource intensive. It is also risky. The results are unknown, and given the limited amount of longitudinal data, hard to know. It is not at all clear yet what motivations and incentives will inspire the individual experimentation necessary to identify

what kinds of collections and tools will actually facilitate the effective integration of online digital materials with instructional technology.

4.3.3. Obstacles to capturing and encouraging re-use of online materials that result from research and teaching. Here we encounter some of the same obstacles that impede greater exploitation of online library collections in computer-based learning. MIT's DSpace and UC's eScholarship Repository both demonstrate that scholars will contribute their online materials to managed collections if the process is simple. Yet we have no clearer understanding of what motivations and incentives might encourage scholars to prepare and submit the online products of their work. Organizational fragmentation also plays a part. What department in a university or college is responsible for managing the digital assets that scholars create: the library or the computing center? And what role do scholarly communities play? Don't the ArXiv of physics pre-print publications and the social science data centers (e.g. at Michigan) suggest that the profession, not the institution, is the logical locus of aggregation?^{xxiv} And if the profession is the logical locus of aggregation, what motivates any institution to step in and subsidize its work? The locus of responsibility is even more difficult to locate for schools than it is for universities and colleges. Individually, schools are likely to be unable to support the requisite infrastructure, yet logical combinations through which scarce resources may be pooled are as difficult to conceive. Should such combinations be based at the county or school-board level, or managed statewide? The question is complicated by the free-rider problem that also impedes universities from developing infrastructure that supports scholarship outside their walls. Repositories that manage learning objects created by and used in the California schools would benefit instruction in Tennessee and Rhode Island even without contribution from those states.

5. Activities and outputs

Work on this proposal will take place in four phases: user needs assessment, product design and implementation, product review and assessment, and product operationalization. Funding is sought to support the first three phases over a twenty-four month period. Through the CDL, the UC libraries anticipate providing long-term programmatic support for the virtual American West collection and tools that result. This commitment is matched by our content-supplying partners who endeavor to maintain open access to the very substantial digital collections on which the virtual American West collection will be based. Work on each of these phases is set out below in some detail and summarized in Appendices 4 (chart of indicators, baselines and targets), and 5 (project timelines and deliverables). Work will be punctuated by three meetings of the CDL's core partners.

- ❑ An initial meeting held in month one of the project will ensure that partners share an understanding of the project's aims, goals, and processes. It will also review content suppliers' current collection content and technical capacities, and any constraints they impose upon our thinking about the American West collection and the technologies we use to create, present, and integrate it into instructional learning environments.
- ❑ A second partners' meeting held in month 5 will review functional, technical, and service operational specifications that are developed out of user needs assessment and develop a detailed implementation path for building the initial prototype of the American West collection and associated tools.
- ❑ A final meeting held in month 12 will review implementation progress, identify problems and strategies for resolving them, and identify development targets for months 18-24 of the project, that is, after the initial release of the American West collection is out for user review and evaluation.

5.1. User needs assessment (phase 1, months 1-4) and product review and assessment (phase 3, months 18-24). Users will play essential roles in refining our understanding of the problems

we seek to address, and the obstacles we will need to overcome in order to solve them. They will also help to define the scope of the American West collection and the functional specifications for the tools we will associate with it to support its use. They will also be closely involved in evaluating the collection and tools once they are developed as working prototypes. Because we anticipate that our users' needs and interests may differ, we will consult with them as grouped below. To ensure that the perspective we gain is not solely bounded by the State of California, we will rely on the very active networks that the Colorado Digitization Program maintains with university faculty and researchers, with public librarians, and with K-12 teachers and librarians in Colorado and selected other western states through its Western Trails project.^{xxv}

- ❑ University faculty and researchers. An interdisciplinary users' group will be assembled, drawing at a minimum on faculty from the University of California, Stanford, and the universities in the purview of the Colorado Digitization Program which together house leading scholars working on aspects of the American West. Many of these faculty from UC and Stanford are already engaged in UC library initiatives, notably in California Cultures (a project that brings together a collection of digitally reformatted library materials to bear on California's historic diversity), the Japanese American Relocation Archive (JARDA – another themed digital library collection focusing on the Japanese-American experience of relocation during World War II), the Online Archive of California (OAC), and in the Jepson Herbaria (in a project that integrates access to distributed online collections bearing on the state's ecological diversity). Additionally, UC hosts a number of research centers that specialize in relevant aspects of the American West.
- ❑ Community college faculty. Given The William and Flora Hewlett Foundation's extensive contacts with the community college sector, the CDL will seek assistance from the Foundation in establishing a user group whose members are drawn from that sector.
- ❑ Research libraries. Drawing on the extensive committee structures and the numerous activists who are critical to the success of the UC libraries' deep collaboration, the initiative will draw upon bibliographers, collection development officers, and digital library specialists who are responsible for building collections that support research and teaching on the American West.
- ❑ Public libraries. The Colorado Digitization Program has extensive experience working with public libraries, helping them gain maximum advantage from online information resources and digital technologies. The CDP will convene a group of public librarians whose collections are built for a far broader and more general user community than that which falls in the purview of research and academic libraries. Public libraries and their users are key constituencies who stand to gain a great deal from the collection development potential that our work may afford. The CDL will use its close working relationship with the California State Library to ensure that California public libraries are well represented in this user group.
- ❑ K-12. The Digital California Project plays a key role in brokering relationships between teachers, librarians, school administrators, and a world of network and information providers. It has offered to convene a group of K-12 librarians and teachers with an interest in the American West.^{xxvi} At a minimum, the group will include staff from school districts where the CDL or its partners have good working relationships, notably from the Oakland Unified School District, the Piedmont Unified School District, and with the schools that are active in the work of the Colorado Digitization Program and the UCLA National Center for History in the Schools, respectively. School librarians and teachers will be combined in a single group in respect of the fact that there may be no clear functional distinction between their curatorial and instructional responsibilities. It may be prudent to distinguish K-8 from 9-12 in light of

the likely distinctions in both the level and nature of their educational efforts, and the distinctive information needs they entail.

