Digital Library Technology Trends

Sun microsystems

PREFACE

This white paper is intended to describe the technology trends in digital libraries, discuss key issues involved in digital library implementation, and provide profiles of some of today's leading digital library programs. As one of the leading technology providers to libraries, museums, and the educational community, Sun Microsystems has played a key role in the evolution of digital libraries around the world and is pleased to support their continued development.

Sun Microsystems will continue to support the library and museum communities as they explore new and evolving information technology (IT) architectures. Sun is also committed to fostering the sharing of best practices among institutions and increasing the cross-fertilization of commercial and library/museum-specific solutions.

We hope that this white paper — together with the Digital Library Toolkit, which is available on the www.sun.com/edu/libraries website — will offer guidance and useful information to our Education, Library, and Museum customers.

Thank you for your continued support!

Art Pasquinelli, Manager
Global Education and Research
Sun Microsystems, Inc.
art.pasquinelli@sun.com

August, 2002

INTRODUCTION

Unlike previous eras, ours is an age of lifelong learning. Society expects that people will continually gain new skills and knowledge. Economically, continuing education is critical for one's ability to maintain earnings potential. Sociologically, education has become increasingly important as the core of professional and personal success. Evening degree programs are making way for Internet universities as students from all walks of life seek more flexible ways to learn.

The library, historically a cornerstone of scholarly endeavor, is reinventing itself in today's networked society to meet these new demands. Instead of a building that holds books, the library is evolving into an electronic portal to a growing global collection of digital content. The doors to this virtual library are now open 24 hours a day, seven days a week, and the library's holdings come to the user when needed. Today's library includes sophisticated tools that make it easy to find the best information resources, delivering them to one's desktop or mobile computing device at the push of a button.

While there are still many challenges to realizing the potential of digital information, digital library technologies and practices have developed enough so they are within reach of every type and size of educational institution. To that end, this publication is intended to provide executives of learning organizations with a vision from Sun Microsystems of what the digital library can be, practical direction on how to approach development of a digital library, and insight into what library industry leaders have learned from their experiences.

Cheryl D. Wright
Digital Library Consultant
cheryl@cherylwright.net

Part 1 – What is a Digital Library?

Evolution of Digital Collections

One of the most precious treasures of the British Library is the irreplaceable 11th century manuscript of "Beowulf," an epic poem describing the heroic adventures of a Scandinavian warrior. The manuscript was donated to the British Museum in 1700 and was nearly destroyed in a fire 30 years later. With scorched edges and brittle pages, the manuscript continued to deteriorate over the next century of handling by scholars. Although the degradation was halted in the mid-1800s by mounting each manuscript leaf in a protective frame, much damage had already been sustained and sections of this great work can now only be known through historical transcripts.

In 1993, the British Library initiated the "Electronic Beowulf Project" to capture, enhance, and preserve forever this cultural artifact in digital form. Not only has the manuscript been captured in its current form, but it is now available for study anywhere. This is only one example of how, throughout the world, libraries, museums and archives are digitizing the important documents and images of our culture, both to preserve them for future generations and to make them more accessible to our own.

As digital imaging technology has advanced, so have opportunities for preservation of irreplaceable artifacts. The Dead Sea Scrolls, written as early as the second century B.C. and discovered fifty years ago, were considerably degraded by their centuries of exposure to the elements. Conditions in the caves where they were discovered had caused the parchment scrolls to deteriorate, obscuring the text and causing parts of the writing to transfer to adjacent parts of the scroll. A specialized camera and filters developed by Eastman Kodak for the space program were used to image fragments of the scrolls and reveal new characters previously invisible in normal light, allowing researchers to reconstruct new parts of the text.

The barriers of space and time in the search for knowledge are also being eliminated. Today, students, researchers, information professionals, and the general public can directly access many of the world's rarest artifacts—from high-quality images of each page of the Gutenberg Bible to a digital likeness of the Mona Lisa—right at the desktop or other Web-enabled device, at any time and from any location. Audio recordings of historic speeches or exotic birdcalls, video clippings from televised news programs, geospacial data, and more are being delivered directly to the desks of students and researchers. Thanks to the creation of digital libraries, scholarly research is accelerating dramatically without the limitations of physical access.

Through the 1990s, digital library projects were largely experimental activities. As shown in Figure 1, many important advances in digital library techniques came about through research sponsored by the U.S. National Science Foundation (NSF) and the U.K. Joint Information Systems Committee (JISC). In 1999 these projects began expanding internationally when NSF linked its digital library research program with similar activities being undertaken by JISC, resulting in the JISC-NSF International Digital Library Initiative. The objectives of this three-year program were to:

- Assemble collections of information that were not otherwise accessible or usable because of technical barriers, distance, size, system fragmentation, or other limits.
- Create new technology and the understanding to make it possible for a distributed set of users to find, deliver, and exploit such information.
- Evaluate the impact of this new technology and its international benefits.

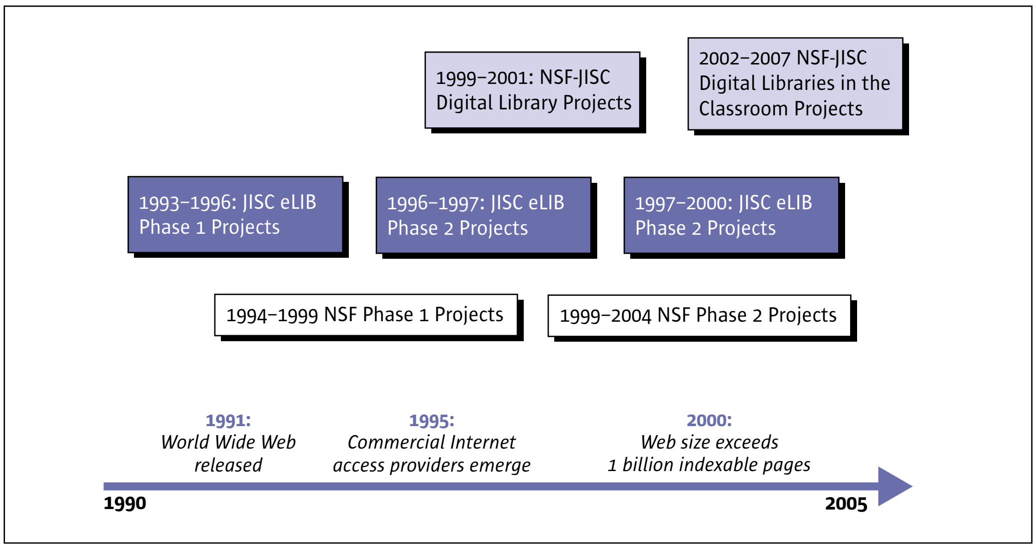

Figure 1. Key U.S./U.K. Digital Library Research Initiatives

Since then, many other groups have become involved in the expansion of digital library technologies and techniques, including the European Union, Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), the International Federation of Library Associations (IFLA), the American Library Association (ALA), the Coalition for Networked Information (CNI), and the Digital Library Federation (DLF).

Library automation software vendors have also introduced the first commercial products developed from the experimental work. These products are aimed at educational institutions that seek easy-to-install and easy-to-maintain solutions based on industry standards and best practices to date.

Defining the Digital Library

Debate rages among librarians and knowledge professionals as to what constitutes a digital library. There are many definitions, ranging from the electronic catalog that describes physical items in a "brick and mortar" library to advanced multimedia environments housing all-digital collections. H. Thomas Hickerson, Cornell University's Associate University Librarian for Information Technologies & Special Collections, believes it is time to erase the line between physical and digital libraries. "A major portion of library activities are technology-supported and have been for years. The Internet has had an incredible impact, but libraries have a history of managing large systems and using technology to deliver bibliographic information," Hickerson says.

Still, there has been tremendous progress in adapting new technologies to library use over the past decade. Bernie Hurley, the Director for Library Technologies at U.C. Berkeley, draws the distinction between traditional library automation and digital libraries. Hurley notes that "digital libraries are different in that they are designed to support the creation, maintenance, management, access to, and preservation of digital content." He sees the work by U.C. Berkeley and other large research libraries as the first generation of digital library development. "Once the specifications settle out, the commercial vendors can more successfully enter the market with competitive products," Hurley says.

Sun Microsystems defines a digital library as *the electronic extension of functions users typically perform and the resources they access in a traditional library*. These information resources can be translated into digital form, stored in multimedia repositories, and made available through Web-based services. The emergence of the digital library mirrors the growth of e-learning (or distance learning) as the virtual alternative to traditional school attendance. As the student population increasingly turns to off-campus alternatives for lifelong learning, the library must evolve to fit this new educational paradigm or become obsolete as students search for other ways to conveniently locate information resources anywhere, any time.

