Digitization and electronic communications are making manuscript and other special collections more available to more people than ever before. Libraries, special collections, and archives now routinely digitize materials for user access, for Web-based exhibits, and, increasingly, for preservation purposes. Although these activities have seemed revolutionary—by providing increased exposure to collections, making collections available to a much wider variety of users than ever before, and revealing hidden treasures—we are participating in a transformation that has only just begun. Most of the digitization and access activities that we engage in still parallel our activities in the physical domain fairly closely. We tend to present digital surrogates in much the same way as we would the physical materials, perhaps as adjuncts to a finding aid, or as a Web exhibit. Or we may design a beautiful Web-based page-turning application that makes that activity as close to the real thing as possible. Although creation of a virtual experience that parallels the physical environment is a worthwhile activity, there are a variety of services and techniques that can be added to digital collections to facilitate research and teaching.

This article will discuss issues surrounding digitization of special collections materials and the ways in which digital surrogates are used. Then, with reference to some projects we are engaged in at the University of California, Los Angeles (UCLA), I will suggest ways in which digitized special collections materials can be drawn more closely into the teaching and research activities of faculty and students. The projects to be discussed include a scholarly publication (an encyclopedia) and an instructional Web site (map-based urban studies) that draw special collections materials from a digital archive dynamically—that is, in real time—and are integrated more actively into scholarly and educational activities than previously possible.
Digitization, Digital Services, and User Feedback

A digital surrogate has a number of attributes of which it is worth reminding ourselves:

- In most cases, the digital file has been created to facilitate access to the information contained by the original item. There are other, more complex reasons to digitize materials—to reveal layers in a palimpsest, for instance—but they are the exception.

- A single digital image is a snapshot of the object, frozen in time, captured in a unique set of conditions (lighting conditions or physical condition, to name two). Each time an item is digitized, a new and different surrogate is produced.

- In most cases, the digital object will be described as if it were the original object, not something new. The only possible addition would be some administrative metadata, but this is probably rare outside a formal digital management system.

- Although some digital preservation activities may be undertaken, it is more likely that most information technology departments will limit their activities to traditional backup services. As digital archives increase, it will become imperative for institutions to gain access to reliable digital preservation services at reasonable cost.¹

- In most cases, the digital files will be used in ways that parallel traditional access mechanisms. These include linking files to finding aids and posting on Web sites, either for illustrative purposes or as part of a more formal Web exhibit.

Digitization is a start, but providing innovative digital services on top of the digital collections is essential to meet the needs of an expanding audience for special collections materials. As digitized material is made more freely available, new populations of users will be drawn to them, bringing with them a much broader range of needs. Meeting these needs will require a much wider range of services than is traditionally offered by special collections and archives.

¹ Institutions with heavy investments in digitized content are becoming increasingly aware of the importance of administrative metadata for the management and long-term preservation of this material. Library information system vendors are also aware of this and are developing more sophisticated digital content management systems with digital preservation services.
Strategies for Teaching and Research with Digital Special Collections

There are a variety of different services that a user of special collections materials might need. Some of the more obvious—and common—ones are: page turning and zooming applications for books and manuscripts; viewers for navigating large objects such as maps and architectural drawings; tools for searching through and organizing bodies of textual material; tools for organizing large collections of like material that may have limited metadata, such as collections of designs or photographs, or images (photographs or drawings) of a technical nature.

When the services added to a digital collection are specific to a collection or group of materials, we can make a distinction between “digitization” and “digital projects.” A digital project implies that there are aspects to the activities that go beyond simply digitizing material and making it available, such as provision of a group of services to support a specific type of access to a collection; or new and innovative ways of working with a collection or group of digital objects. It might involve access to collections whose originals are housed in a variety of different places, or the exchange of digital objects to meet user needs. And, increasingly, it might include services that allow patrons to provide relevant data and commentary to supplement the existing metadata.

Given that only a small proportion of a given collection is typically digitized, the question of “what” is digitized and how it is presented becomes very important. Just as a curator shapes a collection to some degree by deciding what to accept and what to reject, what to weed and what to preserve, so the builder of digital collections shapes the digital collection. This selection process also has an effect on the services that will be available. If an entire collection was digitized, the user would be able to use the collection just as he or she would the physical collection. However, we typically digitize only portions of a collection. Ideally, we would digitize entire collections as a matter of course; but, for financial, technical, and other reasons, this is not usually practical. By digitizing the more “important” or “significant” items in a collection, we are giving those priority, which may have the unfortunate effect of drawing attention further away from documents of unrecognized importance in the collection.