Consultation with our user groups will be iterative and ongoing. It will involve a series of one-day workshops and a period during which users assess the American West collection and tools by applying them in their work.

An initial workshop will be held with each group to refine our understanding of problems and obstacles (as set out above), and to develop functional specifications for the collections, tools, and strategies required to address them. Workshops will be hosted regionally (months 1-4) to minimize travel costs and conducted along similar lines to gain insight into:

- ❑ how participants would define the extent and scope of virtual collections they would build based on materials represented in the American West collection;
- ❑ the obstacles participants perceive in their exploitation of online materials, and the tools, services, or mechanisms that would help them overcome those obstacles;
- ❑ how participants would wish to integrate openly accessible digital materials with locally owned or managed collections;
- ❑ how participants wish to present virtual collections to their users, and what end-user configurability may be required.

Where workshops involve teachers, we will also ask specifically about:

- ❑ what role online collections might play in instruction in general, and in the development of online learning materials that support instruction in particular;
- ❑ what obstacles teachers encounter when thinking practically about using online collections;
- ❑ what tools they would require to enable their more effective use of online collections in instruction.

Where workshops involve teachers and/or researchers, we will also ask about:

- ❑ the level of interest participants have in making the online results of their research and teaching available via persistently managed collections;
- ❑ the motivations and incentives that would raise their level of interest;
- ❑ the tools they would require to make effective use of such collections, both as contributors and consumers.

In order to make the greatest headway at the workshops, we will ask participants to develop some familiarity with the online collections that have been offered to us in support of our work (e.g. from the University of California, the university libraries at Indiana, Michigan, and Virginia, from the Library of Congress, and from the Colorado Digitization Program), and to come prepared to speak about these collections and any others that they know of or use. To assist participants in their preparation, we will prepare and circulate in advance a brief list of questions that we will ask them to think about as they conduct their content review. Questionnaires will be tailored to participants' specific interests and roles.

To assist us in our exploration of the obstacles and incentives surrounding the use of online collections in instruction and the submission of scholar/teacher-produced online materials into managed collections, workshops will involve some onsite demonstrations from which it will be possible to gather quantitative and qualitative data on user reactions to selected approaches that are available via the Internet. After a day of discussion, demonstration, and brainstorming, participants will be asked to rank in priority order the most valuable collection development strategies, content, and tools that have been identified throughout the course of the day.

A second series of one-day workshops will be hosted 18 months into the project. Here, we will introduce the user groups to the prototype American West collection and the tools we have developed to assist in its use. The aim of this workshop series will be to provide a sufficient level of hands-on instruction so user-group members can evaluate the prototypes in their own curatorial, research, and teaching work over the next several months. During that trial period, the project will provide users with access to help desk support.

Product assessment will take place at a third workshop series conducted in months 23 and 24. Here we will ask participants to review the tools and services in light of their six-month's experience with using them. To guide this assessment, we will provide participants with a list of questions to think about as they use the collections and tools in their work. We will also get users' initial reactions to new content and functionality that has been added to the American West since the initial prototype was released in month 18.

The entire assessment process will be thoughtfully designed and informed by two one-day meetings where we will assemble between six and ten experts to advise on and assist in developing appropriate instruments. Some of the experts will be supplied by our partners. Others will be invited to participate to ensure the project benefits from objective external perspective. An initial design meeting will be held in month 1 of the project. It will frame the first series of user workshops and develop instruments and strategies that assist us in teasing out user requirements. A second design meeting will help be held in month 16 to develop the training tools that will introduce users to the prototype American West collection and its associated tools, as well as the assessment instruments that our users will deploy to help us in the products' evaluation.

5.2. Product development (phase 2, months 6-24). Working from the needs assessment, this phase will build functional, technical, and service operational specifications for American West collection and its associated tools, review and select promising technologies and approaches, and build the collection and tools.

5.2.1. Building functional, technical and service operational requirements and initial technology review (months 4-5). Based on the user needs assessment conducted at the initial series of workshops in the first four months, we will:

- ❑ compile formal user functional requirements for the American West and its associated tools;
- ❑ create technical and service operational specifications for the collection and its tools basing these on the statement of users' functional requirements;
- ❑ review existing technologies and approaches against the technical and service operational requirements, with a view to identifying solutions that look to be the most promising;^{xxviii}
- ❑ meet with our content-supplying colleagues to review functional and technical requirements, and identify practical implementation paths for building the American West collection.

We will then turn our attention (months 6-24) to the development of:

- ❑ a virtual collection bearing on the social and ecological diversity of the American West;
- ❑ tools that enable scholars, teachers, and curators to configure the collection in ways that meet their specific requirements;
- ❑ tools that ensure that content in the collection on the American West (or locally configured versions of it) can be integrated with online learning environments; and
- ❑ tools that enable scholars, teachers, and curators to submit into managed digital collections, the online information they create when using the American West collection.

Our work in each of these areas will be based on the functional and technical specifications developed out of our needs assessment, and focus on the most promising technologies that

emerge from the technology review described above. It will also bring us into close consultation with our content-supplying colleagues, each of them leaders in the digital library world with very substantial expertise in collection integration and presentation.

The development work will be conducted in months 6-24, even after an initial prototype is released for user review in month 18 of the project.

5.2.2. Building a virtual collection and the tools that enable its flexible configuration. We will assemble an initial virtual collection that integrates access to materials selected from those on offer the University of California, the University libraries at Indiana, Michigan, and Virginia, the Library of Congress, and the Colorado Digitization Program.