Goals

Digital libraries began to appear on the campus in the early 1990s as research and development projects centered within computer science departments, sometimes funded by government grants. Campus librarians were often uninvolved in these early projects, which focused on digitization technology, metadata schemes, data management techniques, and digital preservation. Dan Greenstein, former director of the Digital Library Federation and now head of the California Digital Library, characterizes this experimental period as the "second-generation digital library," exploring new opportunities and developing new competencies. With this new knowledge in hand, Greenstein suggests that the third-generation digital library abandoned this experimentation and the "build it and they will come" philosophy that characterized early digital collections, focusing instead on fully integrating digital material into the library's collections through a modular systems architecture. "This modular approach is fundamentally liberating since it permits libraries to think creatively about how to build upon services supplied by others."

As these projects matured and led the way toward more practical digital library implementations, the library began to take a more central role. Digital library use shifted to a large and diverse campus audience, and information technology (IT) groups began to partner with the library to develop campus-wide standards for the deployment and operation of digital libraries as an integral part of the education enterprise. This development paralleled the development of heightened student requirements for access to library resources.

With the advent of the Internet, individuals' expectations for access to information have increased dramatically. It is no longer considered practical or acceptable to travel to a specific location during certain hours to locate needed information. Library patrons are not satisfied to locate an item of interest that is housed at yet another physical location, request the item, and then wait days or weeks for the item to arrive at the building where it was requested. Patrons increasingly expect instant access to all the information resources they require, from any location, at any time, and from any device. This is the objective that the digital library is fulfilling.

With digital libraries, an individual can:

- Gain access to the holdings of libraries worldwide through automated catalogs.
- Locate both physical and digitized versions of scholarly articles and books.
- Optimize searches, simultaneously search the Internet, commercial databases, and library collections.
- Save search results and conduct additional processing to narrow or qualify results.
- From search results, click through to access the digitized content or locate additional items of interest.

All of these capabilities are available from the desktop or other Web-enabled device such as a personal digital assistant or cellular telephone. Additionally, the user can customize his or her information request so that search results reflect individual needs and preferences. Sun considers personalization the next killer application, creating a more valuable and richer user experience in the digital library environment.

Key Components

As shown in figure 2, a fully developed digital library environment involves the following elements:

1. Initial conversion of content from physical to digital form.
2. The extraction or creation of metadata or indexing information describing the content to facilitate searching and discovery, as well as administrative and structural metadata to assist in object viewing, management, and preservation.
3. Storage of digital content and metadata in an appropriate multimedia repository. The repository will include rights management capabilities to enforce intellectual property rights, if required. E-commerce functionality may also be present if needed to handle accounting and billing.
4. Client services for the browser, including repository querying and workflow.
5. Content delivery via file transfer or streaming media.
6. Patron access through a browser or dedicated client.
7. A private or public network.

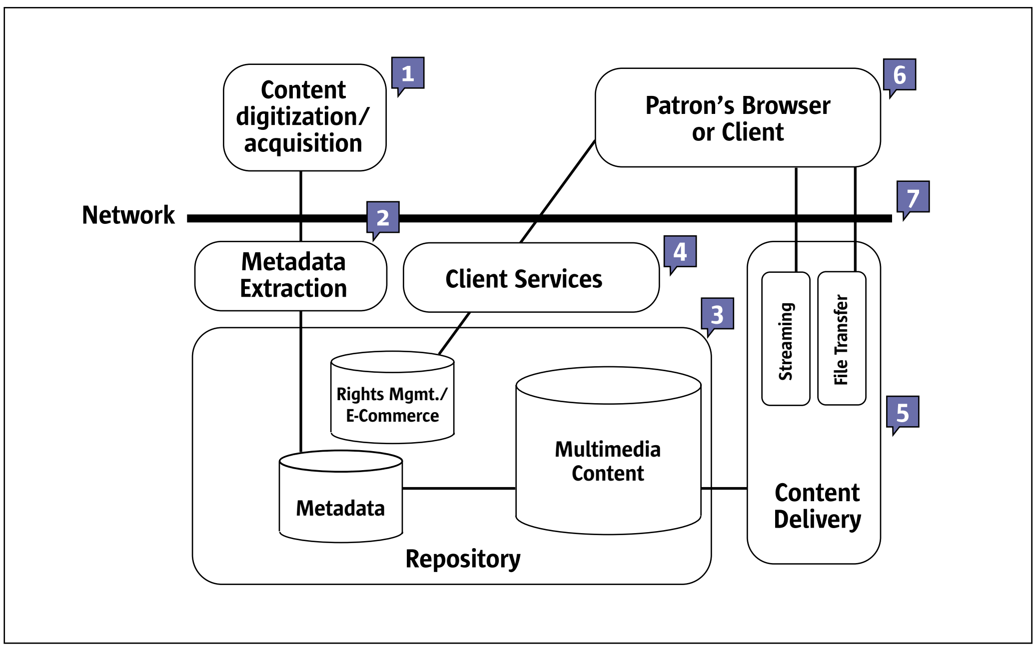

Figure 2. Functional Components of a Digital Library

These components might not all be part of a discrete digital library system, but could be provided by other related or multi-purpose systems or environments. Accordingly, integration is a consistent issue cited by digital library developers.

To interoperate with the existing library infrastructure, the digital library must be designed to work with existing library catalogs and incorporate industry standards, formats, and protocols. The term "digital library" is often used to describe any multimedia management system holding digitized information, but this does not mean it will deliver true library application functionality. Thus, these digital library components must also be tailored to capture, encode, and deliver information according to the standard practices adopted by the library industry. Because of the rapid pace of technological change, some standards are concrete and others are emerging.

Link to e-Learning and the Campus Enterprise

The development of digital libraries must be considered in the overall context of initiatives to unify the IT structures of the campus and to transform the learning process through innovative technology.

Economic, social, and cultural pressures are forcing schools and universities to reinvent themselves. As in the business process re-engineering activities of the last decade that transformed corporate enterprises, education organizations are now viewing themselves in a new light. New types of students and changing student expectations are driving the integration of core campus functions and deployment of student services on the Web. Fragmented, monolithic approaches are falling away as educators realize the need to link learning and administrative resources in a more effective way to become a "knowledge enterprise," the 21st century version of the traditional campus.

During the past decade steep declines in the cost of commodity components, combined with the availability of high bandwidth networks, have made sophisticated IT applications for education affordable. A mix of sophisticated digital and Internet-based services and rapidly expanding global digital content have made possible a "virtual learning environment" (VLE) that delivers the capability to enhance the classroom experience or conduct learning apart from a physical campus. The digital library is a core component of this VLE.

"These developments are extending the role of the library, and changing the relationships between the library and other parts of the academic enterprise," according to Clifford Lynch, Director of the Coalition for Networked Resources and a noted authority on digital libraries. "I think we will see a continued evolution from thinking about digital collections to thinking about networked information services, which will integrate authoring, analysis, and distribution tools that facilitate the reuse and repurposing of digital content. In almost all cases, the collections and services must be integrated into the institutional, national, and worldwide fabric of research and teaching."

The NSF in the United States has joined with the JISC in the United Kingdom to sponsor a major research program called "Digital Libraries and the Classroom: Testbeds for Transforming Teaching and Learning." This program will promote the use of large-scale, distributed digital content and advanced networking technologies for learning. Running from 2002 to 2007, the sponsored projects are expected to demonstrate how integrating recent technical developments with digital content improves the learning experience of students and provides new models for classroom instruction. Evaluating the impact of these projects on student achievement, retention, recruitment, and on institutional structures and practices will be important elements of the initiative.

Part 2 – The Open Digital Library

Technology Platform Requirements

By its nature, digital collection development requires extensive use of technological resources. In the early days of digital library development, when collections were typically small and experimental, a wide variety of hardware and software was utilized. Today, the leading digital library developers are putting substantial collections online. Some of these collections include millions of digital objects; collections are being planned that will require storage measured in petabytes—the equivalent of more than 50,000 desktop computers with 20-gigabyte hard drives.

As digital libraries scale in size and functionality, it is critical for the underlying technology platform to deliver the performance and reliability required. Many digital libraries are considered “mission critical” to the overall institution. In addition, patrons expect high service levels, which means that downtime and poor response time are not tolerable. Moreover, because cost is a foremost concern, scalability and efficiency with a low total cost of ownership are also key requirements. This type of digital library implementation requires a scalable enterprise-level technology solution with built-in reliability, availability, and serviceability (RAS) features.

Storage capacity also must be scalable to adapt to rapid growth in demand, and must be adapted to the mix of media types that may be stored in a digital library, such as:

- Text, which is relatively compact.
- Graphics, which can be data-intensive.
- Audio, which is highly dynamic.
- Video, which is highly dynamic and data intensive.