Minimizing this “selection problem” can be achieved by digitizing complete logical groupings wherever possible—for instance, a complete set of manuscripts, a complete book, all the photographs or all the correspondence within a particular collection. Just as you would not cherry-pick from a collection, we should try to avoid cherry-picking for digitization. Another strategy is to build digital collections using criteria similar to those used to build physical collections. Just as we try to maintain the intellectual integrity of the physical collection and meet the needs of the future
researcher when selecting and building physical collections, we should do the same with digital collections. This approach may reduce the scope of the materials that we digitize, but it would also allow someone interested in that more narrow scope the ability to explore more fully and to use the collection more completely within the digital domain. It would also make additional scholarly and educational services to the digital collection more worthwhile and practical.

Once a body of digital materials has been created, access can be provided in a number of different ways. The most logical is to provide links to digital objects from catalog records or finding aids. The benefit of this approach is that no new system is needed to create and maintain metadata. The drawback is that the services that can be provided in the digital domain are strictly limited. Without a formal management system for the digital materials, it is also difficult to track and maintain them effectively.

Maintaining such a formal management system for digitized materials is going to become very important for collections with significant bodies of digital assets. Digital management and preservation services are needed to ensure the continuing viability of digital collections. Another important benefit of these services is that digital materials can be used very flexibly in a variety of contexts, such as exhibits or online publications.

A clear example of the use of a digital repository in different contexts is provided by the California Digital Library (CDL). CDL maintains two distinct sites for its digital collections: Calisphere\(^2\) and the Online Archive of California (OAC).\(^3\) The same repository, containing the same body of digital content, drives both of these Web sites, but they provide two different modes of access. Calisphere is designed for the general public and K–12 students and teachers. It is a digital repository with an easy-to-use interface. OAC is more research oriented, and, by providing access via archival collection finding aids, it is clearly designed for scholars and others engaged in research. Thus, the same collection can be made available to different audiences with different needs.

If digitization is a first step, and access services are a second step, then, increasingly, the solicitation and use of user feedback is a third possible step in using digital technology to open up and enhance use of special collections materials. A number of user-oriented digital services are already becoming ubiquitous, including blogs, wikis, and “tagging.” Blogs allow users to publish feedback on material they have examined, and for that feedback to become the basis for ongoing discussion. Wikis

---

also allow users of collections to comment but in an environment in which editing of each others’ work is possible. Using wiki technology, a commentary, transcription, or analysis can be considered and edited by a variety of people. Tagging—in other words, the addition of keywords (or informal index terms)—by users is another technique by which users of collections provide feedback, creating new content related to the materials being used.

Although these technologies provide feedback of some sort from users of collections, they mostly exist separately from the collections under discussion. They are a layered commentary for the most part—the “feedback loop” is left open and continuous. For material that appears in blogs and wikis to be permanently related to the source collections, new linking metadata would have to be created by the collection curators. Similarly, “tags” created by users would have to be vetted by a curator to justify its incorporation into the collection Web site or finding aid.4

It is within this new environment of increased user participation that special collections are digitizing their collections and building services, and it is with this context in mind that I will survey some of the digitization activities and digital projects that are being undertaken at UCLA.

Web Services

The term “Web services” probably has a rather diffuse meaning for most readers. We can think of it as “services that are provided over the Web,” which, of course, are almost anything that was previously available via a number of other avenues. The term “Web Services” (and the capitalization is intentional) actually has a very specific technical meaning. The World Wide Web Consortium (the international standards organization for the Web) defines a Web Service as “a software system designed to support interoperable machine-to-machine interaction over a network.”5 It is in essence a formal way in which two computer systems can communicate with each other. For example, the usernames and passwords for an organization might be on one computer system, and the application or resource that users need to use may be on another system. When a user logs into a resource, that computer system sends a message—typically in an XML format—to the accounts management system requesting verification of the user’s credentials. That system will, in turn, send an XML reply verifying (or not) the user’s access to the resource.

4. These are all types of “social software” that serve in various ways to provide communication, accumulate opinions and commentary, and foster new content in a community-based environment. Well-known sites such as Facebook (personal networks), Flickr (image sharing), Delicious (social bookmarking), and YouTube (video sharing) are other examples.
Because the messages being exchanged are in XML, the two computers can be running different software on different operating systems.