Because the online collections that are on offer do not deal either wholly or exclusively with aspects of the American West, our work in this area will focus on an initial exploration with our collaborators into the technologies and approaches that may permit end users to configure the collection in ways that meet their needs.

Although we would anticipate refining the emphasis of our work in light of our needs analysis (see above), the following areas appear to offer particular promise at this stage:

- ❑ Application of proprietary filtering tools (e.g. as used by most Internet search engines) and approaches (e.g. Centroids). These approaches would enable selectivity on the basis of the metadata without further enrichment of it.^{xxix}
- ❑ Capture and local enrichment of available collection and item-level metadata. By capturing metadata content (for example, by using the Open Archives Initiative Metadata Harvesting Protocol, OAI-MHP, or selected Web crawling technologies that are becoming available), we can minimize the number of remote targets that need to be queried by any search, thereby improving system performance.^{xxx} Local management of captured metadata may also provide a key to unlock third-party information resources (e.g. at the Library of Congress, Michigan, etc.) that are hidden because they make up parts of collections that provide limited item-level descriptions. Using Web crawling techniques, it may be possible to create and then gather into the prototype searchable item-level information for collections that do not natively supply it in sufficient abundance or detail. In addition, it may be possible to build more fully functional search services based on locally controlled metadata. Such metadata can be enriched with terminology controls dynamically applied to support specific user communities accustomed to searching aided by highly specialized, subject specific ontologies, and by applying weighting schemes that support different approaches to relevance ranking (e.g. as may be based on frequency of use, authority of underlying source, etc.).
- ❑ Application of “portal building” products currently available on the open market. Most of these products combine some of the technologies and approaches, notably application of the OAI-MHP, that are mentioned above. The market for these products is small but growing, and a number of promising new ones are beginning to emerge including MuseSearch™ from MuseGlobal, ENCompass for Resource Access from Endeavor, WebFeat, MetaLib from Ex Libris, Ltd., and Z Portal from Fretwell-Downing, Inc.^{xxxi} These products promise flexible means of presenting virtual collections, enabling a high degree of institutional and even individual configurability. Although the configurability our users will require is difficult to predict in advance of the needs assessment, our own work in this area suggests that at a minimum it will include personalized approaches to ranking, filtering, de-duplicating, merging, and sorting search results that may be returned when querying a virtual collection,

creating personalized alerts (e.g. to new material as they are added to underlying collections), and creating links to other collections that may be outside the purview of the virtual collection. Users are also likely to want to capture results for reuse in other contexts and applications.^{xxxii}

- ❑ Review of current digital object standards in light of their ability to support selective recombination and collection integration. At present, library and scholarly communities have established a number of standard metadata representations (even though they may not be applied with any degree of consistency), and are making some progress (as represented in work on MPEG21 and METS) thinking about consistent mechanisms for representing format diversity. Because our work will require selective integration of digital collections and collections content developed to different specifications, it will enable some assessment of any minimum requirements that collections should meet to act as interchangeable and recombinant parts in a network of openly accessible information resources. The review will take place with our content-supplying collaborators in light of the possible tension that exists for them between the development of materials that meet immediate local needs and those that meet global ones, including those of unseen and unknown users interested in developing virtual collections.

5.2.3. Building tools for integrating collections content with instructional learning environments

As described above, the CDL is working in partnership with the Interactive University Project at UC Berkeley on a tool known as the Scholar's Box. Briefly, the Scholar's Box enables users to select items discovered in online collections, include these items in a personal workspace where they can be organized, arranged, and annotated, and then output in formats recognizable to a variety of commonly used learning management and office software. Work on the Scholar's Box is at an early prototype stage, but it shows considerable promise. Already, the Scholar's Box is able to:

- ❑ recognize bibliographic references available in UC's Melvyl Union Catalog, incorporating those references in learning materials in a manner that preserves their hotlink to the live Melvyl Catalog database;
- ❑ recognize digital images in most common image formats and a limited set of encoded texts in HTML and XML-encoded formats, and include them in learning materials, with any metadata;
- ❑ arrange and annotate digital objects;
- ❑ output learning materials as selected, arranged, and annotated in formats that can be recognized by many office and learning management software products, such as WebCT, Blackboard, OKI-compliant products, and the Microsoft Office suite of software.

The Scholar's Box is also showing some facility with Web-based collaborative tools, notably weblogging, which will ultimately enable teachers to develop learning materials that encourage dynamic group use.

Work on the Scholar's Box is highly exploratory and draws primarily on a small number of creative technologists and educators who have been able to build only to a prototype or demonstration stage. It lacks the support necessary to make demonstration software robust enough for production-level use. Accordingly, we are seeking support from The William and Flora Hewlett Foundation to leverage this development effort and to translate it into robust operational tools that will enhance both the functionality and use of the American West collection. Our aim is to offer the Scholar's Box as a tool that can be used either directly with the American West collection as assembled by the CDL, or with the virtual collections that are created by our users.

The American West collection will also act as an important testbed digital library collection for a consortium at UCLA that is working with CDL on the integration of digital libraries and instructional technologies. Briefly, that partnership looks forward to a time when instructional technologies converge on a data representation standard, notably the Shareable Content Object Reference Model or SCORM, that is only now in its ascendance.^{xxxiii} Briefly, the partnership is evaluating how digital library collections located at CDL and UCLA can present their contents in a manner that is directly accessible from SCORM-compliant learning management systems without the need for an intermediary. The partnership benefits from UCLA's development of an enterprise-wide instructional technology environment that conforms to emerging standards such as SCORM and that implements tools and architectures that are emerging from the Open Knowledge Initiative – “a collaboration among leading universities and specification and standards organizations to support innovative learning technology in higher education”.^{xxxiv}

5.2.4. Building mechanisms that enable scholars, teachers, and curators to submit materials developed with the American West collection into managed collections. Our work in this area has identified three approaches, each with its own strengths and weaknesses. They are as follows:

- The use of tools that enable data creators to quickly and easily submit any digital information they create (whether peer reviewed publications, online learning materials, research data, encoded texts, or digital images) into managed collections. The DSpace software developed by MIT and Hewlett Packard is representative of this approach and has been successfully deployed at MIT. By reducing complexity for data creators, however, it increases complexity for the information manager who is left with a warehouse of largely undifferentiated content. Unbundling the content and transforming it into well organized, manageable, and readily accessible digital libraries is a challenge that has yet to be successfully confronted.
- The use of tools that enable data creators to quickly and easily submit certain classes of digital information they create (e.g. online documents or digital images) into a managed collection designed specifically for those classes. The ArXiv physics preprint server to which physicists, mathematicians, and other physical scientists submit research papers via relatively easy-to-use means is representative of this approach, as are the many “institutional repositories” that have sprung up on university campuses in emulation.^{xxxv} The University of California libraries have elaborated on this approach. The eScholarship Repository accepts faculty publications, and their image service is being developed to allow submission of personal digital image collections.^{xxxvi} This approach retains the simplicity of submission that is the hallmark of DSpace, without imposing an additional burden on the information manager. Each service is responsible for a single class of digital objects that share format and metadata characteristics, and are consequently easier to manage and represent to users as part of coherent collections. At the same time, the approach, if articulated over a wide range of digital object classes (whether defined by data format or by some other means), will introduce complexity for data creators who will be confronted with having to choose between different managed collections, each focusing on a specific type of material and having its own submission routines.
- The development of digital collections through direct and in-person negotiation with data creators for the acquisition of their materials. This approach is adopted by subject-based data archives, such as the InterUniversity Consortium for Social and Political Research in the U.S. and the Arts and Humanities Data Service in the U.K.^{xxxvii} Through the CDL, it has been adopted by the University of California libraries as a means of acquiring and persistently managing large-scale data resources, such as commercial databases and government and

social scientific datasets.^{xxxviii} Here, collection managers define what materials they will collect and the criteria that guide decisions about which actually to acquire. The criteria may focus on a data resource's intellectual merits and potential re-use. They will also specify acceptable data formats and documentation standards. This approach is especially useful in the acquisition of data resources that are too large and too complicated for submission through easy-to-use tools such as those developed by DSpace and the eScholarship Repository. Its limitations are that it shifts some of the burden and the cost of developing high-quality and manageable data resources onto the data creator. As such, it tends only to attract large-scale data resources that have absorbed considerable financial investment and have some immediately apparent lasting value (such as national polling and census data, commercial opinion polls, large-scale online journal databases, large-scale digital library collections, etc.).

Our work with each of these approaches suggests that all of them may be necessary if we are to capture and manage the full range of digital information that curators, scholars, and teachers produce with (or may wish to contribute to) the American West collection.

- For the foreseeable future, negotiated acquisitions is likely to be the only means of acquiring large-scale and complex data resources.^{xi}
- Tools that manage specific classes of digital material are compelling because they provide a means of developing the digital equivalent of library special collections. The eScholarship Repository articulates the approach particularly well. Individual depositors (in this case, academic departments and research centers at UC), claim a place in the Repository and take responsibility for documenting and organizing the materials they submit. The Repository currently contains more than 100 special collections, each of which can be accessed and searched independently. The collections can also be searched as if they make up a single database. Extending this approach to include other data classes, notably images, potentially enriches the number and variety of special collections that can be made available.
- Tools that permit submission of any and all types of digital content are likely to be useful with learning objects and course materials. These materials are inherently diverse. They comprised digital objects in a multiple range of formats and as such may not be easily represented in or managed through services that capture only a specific class of digital information. Second, many learning materials will be created by instructors with limited technical abilities and who are most in need of very simple submission tools.

Our aim, then, will be to make these three kinds of data capture procedures available from the American West Web site so that we can encourage the submission of materials that are produced as a result of its use. Work will be undertaken in the following five areas:

- Subject to our negotiation of an acceptable license for Bepress, the eScholarship Repository software, we will create an instance of the eScholarship Repository specifically designed for publications resulting from research on aspects of the American West.^{xli}
- We will continue our work on the image service so it is better able to accept personal digital image collections with a view towards being able to accession personal image collections into the American West collection. At present, our work is developmental and undertaken along two paths — one with Luna Imaging Inc., the software company that supplies our image delivery system, and one with the CDL, which is developing a suite of tools to enable the automated acquisition of third-party digital images. Both developments have their advantages. The work with Luna leverages the company's considerable investment and expertise. The work in CDL is built upon scarce technical

- resources, but promises an open source solution that mitigates dependence on a third-party solution.^{xlii}
- We will work on the further development of the DSpace approach, notably by developing mechanisms that enable information managers to disentangle, or rather, organize and provide some collection coherence on materials that may be submitted. We would also expect to tailor an interface so that it is explicitly suitable for those wishing to submit course and learning materials. This work will leverage our partnership with the UCLA consortium which is exploring mechanisms for capturing and persistently managing faculty-produced online learning materials. It is reviewing DSpace and a range of like technologies as part of that exploration.
 - We will continue selectively to acquire access to large-scale data resources bearing on the social and ecological diversity of the American West.
 - We will represent available data capture mechanisms in a single place within the American West Web site, where we will be able to guide potential contributors in their effective and correct use.

These activities and outputs will help us overcome key obstacles. With regard to the technical obstacles, it will provide a practical opportunity to evaluate technologies that can enhance our ability to aggregate extant online collections and to configure and reconfigure them in ways that meet specific local needs. It will enable us to build and make publicly available a promising interim solution that can help bridge the interoperability gap between digital libraries and instructional technologies. Our work will also go some way towards addressing many of the cultural and organizational challenges that are indicated above.