Storage capacity should be expandable in economical increments and should not require redesign or re-engineering of the system design as requirements grow.

An open systems architecture provides both a robust platform and the best selection of digital media management solutions and development tools. The inherent reliability and scalability of open platforms have made them the most popular choice of IT professionals for Internet computing. This computing model features an architecture that is oriented totally around Internet protocols and stresses the role of Web sites for a vast and diverse array of services that follow a utility model.

The Sun™ Platform

The use of new Internet and portal computing models, the explosion of off-campus, lifelong learners, and the growing need to better manage heterogeneous, round-the-clock IT environments has fueled the move to Sun technologies. Sun Microsystems™

continues to lead the way in RAS, resource management, real-world performance capabilities, and cost of ownership, setting a standard that other technology platform offerings cannot match. Among other achievements:

- Sun's 64-bit UltraSPARC® III processor received MicroDesign Resources (MDR) *Microprocessor Report 2001* Analysts Choice Award in the Best Server/Workstation processor category.
- The Solaris™ Operating Environment was declared the Best Server Operating Environment in 2001 by *Network Computing*; it was also ranked as the top UNIX® Operating System by industry analyst firm D.H.Brown Associates.
- In UNIX markets, Sun leads its competitors in both unit shipments and revenue in the entry, midrange, and high-end server segments.

The Solaris™ Operating Environment includes many features that deliver the scalability and RAS required for the digital library environment. Near-linear performance increases are achieved as processors are added to servers, and high-reliability clustering is supported for extremely high uptime environments. According to the results of the *2001 Unix Function Review* performed by D. H. Brown Associates, Solaris 8 Operating Environment is the highest rated operating system among five leading UNIX-based operating systems.

Sun midrange and high-end servers also offer high-availability features such as full hardware redundancy, fault-isolated dynamic system domains, concurrent maintenance, and clustering support. Exceptional performance is achieved through fast processors, outstanding system interconnect performance, fast I/O, and linear scalability.

Sun also offers one of the industry's most economical storage environments, providing modular storage capacity that can be added incrementally. Sun's flexible storage products enable the user to purchase precisely the increments needed. Spindle limits, capacity, and bandwidth can be increased with a simple upgrade. Sun's enterprise storage products are designed to have no single point of failure. For example, Sun StorEdge™ arrays incorporate hot-swappable, redundant components and dual controllers, providing built-in backup systems and easy serviceability.

Total investment protection is achieved through Sun's reliance on a common chip architecture, a common operating environment, and complete binary compatibility. The same application can scale from a single CPU to hundreds of CPUs, spanning generations of product releases without the need for modifications or recompilations.

The Education Portal

Increasingly, IT specialists are implementing campus portals. In a portal environment, the client connects to a portal server via a browser or other device. The portal server presents an authentication screen to the user (whether student, faculty, or staff). With the connection running https, all username and password information is encrypted in transport. If the user name and password is accepted and verified within the directory services, the portal server can then read all dynamic attributes for the customer from the directory services and start to build the primary screen. This screen will include all services and applications that the customer has permissions to access —including library services. Meanwhile, the portal server establishes connections to all the back-end application servers while waiting for the customer to choose where to go next. All of this functionality can create a management nightmare, which is why it is important to support a scalable management and policy base mechanism.

The Sun™ Open Net Environment (Sun ONE) product set (see Appendix) provides this functionality by administering access based on community. It ensures that student information stays with students (via a Student Portal), that faculty sees only applications for the faculty (via a Faculty Portal), and that administrative personnel view only relevant applications and content (via an Administrative Portal). Each portal is customized and tailored to the community. Once the portal server has been configured initially, whenever an application or service needs to be added it is simply a matter of adding the resources, having the services run in the backend, and then adding the access control entries to the portal. The next time the user connects, the service will be available. This provides the flexibility to dynamically modify the customer environment, supporting thousands to tens of thousands of devices or more.

In addition, an easily integrated Sun ONE product set provides modular directory and security services, e-commerce functionality, and campus-wide messaging and calendar services for a total campus infrastructure solution.

Evolution to Web Services

Software is evolving from local desktop and client-server software applications to Web applications. Today, it is very common for organizations to move their local applications to the Web, and in some cases to subscribe to additional application services over the Web from external service providers. The next phase of this evolution is the development of true Web services. A Web application offers a service that requires the intervention of a user, while a Web service facilitates direct program-to-program interaction without user intervention. Web services are functional components that can combine with other services dynamically to create a more useful or powerful application.

The digital library of the future will deliver "smart media services"; that is, Web services that can match "media content" to user "context" in a way that provides a customized, personalized experience. Media content is digital content that includes elements of interactivity. Context includes such information as the identity and location of the user.

Several key technologies must interact to allow Web services to work in this way. Extensible Markup Language (XML) and Standard General Markup Language (SGML) are important standards influencing our ability to create broadly interoperable Web-based applications. SGML is an international standard for text markup systems and is very large and complex, describing thousands of different document types in many fields of human activity. XML is a standard for describing other languages, written in SGML. XML allows the design of customized markup languages for limitless types of documents, providing a very flexible and simple way to write Web-based applications. This differs from HTML, which is a single, predefined markup language and can be considered one application of XML. The primary standards powering Web services today are XML-based. These include:

- Simple Object Access Protocol (SOAP).
- Universal Description, Discovery and Integration (UDDI).
- Web Services Description Language (WSDL).
- Electronic Business XML (ebXML).

These standards are emerging as the basis for the new Web services model. While not all are fully defined standards, they are maturing quickly with broad industry support.

Part 3 – Digital Library Development

An educational institution's digital library environment will likely encompass both local collections and externally provided resources from such sources as subscription services and other libraries. There is often a desire to structure this blend of internal and external resources as a unified collection from the user's point of view, with a single gateway and a comprehensive set of support services. Some resources are "born digital" while others originate in physical form and require transformation into digital form, a process that can be difficult and costly. Most of today's digital library environments are the result of custom development, but as packaged library solutions are now available, the next group of digital library adopters must reach a "make versus buy" decision. Much of the utility and long-term ease of maintenance of digital collections is dependent upon the types and quality of metadata that is stored with the digital objects, but agreed-upon standards for metadata are only beginning to emerge.

The first organizations to build large-scale digital libraries encountered many challenges in organizing, funding, managing, and ensuring the long-term maintenance of the digital library environment. The objective of this section is to discuss some of these challenges and share experiences of the early adopters. Some of these experiences may help others avoid pitfalls and take advantage of evolving best practices.

Funding

The funding issue is probably the single largest barrier to digital library development; this is why most existing collections are essentially pilot projects or R&D activities. The challenge of mounting a large-scale digitization effort, encoding appropriate metadata to ensure ease of discovery and use, and making a commitment to long-term preservation can be daunting to libraries already struggling to pay for existing activities.

Some institutions like Harvard University have been able to generate start-up funding with the philosophy of "build it and they will come." Most will have to shift funding from existing programs or activities—never an easy task. "It's hard to change what you're doing when you have to achieve that change by moving existing funds from one spot to another," comments Cornell University's H. Thomas Hickerson. "Not all library leaders have been willing to do that." Cornell has sought initial seed money for digitization projects, but establishes cost recovery models that will eventually allow the digital programs to become self-funding.

Custom Development vs. Packaged Solutions

Custom development has been the predominant means of constructing digital libraries simply because off-the-shelf solutions have only become available relatively recently. Leading digital library implementers who have gone this route consistently cite the lack of commercially available solutions that met their requirements as the reason for developing custom systems.

There are both benefits and drawbacks to doing custom development. On the plus side, there is a high degree of assurance that the completed system will closely fit the specific user requirements, since there are (theoretically, at least) no competing communities with different needs or opinions about the best way of accomplishing a certain task. The up-front cost of development can be spread over a long period of use, with no further license fees or charges to be paid as time goes on. Future development is completely controllable and predictable, and there is no doubt about continued availability of the system. On the other hand, the cost of developing and maintaining a complex multimedia management environment is very high. Says a veteran digital library implementer, commenting on one particularly well-funded project: “Even a million dollars doesn’t go very far in developing this kind of a system.” Another experienced digital library developer says, “If you’re going to do it by yourself, bring your checkbook.” There is also the threat that evolving standards would result in one-off systems that lack interoperability with other digital environments.

The school of thought advocating the development of commercial digital library solutions revolves around the value of a large number of buyers supporting the research and development required to bring a robust “turn-key” solution to market. The nature of the marketplace will dictate that the system for sale will meet a majority of the buyers’ needs at a more reasonable price than developing unique systems. The continuing force of the market will exert pressure for improvement and evolution of the product to continue to meet users’ requirements, protecting the buyers’ long-term interests. Many libraries simply do not have access to the types of technologists required to develop and maintain these systems, nor do they have any interest in staffing to maintain a complex digital library system. These arguments in favor of commercial systems have motivated many would-be digital librarians to wait... but the wait may be over.