Two projects at UCLA use Web Services software to expose and use digital content hosted by the Digital Library: Hypercities and the UCLA Encyclopedia of Egyptology. Hypercities is a “[W]eb-based learning platform that connects geographical locations with the stories and histories of the people who live there and those who have lived there in the past.”6 Among the primary materials being used to build this learning environment are maps and plans of cities created at various times during their histories and documentary material from a variety of archives and libraries. The application will allow users to explore cities via these maps, through space and time, along with photographs, oral histories, and other documents that are bound to a particular time and place. Although this may sound a little like “virtual reality,” it is important to make the distinction between this platform for the exploration of a place through concrete documentary evidence and the reconstructed environments that are created through virtual reality techniques. Hypercities aims to provide an environment in which both exploration and authorship are possible.

Hypercities is best envisaged as a series of layers. The base layer consists of digitized maps from various times during the city’s past, upon which are layered other digitized materials, including photographs and other archival materials and oral histories. This second layer can be hosted locally or populated through links to other digital archives. A third layer consists of user-supplied content of various types, including uploaded content and links to related material. User groups could be individual historians, groups collaborating on a project, or students in a class. “Through collaboration between universities and community partners in Los Angeles, Lima, Berlin, Rome, Tel Aviv, New York, and other places, HyperCities will develop and offer a participatory, open-ended learning environment grounded in space and time, place and history, memory and social interaction, oral history and digital media.”7

In the first instance, the Hypercities team is building a generic environment that can be used to explore a variety of different cities. Los Angeles will be the first, followed by Lima, Rome, Berlin, New York, and Tel Aviv. In each case, historical maps will be used to form the foundation layer for navigation through both space and time, and dynamic links will be made from points in space and time to relevant documentary evidence. The documents, photographs, oral histories, and other parts of the second layer could be hosted by the Hypercities application, but,

---

where possible, the digital objects referred to will be drawn dynamically from a variety of other digital repositories. For *Hypercities Los Angeles*, both the UCLA Digital Library⁸ and the University of Southern California Digital Archive⁹ will provide content. UCLA materials accessible through the *Hypercities* platform will include photographs from the Department of Special Collections, oral histories from the Center for Oral History Research, and maps from the Bruman Map Collection.

The diagram below illustrates how the *Hypercities* application uses digitized content from both UCLA and the University of Southern California (USC). A user would be interacting with the *Hypercities* application via the Web in the manner in which we are accustomed. However, behind the scenes, the application has access to both the UCLA and USC collections, from which real-time requests for content can be made. For instance, a user may be studying the history of a specific neighborhood of Los Angeles. Historical maps provide a way of browsing geographically, and a timeline provides a means of browsing through time. As the user moves the focus across maps and moves the temporal focus through time, links to content at UCLA and USC appear in the appropriate places. When the user sees something interesting, he clicks the link and the browser focus changes to the appropriate archive and the digital object of interest, as illustrated by the HTTP links in the diagram below.

---

The UCLA Encyclopedia of Egyptology (UEE) uses similar technologies to make use of digital special collections materials in support of scholarship; in this case, creating an online scholarly encyclopedia. As with Hypercities, the Web application—that is, the online Encyclopedia—pulls digital content from the UCLA Digital Library as needed. The Digital Library stores the content of the Encyclopedia as individual elements; it is the application that draws them together and adds the specific navigational and other added services that one expects of an encyclopedia.

A key feature of this project is its innovative approach to data access by making use of hypermedia. Hypermedia provides opportunities for linking information in ways not otherwise possible, enabling it to be seen in new ways and facilitating the creation of new knowledge. Online media provide opportunities that are not available in printed versions. Innovative search functions and data mining that make use of textual, pictorial and, in future, audiovisual formats, have an important heuristic function that by far supercedes that of searches in printed volumes. Rather than only presenting a conveniently accessible body of excellent content, the UEE will ultimately open up new research methods, through combining and interweaving the knowledge of eminent scholars with original data in completely new ways.¹⁰

The content of the Encyclopedia includes the scholarly articles, illustrations, maps, video, and audio. As searches are initiated and as users browse the Encyclopedia site, the Digital Library is queried and responses are formatted to provide the experience that users of an encyclopedia would expect. As is the case with the Hypercities project, individual digital objects are freely available directly through the Digital Library Collection System, but as individual items lacking the context of the Encyclopedia. In addition, the encyclopedia articles are being made freely available in an “Open Version” via the University of California’s eScholarship repository in the PDF format.¹¹ The complete Encyclopedia, with its additional services and context, will be a licensed product. This dual approach to access—open access to the digitized content and scholarly articles, as well as an online publication that provides added value through navigation, context, and additional services—was a key component for the involvement of the UCLA Library. It allows us to adhere to the principle of open access to our collections while supporting a publication that needs an income stream for its continuing viability.