- Our work is inherently consultative and will be guided and assessed at every stage by librarians, scholars, and teachers working in close cooperation. Given the standing and the reputation of the institutions involved in this initiative, we hope to lead by developing a working example, certainly not the only example, of scholar-led (or needs-based) digital library development.
- We also hope to create a powerful wake that can pull other cultural organizations along in recognizing that the greatest reward for their investment in digital collections is not their ownership and control of the material they create, but in the number and variety of rich, interesting, and enabling services that can be built upon it.

Other obstacles will only be overcome with additional and very powerful influences. For example, amongst the constraints to faculty and teachers making more effective use of digital technologies, we can only remove one: the limited availability of tools that enable easy exploitation of those technologies. Other constraints, such as the nature of the scholarly reward structure, and the absence of time can only be relaxed through the concerted effort required to change institutional and professional practices and public policy. Having admitted these very significant limitations, however, we would point to the efficacy that simple tools can have in changing organizational and individual behavior. The ArXiv preprint server, the UC eScholarship Repository, and MIT's implementation of DSpace have had profound impacts. In each case, new behaviors initially apparent in only a few users have become more pervasive and widespread.

Constraints that impede institutions from investing in services from which others will benefit (the free-rider problem) are also likely to persist for some time. Yet here too, we hope to have an impact. The University of California libraries are committed to leading this extensively collaborative and highly consultative effort. We have made a further commitment to sustaining the collections and tools that result as public goods. The commitment is matched by very

substantial investments past, present, and future. There is an element of self-interest. The collections, tools, and services will promote research and teaching at the University of California, extending the information resources we make available to faculty and researchers, and providing them with the means of disclosing to their peers the results of their work.

Yet while we serve our own self-interest, we do not shy away from the very substantial additional effort that will be required to make our collections and services available to a broader public:

- to build collections that disclose the wealth of cultural resources and scholarly knowledge that is available within the University, and are openly and freely available;
- to develop tools that enable public libraries, community colleges, and schools both inside and outside California to configure online collections so they meet local needs;
- to build tools that enable online collections to readily and easily be used in online learning;
- to maintain services that allow us to capture, persistently manage, and encourage widespread re-use of the research and teaching that results from work with online information.

We make this commitment at a time of unprecedented budget cuts. State funding for the University of California is down 25 percent from 2001/02 and as a result, the UC libraries are curtailing collection expenditure, reducing opening hours, and consolidating or closing public service points. We do this because our faculty demand and deserve the level of service that is entailed, because we take seriously our role as stewards of our digital cultural and scholarly heritage, and because we believe that much in our public service mission.^{xliii} We do this, finally, by way of example for our peer institutions – as a way of indicating what it means to be a leading cultural and educational center in a networked digital age, and as an open invitation to them to join us, because humbly, we know that we will accomplish so much more in the company of others.

6. Inputs

6.1. Collections. We will collaborate closely with colleagues at the following institutions in order to integrate the following online materials into the virtual collection on the American West:

- Library of Congress, American Memory. This is the pre-eminent online collection in the world. It is “a gateway to rich primary source materials relating to the history and culture of the United States. The site offers more than 7 million digital items from more than 100 historical collections” of the Library of Congress. American Memory contains many materials bearing on the American West including, no fewer than 25 online collections dealing exclusively with the American West and the Midwest.^{xliv} LC has also been a leader in the development and implementation of the OAI and Z39.50, network search and retrieval protocols that enable cross-collection federation, and in the design and application of digital library standards.
- The Colorado Digitization Program. In this statewide program, heritage and educational institutions work together to create and bring together digitized materials from all corners of the state in order to “ensure public access to the rich resources in Colorado, promote the economic and efficient delivery of full text and graphic resources to the people of Colorado, and contribute to the national effort to develop digital libraries and museums.” More than 20 institutions participate in the CDP, which has assembled tens of thousands of digital objects.^{xlv} CDP has also been a leader in collection federation. Its own collection is created by virtual aggregation across its partners’ holdings. Further, CDP has been powerfully represented nationally in a high-profile effort led by the federal Institute for Museum and Library Services to identify good digital library practices as those supporting interoperability and persistence.^{xlvi}

- ❑ University of Indiana library's Wright American Fiction collection. The collection includes works of "19th century American fiction, as listed in Lyle Wright's bibliography *American Fiction, 1851-1875*. There are currently 2,887 volumes," including the great works of American Western fiction from that period.^{xlvii} Indiana is also a leader in the application of text encoding standards and in the exploration of data interoperability for texts and other formats.^{xlviii}
- ❑ University of Michigan Library, Making of America (MOA) and other digital collections. The University of Michigan is a leader in the digital reformatting of printed monographs and serials. To date, it has made nearly 20,000 19th century American imprints openly available online, including those in its flagship collection, Making of America. The Making of America is a digital library of primary sources in American social history from the antebellum period through reconstruction. The collection is particularly strong in the subject areas of education, psychology, American history, sociology, religion, science, and technology.^{xlix} Michigan's digital library system, DLXS, has been widely deployed by libraries around the world, and it is a leader in interoperability implementation. Michigan was one of the first research libraries to deploy the OAI metadata harvesting protocol in a production environment. It is currently developing a new protocol (CGM) for distributed full text search and display with the Cornell and Goettingen Universities.¹
- ❑ University of California collections. Through the CDL, the university libraries have made available some 500 openly accessible online collections created by the UC libraries, museums, archives, academic departments, and research centers.^{li} The UC libraries' leadership in the development of digital libraries is documented in Section 2.^{lii}
- ❑ The University of Virginia library's digital media centers. The centers include the electronic text center – one of the pre-eminent text centers in the world – and are located in one of the most innovative digital libraries.^{liii} The centers maintain online archives of hundreds of thousands of openly accessible digital objects including tens of thousands of SGML and XML-encoded electronic texts. Many of the objects bear directly on the American West. Meantime, the Virginia library has been a leader in the development of interoperable digital library collections and services, and is perhaps best known for its groundbreaking work on FEDORA, a digital object repository and management system.^{liv}

6.2. Services and technologies. The CDL will be responsible for the project's leadership and implementation. Accordingly, the project will leverage a number of programs that the CDL sustains on behalf of the UC libraries. Programs that will contribute directly to this work are listed briefly below and described more fully in Appendix 1 (project organization and staffing).