At least four library automation companies now offer digital library solutions, and other general multimedia management technology companies are starting to target the educational institution/library market. Stanford University Librarian Michael Keller’s advice is: “Don’t reinvent the wheel. Don’t write code or try to be leading edge unless you have strong support from your computer science department. Buy off the shelf from reliable suppliers.”

The following section introduces several commercial entries from Sun software partners in this emerging digital library software market.

Packaged Solutions

These products are available from established library automation vendors:

Endeavor Information Systems, the vendor of the integrated library system "Voyager," has introduced its digital library offering "ENCompass." Endeavor involved a small group of customers as product co-developers. Cornell University, supported by a hardware grant from Sun Microsystems, Inc., became the first co-developer. Other co-developers now include the Getty Research Institute, Kansas State University, and the University of Pennsylvania. The first commercial release of ENCompass was in March 2001. Endeavor is currently working on the third release of this product.

Endeavor describes ENCompass functionality as including:

- Search and discovery.
- Object management.
- Collection management.
- License and rights management.
- Linking.

Information about ENCompass products can be found at Endeavor's Web site, www.endinfosys.com.

Ex Libris, provider of the integrated library system "ALEPH," has entered the digital library market with "DigiTool." Ex Libris has also taken the approach of involving users. In October, 2001 the University of Maryland Libraries was announced as a Premier Partner to further development of the DigiTool digital asset management software. This Ex Libris program actively fosters participation in the ongoing design, development, and testing of new features within DigiTool.

Ex Libris describes DigiTool as a means of ensuring that digital data can be acquired, manipulated, shared, searched, and distributed in conformance with industry standards:

- Data structures such as Dublin Core (DC), MARC21, MAB, Text Encoding Initiative (TEI), and Electronic Archive Description (EAD). Future support is planned for Visual Resources Association (VRA), Computerized Interchange of Museum Information (CIMI), and Records Export for Art and Cultural Heritage (REACH).
- Data content of pages in TIFF, GIF, JPEG, or PDF formats; text in SGML, HTML, or XML formats.
- Data interchange via Z39.50, SQL, HTTP, Dienst, and Open Archive Initiative.
- Information industry syntaxes (SGML, HTML, XML, and Resource Description Framework [RDF]).

DigiTool product information can be found at Ex Libris' Web site, www.exlibris.co.il or www.exlibris-usa.com.

Sirsi, vendor of the "Unicorn" integrated library system, has a digital library offering called "Hyperion™ Digital Media Archive." The Hyperion Digital Media Archive is a tool for building, storing, and maintaining collections of digitally captured material—most often scanned documents, photographs, maps, drawings, artwork, sound, and movie clips. Hyperion can also serve as a repository for computer-generated material such as technical documents, multimedia presentations, and engineering drawings. More information on Sirsi's Hyperion Digital Media Archive can be found at www.sirsi.com.

VTLS, developer of the "Virtua" integrated library system, has a slightly different approach, offering "Visual MIS" (Multimedia and Imaging Solutions), professional services for digitizing and making available a variety of materials such as documents, maps, glass plates, slides, photographs, catalog cards, manuscripts, and microfilm. VTLS believes that outsourcing the digitization process ensures high-quality, professional scanning results without the necessity for a long-term investment in equipment, staff and time. The Visual MIS services address all aspects of the digitization process from collection analysis to database creation and interface design. For more information, see the VTLS Web site at www.vtls.com.

The following offerings are examples of products from mainstream multimedia vendors who are targeting the education market:

Artesia Technologies is a generalized multimedia management system vendor targeting the library market. Artesia's "TEAMS™" product is being used by Stanford University as the basis for its digital library initiative. Artesia describes TEAMS as a digital asset warehouse that serves as the hub for media files as well as for the vital information related to them. TEAMS integrates with other asset-intensive applications like Web content management, digital rights management, customer relationship management, syndication, and e-learning. TEAMS product information can be viewed at www.artesia.com.

Luna Imaging's "Insight®" high-resolution digital imaging software is used to digitize image collections, license digital images from existing image collections, or share content with other institutions. The Andrew W. Mellon Foundation is adopting Luna's Insight software as the platform to distribute ArtSTOR, a new independent, not-for-profit organization that will make a large body of digital resources for the study of art, architecture, and other fields in the humanities available for subscription. More information is available at www.luna-imaging.com.

MuseGlobal, Inc. specializes in information connectivity solutions. Its flagship product, MuseSearch™, is a broadcast search technology that allows unlimited numbers and types of information sources to be searched simultaneously with a single user query. It also incorporates linking features to take users from record results to category indices, content, and references. More information is available at www.museglobal.com.

Metadata

The key to locating, using, and preserving digital content is metadata, or structured data about digital objects and collections. Many digitization efforts have been unsuccessful due to inadequate metadata. There are three different types of metadata, all essential to ensure the usability and preservation of the collection over time.

1. **Descriptive Metadata**

This metadata provides information that a) allows discovery of collections or objects through the use of search tools, and b) provides sufficient context for understanding what has been found. When collections become large or when searching multiple collections (such as over the Internet) the discovery of objects of interest becomes a "needle in a haystack" exercise. Without agreed-upon metadata standards and the discipline of capturing and storing appropriate descriptive metadata, all but the smallest digital collections would be useless.

Metadata for individual objects varies by the type of object, but would include such things as its title, what it is, who created it, contributors, language, when it was created, where it is located, the subject, etc. At the collection level, users should be able to determine the scope, ownership, any access restrictions, and other important characteristics that would assist in understanding the collection. Probably the best-known descriptive metadata standard for libraries is MARC (MAchine-Readable Catalog) used for cataloging books and other publications. MARC has served the traditional library well, but was not designed for describing images, sound files, and other new media types.

An important emerging descriptive metadata standard for images and other multimedia objects is Dublin Core, a group of 15 items of information designed to be simple to understand and use. Dublin Core was designed to provide a very widely accepted mechanism to allow discovery, but with the option for different communities of users to adapt and customize it by adding more fields of particular importance to the community. In this way, the same base standard can be used for a wide variety of purposes and business models. Figure 3 describes the elements of the Dublin Core metadata standard.

Name	Identifier	Definition	Comment
Title	Title	A name given to the resource	Typically a Title will be a name by which the resource is formally known.
Creator	Creator	An entity primarily responsible for making the content of the resource.	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Subject and Keywords	Subject	The topic of the content of the resource.	Typically, a Subject will be expressed as keywords, key phrases, or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Description	Description	An account of the content of the resource.	Description may include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
Publisher	Publisher	An entity responsible for making the resource available.	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Contributor	Contributor	An entity responsible for making contributions to the content of the resource.	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Date	Date	A date associated with an event in the life cycle of the resource.	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.
Resource Type	Type	The nature or genre of the content of the resource.	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types [DCT1]). To describe the physical or digital manifestation of the resource, use the Format element.
Format	Format	The physical or digital manifestation of the resource.	Typically, Format may include the media type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).

Figure 3. *(continued on following page)*

Name	Identifier	Definition	Comment
Resource Identifier	Identifier	An unambiguous reference to the resource within a given context.	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Source	Source	A reference to a resource from which the present resource is derived.	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
Language	Language	A language of the intellectual content of the resource.	Recommended best practice for the values of the Language element is defined by RFC 1766 [RFC1766] which includes a two-letter Language Code (taken from the ISO 639 standard [ISO639]), followed optionally, by a two-letter Country Code (taken from the ISO 3166 standard [ISO3166]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.
Relation	Relation	A reference to a related resource.	Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.
Coverage	Coverage	The extent or scope of the content of the resource.	Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]), and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.
Rights Management	Rights	Information about rights held in and over the resource.	Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

Figure 3. Dublin Core Metadata Element Set[1]

[1] **This is the reference description, version 1.1 of the Dublin Core Metadata Element Set. See the Dublin Core Home Page (http://dublincore.org) for further information concerning the Dublin Core Metadata Element set.**

2. Structural Metadata

The second type of metadata is structural metadata. This describes the associations within or among related individual information objects. A book, which consists of pages and chapters, is one of the most straightforward examples of structural metadata. The structural metadata would explain how individual page images make up individual chapters, and how chapters make up the book. There could also be individually imaged figures, and structural metadata could also relate these to chapters or to a list of all figures in the book. Structural metadata aids the user in navigating among individual objects that comprise a compound object.

3. Administrative Metadata

Administrative metadata facilitates access, management, and preservation of the digital resource. It can describe the viewer or player necessary to access the object, automatically opening that viewer or player when a user selects that resource. It can describe attributes such as image resolution, file size, or audio sampling rate. It can provide a record of how and when an object was created as well as archival and rights management information.