The following diagram shows the relationship between the open and the licensed versions of the Encyclopedia. A user could potentially find most of the content of the publication in either the UCLA Digital Library or UC eScholarship Reposi-

---

tory, but the complete publication, which would bring textual and visual content together in context and provide navigation, linking, and other services, would only be available in the licensed version.

The other component shown here is the University’s Digital Preservation Repository (DPR). This is a repository available to all UC campuses for the long-term preservation of their digital content. Digital objects are deposited using the Metadata Encoding and Transmission Standard (METS), which is the standard format for the exchange of digital objects between repositories. Clearly descriptive metadata plays an important role: it is the basis for the Web Services links between the Encyclopedia and the Digital Library and is encoded and stored as an integral part of the digital objects deposited in the DPR. Without good metadata, applications such as Hypercities and UEE would not be possible. Adequate metadata is also necessary to manage the relatively expensive task of long-term digital preservation.

**Metadata to Support Digital Services**

The two forms of metadata usually created in special collections are finding aids for archival and manuscript collections and catalog records for published materials. These two forms of description are very different, both in format and content. Catalog records typically have the following attributes:

- A record usually describes a single item, potentially in very great detail. Although a catalog record can vary in its level of detail, ideally all items in the catalog would be described at a similar level of detail to provide uniform access to all materials.

- There are limited mechanisms for dynamic linking to other items, either those related hierarchically (part of...; contains...) or as sibling terms.
In general, finding aids have the following attributes:

- The primary organizing principles are intellectual (for example, works, related works, and projects) and physical (such as correspondence, maps, and plans) properties and relationships.

- Related items are gathered together and listed as series and sub-series, and so on, sometimes in complex structures that are many layers deep.

- Some collections warrant description in great detail, whereas others are described in a summary fashion. It is clearly impractical to describe every collection at the same depth, especially in large collections.12

Metadata for codices and other valuable manuscripts tend to fall outside this dichotomy. Due to their high value—both intellectual and monetary—they are often afforded extensive descriptive treatment, perhaps extending to a monograph in length. The description of codices and antiquarian books generally includes discussion of the physical object as well as its intellectual content, and it is hard to generalize standard practice, or to squeeze the descriptive content into the elements used for the description of digital surrogates. Codicological studies are hard to reduce to the MARC13 standard of the library catalog or the EAD/DACS14 standards of archival materials, or by extension to standards used in digital collections, such as Dublin Core and MODS.15

Digitized library collections are most likely to come from archival and special collections, so the source of metadata is often catalog records, finding aids, and codicological descriptions. To provide for a rich online experience with these materials, it is desirable to make the type and depth of description as uniform as possible. If this is not the case, those materials with the richest description are more likely to be found by users, thus biasing use of the collections in their favor.

12. The reasons for deciding on a given level of description can vary a great deal. In some cases, the importance of the collection may be a factor. In other cases, there may not be resources available for the research that may be needed. It may also be judged that the likely users of the collection do not need detailed descriptions to meet their needs.


15. The Dublin Core Metadata Initiative is an organization developing interoperable metadata standards to support a broad range of purposes and business models. The Metadata Object Description Schema (MODS) is a bibliographic elements set designed for general use but particularly for library applications. See http://dublincore.org/ and www.loc.gov/standards/mods/ for more information (Accessed March 3, 2009).
The user is best served if descriptions are uniform—as is ideal in a library catalog—especially given that online exposure will allow for use of materials in innovative ways. Some simple and obvious examples include digitization of large numbers of manuscripts that will allow for direct comparative study of manuscripts from widely diverse traditions, perhaps resulting in unforeseeable insights; digitized images from archival collections—again, from widely divergent traditions—that could be the inspiration for new artistic or speculative creativity; and archival materials that have historically been of interest to one group of scholars (for instance, sheet music and musicologists) might now be used as evidence by an entirely different group (for example, sheet music illustrations and social historians).

Metadata standards and systems for digital library projects have elements in common with both the traditional library catalog record and with finding aids. Most assume that digital items will be described at the item level. They do, however, include structural and relational metadata that provides for the description of complex digital objects, and for the definition of relationships between objects. Creating metadata for digital library purposes is a time-consuming and expensive process, often involving repurposing and enhancing metadata to meet the needs of the online environment and the services we wish to provide our users.

Online use of our collections is likely to be much broader in scope than is the case in the traditional special collections reading room—involving more divergent use and a broader range of users—and our metadata needs to be rich enough to meet those needs, including users who are new to us. The example of digitized sheet music is instructive. Whereas sheet music description has usually catered to the needs of the musician by focusing on musical attributes, digitized sheet music collections are being used by a wider range of scholars, some of whom are much more interested in the social aspects of sheet music publication. Also, as noted above, users may be interested in sheet music cover artwork completely independently of the music itself. This expansion of the user base will obviously be even more significant as we acknowledge uses beyond traditional scholarship. For instance, digitized collections are widely used by K–12 teachers, as well as by the general public for any number of other reasons.