- ❑ Built content – builds openly accessible digital collections through a variety of means.
- ❑ Digital library services – provides content access and integration services.
- ❑ Preservation – persistently manages digital library information on behalf of the UC libraries.^{lxiii}
- ❑ eScholarship – supports scholars in exploring new means for disseminating the results of their research worldwide audiences via the Internet.
- ❑ Technology – provides the technical infrastructure the CDL requires to supports its wide range of content development, online service, and exploratory initiatives.
- ❑ The Scholar's Box – a development initiative led by Berkeley's Interactive University Project, with support of the CDL and other organizations.
- ❑ Education and strategic innovation – the program includes an assessment unit that monitors the information requirements of our diverse user community, feeding back findings in ways that influence existing content and services and shape new initiatives.
- ❑ Public information – brings together hundreds of openly accessible online collections that have been created at UC, liaising with public libraries, schools, and colleges to encourage their use.

The key challenges the CDL confronts with this program are the same challenges that it faces in all its work. Like so many digital libraries, the CDL evolved as a variety of highly focused content and service development activities, each with its own organizational and technological base.^{lxiv} This approach produced rapid growth in the library's collection and service portfolio. It also introduced organizational inefficiencies where commonly required technology functions (content management, searching and indexing tools), and skills (assessment, design, communication, product review, and licensing) were redundantly supplied by different programs.

Beginning with a strategic review launched in May 2002, the CDL began a major restructuring effort in order to iron out these inefficiencies. The organizational aspects of that restructuring are completed. Staff are organized in functional units whose members support a variety of content and service programs. The technical restructuring is still underway and involves the development of a suite of common tools that can be deployed effectively by different projects. While the technical transition is underway — it is expected to continue for at least another twelve months — the CDL technology division needs to do two things simultaneously: develop the common technology infrastructure and transition old infrastructure to new as it becomes available; and maintain existing systems by providing them with any incremental development they might need. New initiatives such as the one proposed here, provide practical opportunities for us to develop and evaluate essential pieces of the common infrastructure.

The CDL's key strengths reside in its highly motivated and energetic staff, in the extensive networks it maintains systemwide and nationally with libraries, vendors, technologists and technology researchers; in its grounding in needs-based development and assessment; and in its approach to project management, which ensures that every project has access to the appropriate range of specialist staff that it requires, irrespective of the CDL divisions in which those staff are located.

7. Evaluation

Evaluation and assessment are essential components of the CDL's work. New content and service initiatives are informed extensively by (and often arise out of) qualitative and quantitative assessment of our users' information requirements. The use that is made of existing services and collections is also closely monitored in a manner that supports evaluation of their performance, acceptance by users, and further development. In assessing users' needs and the use and usability of our collections and services, we use a variety of assessment strategies (both qualitative and quantitative) and systems, since no one system meets the needs of the all of the online services that we offer.^{lxv}

Needs assessment and use and usability testing in this project will involve a variety of approaches, including focus groups, onsite demonstrations, and online "Webinars" with members of our user groups to collect quantitative and qualitative data on reactions to selected Web-based information content and applications. Methodology will include capturing user input through audio or videotape, Web transcriptions, written questionnaires, and informal notes. We will rely upon focus groups, structured surveys, user protocols (watching users conducting a number of specified tasks with a specific application), and observation in order to evaluate the outcomes of our work. Use of the American West collection will also be monitored through transaction log analysis, and by monitoring contributions to repositories and the kinds of virtual collections that are based on the American West. Technical development work at CDL is also subject to a formal assessment and review at various stages, focusing on performance (both in terms of speed and of load), interoperation with other essential systems, and prospective maintenance and development paths.

Finally, we offer the following three reasons why the work should be supported this year.

1. The University of California libraries' formal and financial commitment to building to a layered library service model – one that enables them to disclose the wealth of holdings online along with a suite of tools that enables their configuration to suit a virtually infinite number of research, educational, and civic uses inside and outside the University of California.
2. The momentum that has gathered in support of this activity including the extensive network of essential partnerships that are essential to its progress.
3. The considerable early investment in the user needs' assessment and technology design that is underway in support of this developmental direction.

ⁱ University libraries include those at Berkeley, Davis, Irvine, Merced, Riverside, San Francisco, Santa Barbara, Santa Cruz, UCLA, and the California Digital Library.

ⁱⁱ The University of California libraries claim 32 million volumes; Harvard libraries, by comparison, have 14 million. The digital collections extend to some 8,000 electronic journal titles, 250 reference databases, 4,500 government and statistical databases, 200,000 digital images, 7,000 finding aids (for archives and special collections in UC and the State of California), all books published in English before 1700 and several major historical newspaper archives.

ⁱⁱⁱ See Daniel Greenstein, et. al. "New models of library service: Deep resource sharing and collaboration at the University of California" from <http://www.slp.ucop.edu/documents/newmodelsoflibraryservice4-17-03.pdf>

^{iv} The extent of the digital collections is referred to above. The libraries' foray into the development of shared print collections has attracted international interest as they break with traditional and increasingly uneconomical collection development strategies that have tied access directly to libraries' highly redundant ownership of physical materials. See Bernard F. Reilly, Jr., *Developing print repositories: Models for shared preservation and access* (CLIR, June 2003), from <http://www.clir.org/pubs/abstract/pub117abst.html>. See in particular the forward by Abby Smith.