An important emerging standard for interoperability of digital collections is the Metadata Encoding and Transmission Standard (METS), which provides a uniform framework for managing and transmitting digital objects. The Making of America II project (MOA2) developed an encoding format for descriptive, administrative, and structural metadata for textual and image-based works. Supported by the Digital Library Federation and the Library of Congress, METS builds upon the work of MOA2. It provides a format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users). Leading academic and research libraries are citing METS as an important standard for digital library interoperability, and seem to be rallying behind this standard.

Identity and Rights Management

As libraries increasingly focus on building gateways that direct patrons both to the library's own content and to networked digital resources owned or controlled by other entities, rights management becomes a major concern. "In constructing a digital library service environment, the library becomes responsible for configuring access to a world of information of which it owns or manages only a part," says the California Digital Library's Dan Greenstein. "Accordingly, the digital library is known less for the extent and nature of the collections it owns than for the networked information space it defines through a range of online services." In this type of cooperative resource-sharing environment, mechanisms for identifying information users and their access privileges are essential.

With increased use of the Internet to buy, sell, or license the use of documents, images, video, and other copyrighted content in digital form, comes the need to protect that content from unauthorized use once it is outside the control of the publisher or distributor. Digital content is very easy to copy and disseminate with little or no loss of quality. Accordingly, the ability to positively identify library patrons and their associated access privileges is required to assure that materials are being accessed and used by the appropriate parties. From a more practical perspective, it is also useful to integrate these identification and authorization mechanisms with other campus requirements, providing an environment where individuals can use a common facility for a "single sign-on" identification process shared with other campus computing applications.

Ideally, a common authentication/authorization process would be capable of positively identifying any computer user, linking that user to groups of other users with similar access requirements, and passing the user's identity and access profile to all campus computing applications that require such intelligence. A student enrolled in a particular class should be able to seamlessly access reserve materials in digital form (held in the digital library system) as well as online course materials available through the curriculum management system, the student's enrollment in the course being verifiable through the course registration system. The Sun ONE framework discussed in Appendix A provides these common identity and policy facilities—deployed to an open directory, available to all campus computing applications, and delivering this seamless ability to give students access to all resources to which they are entitled.

Preservation

Libraries place high value on preserving the irreplaceable contents of their collections, making the historical and cultural artifacts of our civilization available for future generations. Stanford University's Michael Keller comments, "Libraries are communications devices from preceding generations to succeeding generations. Debates about preservation of scholarly publications are tremendously important."

Fire prevention systems, environmental controls, and microfilming initiatives guard against catastrophic loss of irreplaceable items. Even the move from acid to alkaline paper in publishing has worked for long-term retention of printed items. In 1991, the American Library Association published a preservation policy outlining the responsibilities of the library profession for preserving access to information of all forms. In 2001, that policy was updated to reflect changes brought about by the Internet. The preamble of the revised policy states:

"Librarians must be committed to preserving their collections through appropriate and non-damaging storage, remedial treatment of damaged and fragile items, preservation of materials in their original format when possible, replacement or reformatting of deteriorated materials, appropriate security measures, and life-cycle management of digital publications to assure their usefulness for future generations."[2]

With digitized resources, it would seem that preservation would be much easier to achieve, especially since an unlimited number of copies can be created of the individual object. There are, however, a number of issues that complicate the maintenance of digital objects over long periods of time:

- **Deterioration of media.** Stone tablets have a considerable advantage over today's digital media for long-term storage. The problem of deterioration limits the useful life of today's digital media to between 5 and 50 years, while librarians debate how to retain the artifacts of our civilization for thousands of years to come.
- **Evolution in type and format of media.** In addition to the concerns of physical deterioration of media, we must be aware of the challenges posed by changes in media type and format. In the relatively recent history of personal computer use, we have seen an evolution from 5-1/4" floppy disks to higher density 3-1/2" diskettes to high-density zip disks, all requiring different physical disk drives and reader software. Optical storage technology is next, and there will undoubtedly be continual change in storage technologies.
- **Changes in applications and operating systems.** New software constantly appears on the horizon, rendering older versions obsolete. The hardware required by applications and operating systems software also changes over time. Any information stored in a given software/hardware environment will eventually be rendered obsolete through technological obsolescence, most likely before deterioration of the actual storage media occurs. Today, locating a system capable of reading 20-year-old Visicalc spreadsheet files or Multimate word processing files is a difficult proposition; in 50 years' time it will likely be impossible.
- **Preservation of processing results.** Some digital resources exist only fleetingly as a program runs and cannot be preserved as static objects. Preservation of such resources requires maintenance of the program and the surrounding operating environment in operable form.

Of course, not all digital resources have the same preservation requirements. "We are more concerned about persistence and authenticity of the collection of images than the long-term preservation of the surrogate," explains David Bearman, director of the Art Museum Image Consortium (AMICO). "Museums are in the business of pre-

[2] American Library Association Preservation Policy, Revised 2001.

serving the real thing." The skill set of the archivist, establishing criteria for selection of content to be preserved, will become even more essential as the amount of digital content expands.

Today, most digital libraries follow a schedule of copying archived digital resources to address the issue of media deterioration. As organizational standards for formats and applications evolve, many libraries also convert the objects over time to maintain readability. Specific preservation policies and best practices, however, are not well developed. The Research Libraries Group (RLG) is an international member alliance, including universities and colleges, national libraries, archives, historical societies, museums and independent research collections, and public libraries dedicated to "improving access to information that supports research and learning." In early 1998, RLG funded a study by Dr. Margaret Hedstrom and Sheon Montgomery to determine the state of digital archiving among 54 of its member institutions. Only 13 institutions reported they had established methods in place for digital preservation.

The National Library of Australia sponsors an initiative called "Preserving Access to Digital Information" (PADI) with the goal of providing mechanisms that help ensure that digital information is managed with appropriate consideration for preservation and future access. PADI recommends these strategies for long-term preservation of digital collections:

1. The migration of digital information from one hardware/software configuration to another or from one generation of computer technology to a later one offers one method of dealing with technological obsolescence.
2. Adherence to standards will assist in preserving access to digital information.
3. Technology emulation potentially offers substantial benefits in preserving the functionality and integrity of digital objects.
4. Encapsulation, a technique of grouping together a digital object and anything else necessary to provide access to that object, has been proposed by a number of researchers as a useful strategy in conjunction with other digital preservation methods.
5. It is universally agreed that documentation is an important tool to assist in preserving digital material. In addition to the metadata necessary for resource discovery, other sorts of metadata, including preservation metadata, describing the software, hardware and management requirements of the digital material, will provide essential information for preservation.[3]

Sun Microsystems has teamed with Stanford University, the National Science Foundation, and the Andrew W. Mellon Foundation to develop an approach to preserving

---

[3] See the PADI website at www.nla.gov.au/padi for additional information.

web-published materials (with a focus on electronic journals). The resulting technology, LOCKSS™ ("Lots of Copies Keeps Stuff Safe"), depends upon cooperation among multiple organizations with a common goal of maintaining access to the same content. A LOCKSS system run by each organization manages multiple web caches of the content and continuously compares the copies. If any cache appears to have been lost or has suffered disruption, the content is automatically replaced from the publisher or other caches.

A beta test is being conducted by more than 50 libraries worldwide, with a revised version of this software expected to be available through open source late in 2002. More information is available at http://lockss.stanford.edu/.

Portals

Deployment of portal technology has occurred side-by-side with digital collection development and access. As seen in Figure 4, portals are unified views into a set of disparate information sources and research tools, with the goal of providing users a simple way to locate and access all the information content they need and have authority to access. As libraries create, license, or negotiate access to more and more digital content, the need for an easy-to-use interface becomes increasingly important.