Managing metadata in special collections has become very complex. We are managing at least four different systems: library cataloging, finding aids, detailed descriptions of codices and other antiquarian materials, and now metadata for our digitized collections. As each has its own internal standards, it is important that we have coherent standards for mapping these data from one to another, as well as well-defined rules for supplementing metadata as needed. Ideally, we would like all these different types of description to coexist through sophisticated mapping tech-
niques that will allow us to layer new services over them, services that will make up a “virtual research environment.”

Conclusion: Virtual Research Environments

*Hypercities* and the *UCLA Encyclopedia of Egyptology* are just two possible exemplars for systems that support online research and instruction. Each relies on dynamic access to special collections materials, along with a suite of specialized tools and services to enrich the experience of users and to provide them with tools to engage with primary research materials in productive ways. At UCLA, we use the term “Virtual Research Environment” (VRE) to describe the direction in which we hope to move the Digital Library. The *Hypercities* and *UEE* both have elements of a “virtual research environment.” *Hypercities* provides the ability for users to add content, to annotate, and to collaborate with other users. The *Encyclopedia* is also collaborative, but, in many respects, it is a traditional online publication. It does, however, use the UCLA Digital Library as a dynamic source for its content, content that is available through the Digital Library portal and potentially in other contexts as well. A generic VRE would contain many of the elements of these two projects, with additional tools that would allow users to undertake personal and collaborative investigations.

In many ways, this new type of research environment will be a generalization of some of the tools being developed for projects such as *Hypercities* and *UEE*. It is an environment in which users of our collections have access to large quantities of digitized material using a suite of tools that allow them to undertake research efficiently and collaboratively. Examples of the activities we aim to support include:

- Searching, retrieving, viewing, and manipulating digitized materials across collections in various institutions.
- Saving, organizing, annotating, and sharing digital documents from a secure online environment.
- Incorporating collaborative tools that allow groups of researchers or students to undertake joint investigations or instruction.

There are many existing tools—such as wikis and blogs—that facilitate these activities, but they are mostly used one tool at a time and don’t necessarily integrate well. A fully functioning VRE would allow users to engage meaningfully and productively with all types of archival materials—manuscripts, photographs, antiquarian books,

---

16. The term is also the title of a program of the Joint Information Systems Committee (JISC) in the United Kingdom. See [www.jisc.ac.uk/whatwedo/programmes/vre2.aspx](http://www.jisc.ac.uk/whatwedo/programmes/vre2.aspx) for more information (Accessed March 3, 2009).
sound and moving images materials, maps and plans, artwork of various types—and use them to engage in a wide range of scholarly activities, both solo and in concert with collaborators worldwide. Users should be able to jointly examine and manipulate objects, experiment with them, and exchange ideas in a number of different formats. Ultimately, one hopes that online publications would result, with links back to original evidential content, as well as to annotated or otherwise manipulated or derivative digital content that would illustrate the arguments being made.

An effective research environment needs to do a lot more than simply provide access to digitized content, although that is an important function and one that may meet the needs of many users. But, as the next few generations of scholars start using our collections, we will find that they are increasingly comfortable working online and also working collaboratively. They will certainly expect our collections to be available online, but they will want much more than that. They will expect to be able to work with the materials in the online domain much as they do in the physical domain. Tasks such as taking notes, annotations, sharing, discussing, organizing, writing, and publishing will all be possible in the digital domain. We, as special collections professionals, should be prepared to meet those needs.
50th Annual RBMS Preconference

Seas of Change: Navigating the Cultural and Institutional Contexts of Special Collections

June 17-20, 2009 | Charlottesville, Virginia

With major sponsorship from the University of Virginia Library and the Antiquarian Booksellers Association of America

As the 50th anniversary of the RBMS preconference, this year represents an important moment in the history of RBMS and its affiliated professions. In addition to celebrating our achievements, we will also look broadly at how special collections librarianship has evolved over the past half century with respect to changes in social, cultural, technological, economic, and academic environments, and – more importantly – how we will need to respond to such changes in the future.

A variety of workshops, seminars, short papers, discussion topics, tours, receptions and a booksellers’ showcase will complement the main program. For complete schedule, registration, and housing information, please visit:

http://rbms.info

Image of the H.M.S. Beagle from an original watercolor by Conrad Martens in the PaulVictorius Evolution Collection, courtesy of Special Collections, University of Virginia Library