^v The University of California libraries are unique in maintaining a large-scale and core-funded digital preservation program. A small number of other programs exist at selected national libraries, including the Library of Congress, the National Library of Australia, the British Library, the Royal Library of the Netherlands, and the Bibliotheque National in France.

^{vi} From <http://escholarship.cdlib.org/>

^{vii} The University of California libraries have played a leading role in the development and widespread adoption of a large number of information standards and protocols that are now in common use, including Z39.50, METS, EAD, and OpenURL.

^{viii} The Web site was designed and developed in close consultation with educators, librarians, and others working at all levels in the California educational system, and has been used and well received in a number of educational settings and in teaching in several disciplines. See <http://www.californiadigitallibrary.org/>

^{ix} For the Interactive university Project see <http://interactiveu.berkeley.edu:8000/IU/>. Work on instructional technologies also benefits from a collaborative effort with the Center for the Studies in Higher Education which is evaluating what use faculty make of digital library collections. That effort benefits from the generous support of The Andrew W Mellon and William and Flora Hewlett Foundations, respectively.

^x For more on this topic, see Lorcan Dempsey, *The Recombinant Library: portals and people* (forthcoming).

^{xi} <http://www.nsdlib.org/>; <http://www.americansouth.org/>

^{xii} Jerry Campbell, "The Case for Creating a Scholars' Portal to the Web. A White Paper", *ARL Bimonthly Report*, 211(August 2000) from <http://www.arl.org/newsltr/211/portal.html>; for the DNER see <http://www.jisc.ac.uk/dner/>

^{xiii} See Neil McLean and Clifford Lynch, Interoperability between Information and Learning Environments. Bridging the Gap: a joint white paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information (June 2003), from http://www.imsglobal.org/DLims_white_paper_publicdraft_1.pdf

^{xiv} See <http://ocw.mit.edu/>

^{xv} See the press release announcing the alliance at <http://www.imsproject.org/pressrelease/pr030313.cfm>. For CNI see <http://www.cni.org/>.

^{xvi}

^{xvii} In the state of California whose schools are ranked last in the United States by almost every measure that can be applied, nearly 75 percent of all public schools are connected to the Internet, thanks to the Digital California Project (<http://www.cenic.org/DCP.html>). In the next two years, that figure will grow to nearly 100 percent. Network penetration is not as good in the developing world, but is growing at a considerable rate.

^{xviii} In California, the Digital California Project was launched in part to identify sources of information that can utilize the extensive public school network that it has put into place. Yet it is sobering to point out that the DCP budget has fallen from a high of \$31 million annually in 2001/02 to \$15 million for 2003/04. Internationally, UNESCO has taken a leading role in this area. See <http://www.unescobkk.org/education/aceid/higher-edu/ODL/>

^{xix} Google and other search engines automate the process of metadata creation by “crawling” Web pages and grabbing basic information from them based on their location (URL) and any words that are represented on their pages.

^{xx} See <http://www.davidrumsey.com/>

^{xxi} See Abby Smith, Strategies for building digitized collections (CLIR, 2001), from <http://www.clir.org/pubs/abstract/pub101abst.html> and her *Why Digitize?* (CLIR, 1999), from <http://www.clir.org/pubs/abstract/pub80.html>

^{xxii} Daniel Greenstein and Suzanne Thorin, *The Digital Library: A biography* (CLIR, 2002) from <http://www.clir.org/pubs/abstract/pub109abst.html>

^{xxiii} See <http://www.imsproject.org/>, <http://www.diglib.org/>, and <http://www.arl.org>, respectively.

^{xxiv} <http://arXiv.org>; <http://www.icpsr.umich.edu/>

^{xxv} Western Trails extends the work of the Colorado Digitization Program to a collaboration involving Colorado, Nebraska, Kansas and Wyoming. See <http://www.cdpheritage.org/westertrails/>

^{xxvi} For the Digital California Project see <http://www.cenic.org/DCP.html>

^{xxvii} http://k20.internet2.edu/projects/mu_pacificlighthouse.html

^{xxviii} In this process we anticipate emphasizing the needs and priorities that are commonly identified by user communities whose interests and requirements are likely to differ rather substantially.

^{xxix} Alta Vista’s family filter (<http://www.altavista.com/web/ffset>), Google (whose filtering options are available from <http://www.google.com/preferences>) and Lycos Search Guard (from <http://searchguard.lycos.com/>). For use of centroids in searching see Junliang Zhang and Javed Mostafa “Information Retrieval by Semantric Analysis and Visualization of the Concept Space,” *Dlib Magazine* (October 2002), from <http://www.dlib.org/dlib/october02/zhang/10zhang.html> and John Kirriemuir, Dan Brickley, Susan Welsh, John Knight and Martin Hamilton, “Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach” in *Dlib Magazine* (January 1998), from <http://www.dlib.org/dlib/january98/01kirriemuir.html>

^{xxx} The CDL is rapidly gaining experience with these technologies in support of its work on behalf of UC libraries to preserve Web-based information. At present, it is working with Stanford University Computer Science Department and deploying its WebBase technology as a data-capturing agent. See <http://www-diglib.stanford.edu/~testbed/doc2/WebBase/>. It is also either experimenting with (or preparing to experiment with) comparable tools available from XSB (<http://www.xsb.com/index.html>), the Internet Archive (<http://www.archive.org/>), and the San Diego Super Computer Center (which deployed its tools in to harvest collections referenced by the NSDL). For the OAI metadata harvesting protocol, see <http://www.openarchives.org/>

^{xxx} See <http://www.museglobal.com/Products/MuseSearch/>; <http://encompass.endinfosys.com/>; <http://www.webfeat.org/>; <http://www.aleph.co.il/MetaLib/>; and <http://www.fdgroupp.co.uk/fdi/company/home.html>.