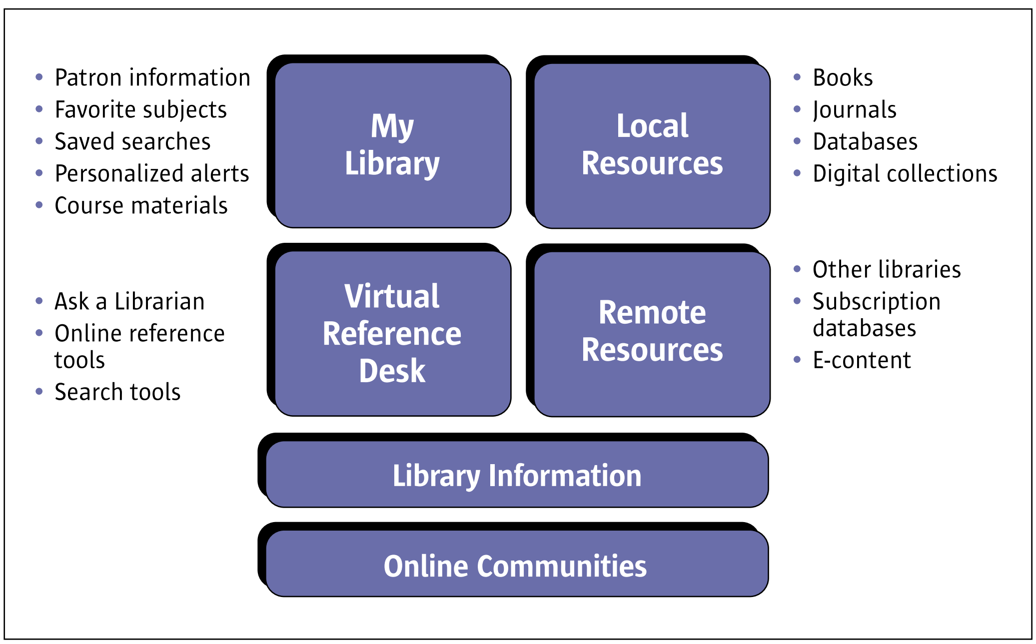

Figure 4. Library Portal Components

Library portals typically include an online catalog of materials as well as gateways to collections of digital resources accessible to the user. Broadcast search tools allow library users to search all of these sources simultaneously with a single query. Portals may include electronic reference services ("ask a librarian"), personalization features ("my bookshelf," custom intelligent searches), and other research tools. Enriched content, such as author biographies and book reviews, tables of contents, and jacket images can be provided to supplement the online catalog. Some libraries have built interactive features into their portals, allowing development of virtual communities.

Portals also provide an excellent opportunity to capture use statistics. "We still know very little about how many of the new systems are being used, who is using them, and who is not using them," says CNI's Clifford Lynch. "Good instrumentation will help you to refine the systems that you develop over time to be more responsive to your user community, and will also allow you to contribute to the collective knowledge base about how to best design and engineer such systems in the future." It must be recognized, however that concern about access to personal information continues to grow and any plan to capture data about individuals' activities must assure them adequate privacy.

Part 4 – Digital Library Program Profiles

Art Museum Image Consortium

www.amico.org

In 1997 a group of museum directors agreed to collaborate on making more museum collections available to educational institutions through digitization. The resulting non-profit organization, the Art Museum Image Consortium (AMICO) now includes more than 35 major museums in the United States, Canada, and the UK.

As the museums grew their digital collections and usage increased, they found that the task of dealing with requests by students and faculty took increasing amounts of time. AMICO provides a more efficient means of dealing with the work of rights management and access by clearing a licensed set of rights and delivering access to the consolidated collections through a variety of distribution mechanisms.

The AMICO Library™ includes multimedia objects portraying about 100,000 fully documented works of art from member museums. "Museums are focused on use, not discovery," according to David Bearman, AMICO's Director of Strategy and Research. "To make scholarly use of digitized works of art, you need to maintain a lot more data than what is needed for mere discovery," he says. More than three million AMICO Library subscribers access not only the images, but also find detailed provenance information, curatorial text, multiple views, and other related multimedia for many of the works.

AMICO maintains the half-terabyte master repository in a custom-developed relational database that is growing by 20,000 works per year. Current challenges include expanding the multilingual capabilities of the user interface to encourage more internationalization of membership and use, as well as encouraging the inclusion of more multimedia items in the collection. Clearing digital rights for works sought for inclusion in the AMICO Library also remains a continuing challenge.

University of California at Berkeley

http://elib.cs.berkeley.edu
http://sunsite.berkeley.edu

The University of California at Berkeley has been in the forefront of digital library innovation for many years. Projects begun at Berkeley included the creation of specifications for encoding electronic finding aids that are used to access special collections and archives. These specifications later evolved into the XML-based Encoded Archival Description (EAD).

The Making of America II project, which proposed a standard encoding for digital objects, was also lead by Berkeley. This project has since evolved into the Metadata Encoding and Transmission Standard (METS).

U.C. Berkeley has also become a significant resource for digital library researchers and developers worldwide. In 1996, with the support of Sun Microsystems, U.C. Berkeley created a SunSITE (Sun Software, Information, and Technology Exchange). Since that time the U.C. Berkeley SunSITE has provided:

- Catalogs and indices to electronic content.
- Links to significant digital collections.
- Digital library information resources.
- Links to significant digital library research and development.
- Information on tools used to build digital libraries.

Current work involves development of a modular object management environment, called GenDL (for "Generic Digital Library"). GenDL includes three components that can be thought of as three separate systems:

- The Web-based content management system, where creation and maintenance of the digital content is controlled. Descriptive, administrative, and structural metadata is created and linked to digitized or born-digital content. The resulting digital library objects are then encoded to the METS standard.
- The preservation repository (a joint project with the California Digital Library), where the digital content is managed to ensure its integrity and longevity.
- The access system, which is used to discover, display, and navigate objects that may have complex internal organizations (i.e., structural metadata).

Figure 5 shows the general architecture of GenDL.

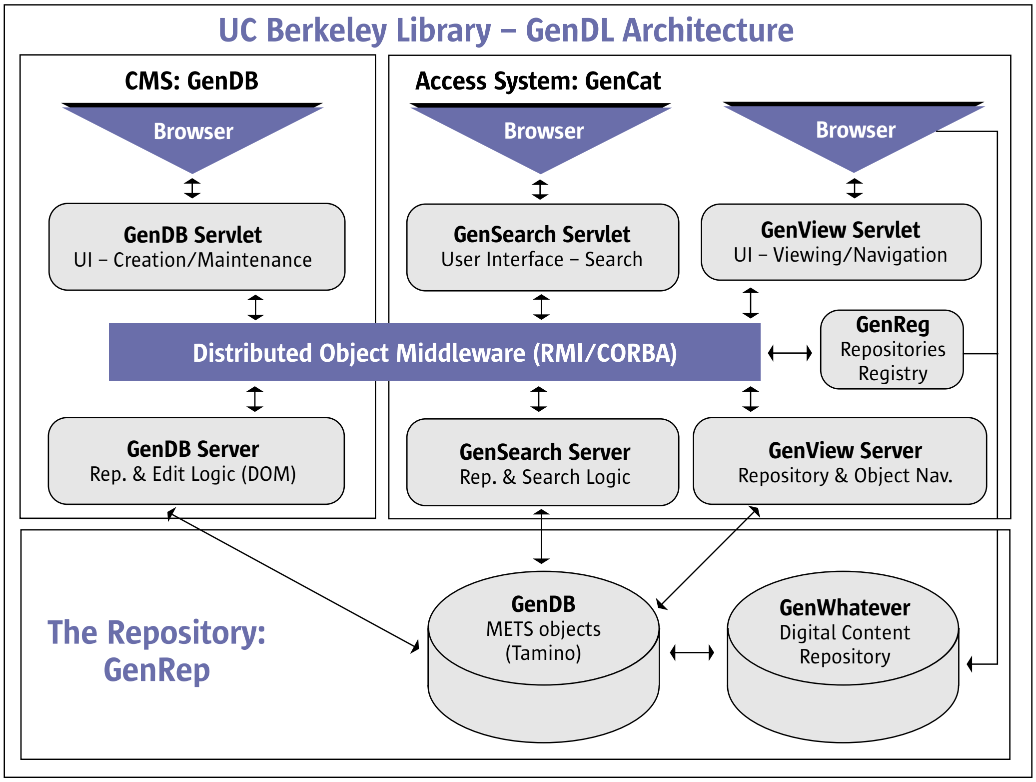

Figure 5. GenDL Architecture

The research and development thrust at U. C. Berkeley does not discount the efforts by software vendors to develop commercial tools that tackle these problems. "A large part of what we do in our initial development is to define the requirements and procedures that need to be supported in a digital library system," comments U. C. Berkeley Library's Director of Library Technologies, Bernie Hurley, pointing out that this allows commercial vendors to focus on introducing successful products.

Cornell University

http://campusgw.library.cornell.edu

"An article in the *Chronicle of Education* has claimed that the libraries are empty. That isn't true at Cornell. We're working very hard to keep that from happening. Although we are aggressively building digital collections and services, at the same time we are supporting greater ranges of use of library space," according to H. Thomas Hickerson, Associate University Librarian and head of Cornell University's digital library programs. Through the library, Cornell provides advanced information technology tool sets to students and supports faculty innovations in using technology to aid in teaching.

Having undertaken one of the world's most aggressive digitization programs, Cornell has been a developer of best practices and a leader in the field of digital libraries. As a Sun Center of Excellence for Digital Libraries, Cornell has served as a co-developer with Endeavor Information Systems for the product "ENCompass," a specially tailored multimedia management system intended for library implementations. Although Cornell has gained international recognition for its computer science advances in digital library technology, this university also recognizes the value of off-the-shelf products. "We would rather have this product supported by someone else. We don't want to maintain our own system any more than we have to," says Hickerson.

Cornell has attained recognition for the substantial digital collections it has developed. One of the best known is the "Making of America" collection, a multi-institutional digitization initiative making available primary sources of 19th century American culture and history, including popular magazines like *Harper's, Atlantic Monthly*, and *Scientific American*. To date, more than 900,000 pages, which are full-text searchable and freely accessible over the Internet, are in the collection. More than 40 other digital collections have been established, ranging from audio recordings of rare birdcalls to digital facsimiles of essential works in the literature of witchcraft and demonology, drawn from Cornell University's Rare Book Collection.

There are several strategic digital library priorities. One is to establish a central repository for all information resources deemed worthy of long-term maintenance. The repository's aim is to support the total lifecycle of those materials.

A Sun case study on Cornell's digital library program is available at http://www.sun.com/products-n-solutions/edu/success/cornell.

Harvard University

http://lib.harvard.edu

Harvard University has a very large and decentralized library system including more than 100 libraries and 35 major research collections. There is no hierarchy tying these libraries into a single organizational group. Harvard University's Librarian reports to the University President with a mission to provide services, including information technology, to the libraries. Accordingly, Harvard's digital library activities are designed to provide a common technology infrastructure, consulting assistance, and guidance on policy issues to all of the Harvard libraries.

Unlike many universities, Harvard has established major funding to support the development of a comprehensive infrastructure for digital libraries. The $12 million, 5-year program is called the "Library Digital Initiative," deliberately emphasizing the library orientation of the program. The goal of centrally funding this initiative to serve the decentralized network of libraries is to create incentive for participation in a common, standardized solution featuring a robust production IT environment and common practices for digitizing, reformatting, metadata creation, digital licensing, preservation, and migration of objects, etc. "We are tackling new and hard issues, and by developing this locus of expertise, setting up digital collections becomes easier for all the libraries," says Dale Flecker, Harvard University Library Associate Director for Planning and Systems.

The central Oracle repository supports all types of multimedia objects and a set of catalogs. A portal sits over the catalogs, allowing cross-catalog searching. Although born digital information is at the heart of the program, a program of internal grants has also been established to foster additions to the collection. Figure 6 gives a graphic representation of Harvard's digital library architecture.

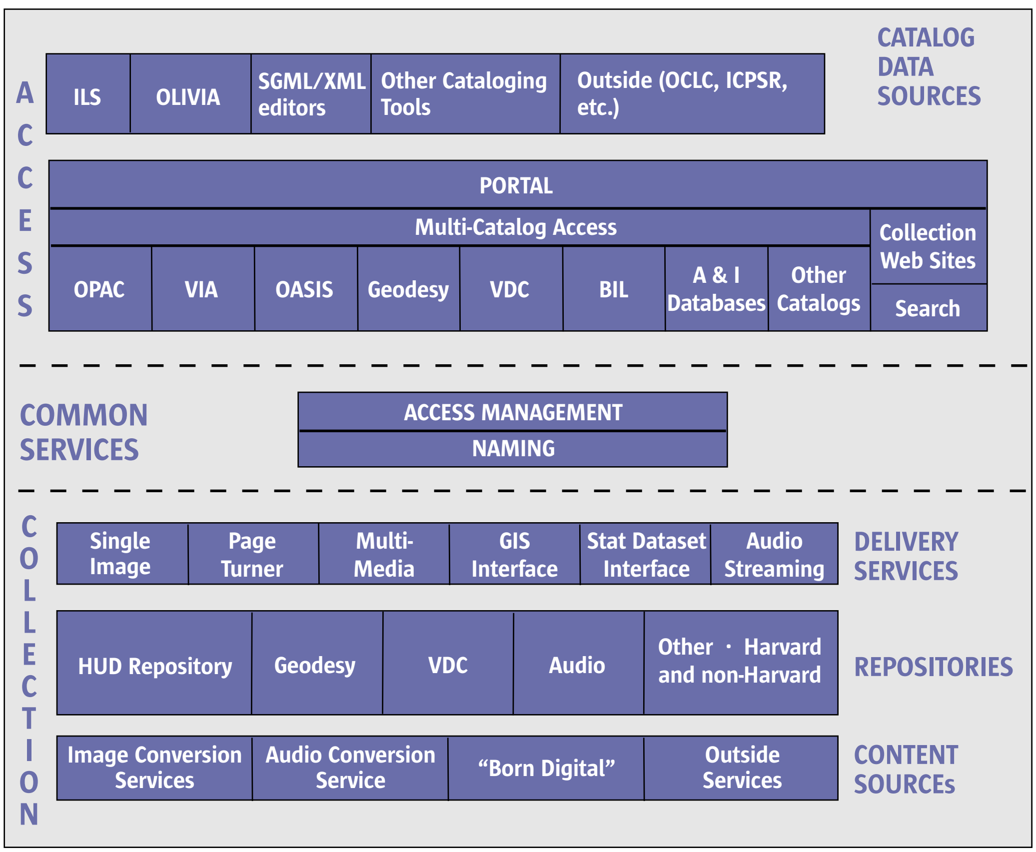

Figure 6. Harvard University Digital Library Architecture

There are active discussions under way with learning management groups about how to make digital library resources more directly relevant to learning and teaching. It is expected that there will be a lot of activity devoted to the interface with learning systems in the future. There will also be increased emphasis on scholarly communication and the libraries' role in capturing the university's intellectual output.

JSTOR

http://www.jstor.org

The problems associated with storing back issues of printed journals and locating articles of interest in those back issues led to the creation of JSTOR, an electronic archive of scholarly journals. JSTOR began as a pilot project sponsored by the Andrew W. Mellon Foundation in 1990. Ten journals in economics and history were selected initially, and all back issues were digitized. High-resolution images of each page were captured and linked to a text file generated with optical character recognition software. Table of contents indices were developed as well, to provide search and retrieval of journal content of interest. When five test libraries showed enthusiasm about the space savings and ease of access, JSTOR was born.

In 1995 JSTOR was established as an independent non-profit organization with the goals of providing a trustworthy archive of important scholarly journals, reducing the cost of accessing these materials, and assuring their long-term preservation. JSTOR now provides 1,100 user institutions worldwide access to 169 journals. Originally written in Perl, the Web interface has been ported to Java. Java's object-oriented approach makes it easier for JSTOR's geographically distributed developers to quickly write small pieces of reusable code. More than five million requests per month are satisfied through three sites at Princeton, the University of Michigan, and the University of Manchester. The entire JSTOR archive requires 2.2 terabytes of storage for more than eight million pages of journal content. Usage has been growing at a rate of 50% per year, with rapid growth among international users. There is also increasing emphasis on allowing access to the JSTOR archive by secondary schools and public libraries.

A Sun case study on JSTOR is available at http://www.sun.com/products-n-solutions/edu/libraries/JSTOR-ss.pdf.

University of Maryland

http://www.lib.umd.edu

The University of Maryland has opportunistically grown its unique digital collections, beginning with the Performing Arts Library and the Broadcasting Archives. Both were highly motivated to make their holdings more accessible and usable. Target collections range from original piano roll recordings of early 20th century concert pianists, to a complete wire-recorded collection of the Arthur Godfrey Hour radio broadcasts, to early rock 'n' roll recordings.

A 1996 implementation of a commercial multimedia management platform had mixed results, due to the difficulty of integrating the product into the existing environment and the lack of specific library functionality. The University of Maryland has since taken a different approach, becoming an Ex Libris "Premier Partner" for the Ex Libris "DigiTool" product. This involves full implementation of the Ex Libris digital asset management product, as well as ongoing participation in product design, development, and testing. The implementation will emphasize streaming audio and video, as well as images. The university's goal is an environment that handles all types of digital objects.

Advanced object management technologies are envisioned to assist users in accessing the Performing Arts Library audio collections. An example of an enhanced digital service is the capability of displaying sheet music images simultaneous with the music being played.

"It's good to have a vendor who works with library problems as a mainline business. Other commercial products don't necessarily map to the library business," says University Dean of Libraries Dr. Charles Lowry. Lowry also cites the ability to integrate the digital library software into the Web-based user interface as a benefit, so users can access the physical collection catalog in the same way as the digital collections. It is also important to the University of Maryland that the digital library software supports National Information Standards Organization (NISO) and National Institute of Standards and Technology (NIST) standards and protocols

Stanford University

http://www-sul.stanford.edu
http://highwire.stanford.edu

Stanford University is well known for its pioneering work in digital library development and electronic publishing. Stanford has participated in both phases of the NSF Digital Library Initiative, provided online publishing services to nearly 100 scholarly societies, developed extensive digital collections, and is currently implementing an advanced digital repository.