^{xxxii} This insight into users' needs is gained from our evaluation of the CDL SearchLight service – a discovery tool that allows users to search across a range of UC journal and other databases and bibliographic catalogs if they comprise a virtual uniform database. In fall 2001, the CDL evaluated the service and discovered that although it was well liked, its users required additional functionality that the service architecture (indeed the then current state of the technology industry) permitted. The SearchLight service is available from <http://searchlight.cdlib.org/cgi-bin/searchlight>. The report on its assessment is at http://www.cdlib.org/libstaff/education/evaluation/searchlight_focusgroup%20_summary_dec2001.pdf and the list of additional desired features is available from http://www.cdlib.org/libstaff/system_services/projects/searchlight/future_functions.rtf.

^{xxxiii} SCORM is distributed through the Advanced Technology Learning Network. For more information see <http://www.adlnet.org/index.cfm?fuseaction=scormabt>

^{xxxiv} See <http://web.mit.edu/oki/>.

^{xxxv} For the preprint server see www.arXiv.org. For work on institutional repositories see Clifford Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly report 226 (February 2003), from <http://www.arl.org/newsltr/226/ir.html>; and Raym Crow, "The Case for Institutional Repositories: A SPARC position paper" (August, 2002), from <http://www.arl.org/sparc/IR/ir.html>

^{xxxvi} The former is completed and being used successfully. To date it has attracted more than 2,000 faculty publications and is growing at a steady rate. The image repository is under construction and will be used to help extend a large collection (some 200,000) of digital image surrogates for art historical works, notably by enabling libraries and faculty to contribute their own image surrogates to the collection.

^{xxxvii} See <http://www.icpsr.umich.edu/>; <http://www.ahds.ac.uk/>.

^{xxxviii} See <http://countingcalifornia.cdlib.org/>.

^{xxxix} Indeed, our work building the initial collection on the American West is a form of negotiated deposit as it will require intensive discussions and co-development effort with our content suppliers.

^{xl} Indeed, our work building the initial collection on the American West is a form of negotiated deposit as it will require intensive discussions and co-development effort with our content suppliers.

^{xli} For the Berkeley Electronic Press see <http://www.bepress.com/>. In order to implement a level of quality control, sites on the Repository will be established and managed by cultural or educational institutions (a public library, a California high school) or by departments within those institutions (the history department of a CSU campus), rather than by individuals.

^{xlii} See <http://www.lunaimaging.com/insight/index.html>. CDL's own digital image delivery service underpins the Online Archive of California at <http://www.oac.cdlib.org/>.

^{xliiii} Indeed UC's public service mission has only grown in its importance as other California agencies that share our goals (the California State Library, the Digital California Project) are financially even less able to meet them than we are. Funding has been drastically reduced for the California State Library and the Digital California Projects both of them leaders in the provision of publicly accessible online education al and cultural materials for the people of California.

^{xliiv} Located from <http://memory.loc.gov/ammem/collections/finder.html>

^{xli v} See <http://www.cdpheritage.org/>

^{xli vi} See "A framework of guidance for building good digital collection" (November, 2001) from <http://www.imls.gov/pubs/forumframework.htm>

^{xli vii} See <http://www.lettrs.indiana.edu/web/w/wright2/>

^{xli viii} Perry Willetts, TEI Text Encoding in Libraries. Guidelines for Best Encoding Practices (July, 1999). The guidelines have been endorsed by the Digital Library Federation and are available from <http://www.diglib.org/standards/tei.htm>

^{xli x} From <http://www.hti.umich.edu/m/moagrp/>

¹ For DLXS (Digital Library Extension System), see <http://www.dlxs.org/> For Michigan's work with OAI, see <http://oaister.umdl.umich.edu/o/oaister/>; for CGM see <http://www.library.cornell.edu/mathbooks/cgmverbs.xml>

^{li} The collections are made available from <http://www.californiadigitallibrary.org/>

^{lii} cf. D Greenstein, et. al. "New models of library service", *op.cit.*

-
- ^{liii} <http://etext.lib.virginia.edu/uvaonline.html>
- ^{liv} See <http://fedora.comm.nsdlib.org/>
- ^{lv} Located from <http://memory.loc.gov/ammem/collections/finder.html>
- ^{lvi} See <http://www.lettrs.indiana.edu/web/w/wright2/>
- ^{lvii} From <http://www.hti.umich.edu/m/moagrp/>
- ^{lviii} For DLXS (Digital Library Extension System), see <http://www.dlxs.org/> For Michigan's work with OAI, see <http://oaister.umd.umich.edu/o/oaister/>; for CGM see <http://www.library.cornell.edu/mathbooks/cgmverbs.xml>
- ^{lix} The collections are made available from <http://www.californiadigitallibrary.org/>
- ^{lx} cf D Greenstein, et. al. "New models of library service: Deep resource sharing and collaboration at the University of California" from <http://www.slp.ucop.edu/documents/newmodelsoflibraryservice4-17-03.pdf>
- ^{lxi} <http://etext.lib.virginia.edu/uvaonline.html>
- ^{lxii} See <http://fedora.comm.nsdlib.org/>
- ^{lxiii} The program entails strategic partnerships with the San Diego Super Computer Center, the Stanford University Computing Science Department, and the UC Berkeley library. It's members also work closely with the Library of Congress's National Digital Information Infrastructure Preservation Program. The program is exploring a variety of Web-crawling techniques as means of capturing selected Web-based materials in order to ensure their persistent management.
- ^{lxiv} See Thorin and Greenstein, *op.cit.*
- ^{lxv} See Denise Troll Covey, *Usage and Usability Assessment: Library Practices and Concerns* (CLIR, 2002) from <http://www.clir.org/pubs/abstract/pub105abst.html>