Stanford has entrusted the responsibilities of University Librarian, Publisher of the University Press, Publisher of the HighWire Press, and Director of Academic Computing to a single individual at Stanford. This assignment has fostered functional synergy that has integrated many information management activities across these disciplines and facilitated coordination between related activities such as digital repository and learning management system development. "We don't need to invent a separate digital library program at Stanford because the digital aspects are an accepted and pervasive part of library support for research and teaching," says University Librarian Michael Keller.

After putting up many digital collections on custom-developed systems, the Stanford Libraries have selected the "TEAMS" multimedia management software from Artesia Technologies as the base for its new digital repository. It is also utilizing Luna Imaging's "Insight" software for managing high-resolution images, and it will utilize the METS standard for managing multimedia objects. "METS is tremendous, an XML-implemented generalized metadata environment that lets you describe and interrelate any digital entity. We'll be able to store objects one time in one repository, encode them using different descriptive metadata, and make them available to all different types of applications on campus," explains Stanford Libraries' Chief Information Architect Jerry Persons. The repository will handle authentication using Kerberos certificates and will include rights management and charging mechanisms. Figure 7 illustrates the architecture of Stanford's Digital Repository.

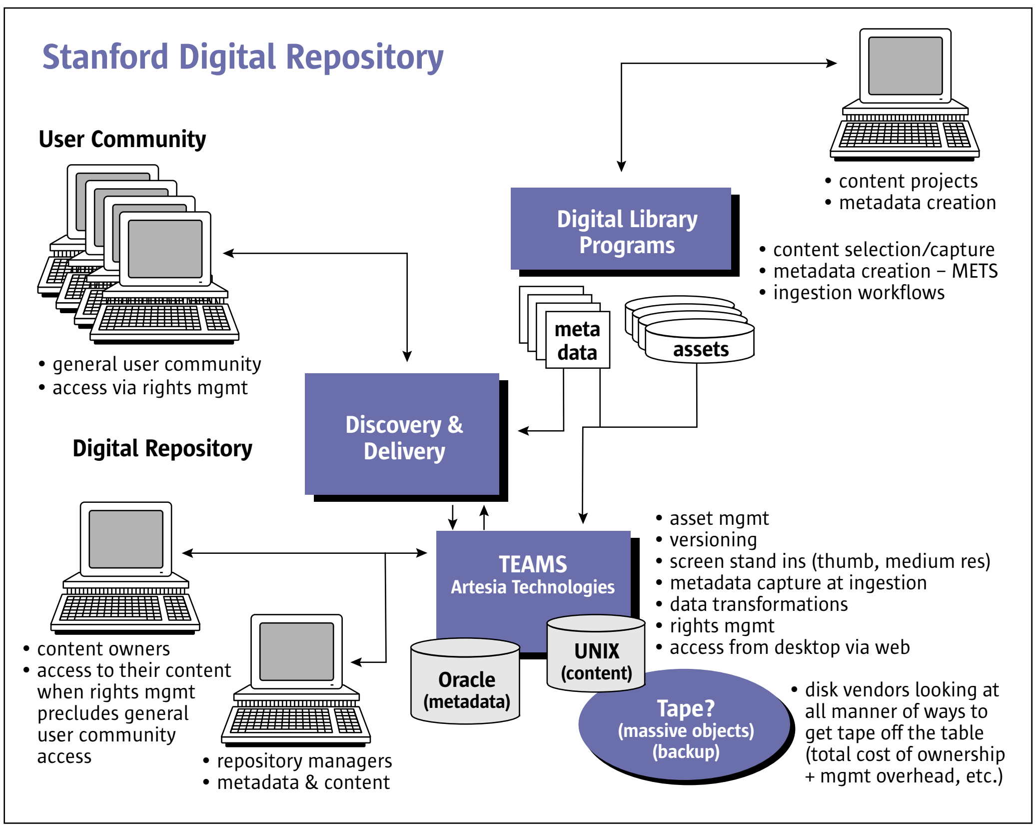

Figure 7. Stanford Digital Repository Architecture

Another well-known Stanford venture is HighWire Press, a very successful electronic co-publisher of more than 300 scholarly journals. HighWire Press was started in early 1995; its goal was to assist scientific societies in the transition to new technologies for scholarly communication by providing a common infrastructure. Small not-for-profit publishers are able to take advantage of economies of scale in making the transition to online delivery, thus giving them parity with for-profit publishers.

HighWire is known for its advanced functionality, creating value beyond what would be available in a physical publication, through:

- Links among authors, articles and citations, particularly featuring Toll-Free Linking, which allows readers of an article also to read the full text of other articles cited by the first without charge, whether or not that reader has subscription rights to the journals in which the cited articles appear.
- Advanced searching capabilities, high-resolution images and multimedia.
- Interactive functionality.
- More content (including more articles) than can be physically published in hard-copy.

Originally written in Perl and C++, HighWire migrated to Java in 1998. With the challenge of rapidly scaling the number of publishers supported, Java allowed HighWire programmers to develop application code much faster and more effectively. Performance and availability is critical; HighWire's web site serves 180,000 users and averages nine million hits weekly.

A Sun case study on HighWire Press is available at http://www.sun.com/products-n-solutions/edu/success/paperless.

Future concerns at Stanford include:

- How to improve the digital environment for teaching online, supplementing the teaching, and managing the teaching process.
- How to do data mining across all the different formats of digital information they have to improve the research process.
- How to make it possible for students and faculty to customize their views and take them wherever they go.

Part 5 – Future Directions

The pace of change in digital library technology and its applications has accelerated in recent years as the focus has begun to shift from R&D to full-scale deployment. Several key trends are emerging and will continue to gain momentum:

- The shift from text and image-based systems to audio and video will continue. As network bandwidth becomes more economical and streaming technologies improve, increasing numbers of institutions will have access to the practicality of full multimedia solutions.
- Broadly accepted best practices will emerge for digitization, rights management, preservation, metadata encoding, and other key digital library processes. The library discipline is highly collaborative and has a history of sharing successful approaches. Debate about these issues will recede as proven techniques mature and spread.
- Standards will move from the discussion and trial stage to widespread adoption. As library industry thought leaders work through existing standards organizations to agree on common approaches, software vendors focusing on the library market will incorporate these agreed-upon standards into their product lines.
- As the center of digital library activity shifts from computer science experiments back to mainstream library implementations, the next generation of digital library development and deployment will focus on standardization, usability, and productization—providing greater usability for library patrons, increased interoperability among digital collections, and more cost-effective choices for institutions just beginning digitization programs.
- New technology services will enhance the scholarly communication process just as they have assisted consumers in making purchases from online merchants and auction houses. Referring to the user-provided book reviews and evaluations of sources, CNI's Clifford Lynch comments, "Recommender systems and reputation management systems allow us to augment traditional information review approaches with new methods that mimic social practices and offer additional tools for helping users to deal with the ever growing flood of information." He cites an initiative where a thousand highly rated life scientists were asked to post their recommendations on new articles, which instantly created an expert panel that became an authoritative source for those interested in keeping current with the most important developments in life science.
- Growing dependence on digital information resources will create market pressure for the creation of cooperative solutions for long-term preservation. The charter of library consortia will enlarge to accommodate development of mass

storage facilities serving a range of institutions. This will be more cost-effective and more reliable than individual solutions.

- Even now, course management systems providers and textbook publishers are starting to work together. As more texts are published electronically, this linkage will become stronger. In time, digital libraries and learning management systems will be routinely integrated, repurposing the same digital content as both course content and reference material. E-learning will continue to blur the lines between course content, textbooks, and reserve materials provided by libraries as these all become digital content managed in common repositories or through common gateways. Digital libraries will be routinely linked to campus e-learning and administrative systems to provide a one-stop virtual campus.

It is clear that digital library technology is becoming an essential enabler of library services. In fact, one recent posting to a digital library e-mail list commented that the term "digital library" is starting to sound as anachronistic as "horseless carriage." Just as the term "electronic banking" lost favor and was subsumed by the more generic "banking," so we will see "digital libraries" discussed less and less often as employment of these technologies becomes universal. It is certain that all libraries of the future will be characterized by technology-based information services that extend and enhance the traditional mission of libraries in our society.

Appendix
Sun™ Open Net Environment (Sun ONE)

Based on these industry standards, Sun has defined an open software architecture to support interoperable services on demand across diverse devices. The Sun™ Open Net Environment (Sun ONE) provides an ideal model for deploying the rich media services that comprise the digital library functionality. Sun ONE provides an open framework for deriving information about a device user's identity, role, and location as well as security, privacy, and permissions requirements that allow the delivery of smart media services within the overall context of the organization's IT environment. Figure 8 summarizes the components of the Sun ONE smart Web services architecture.

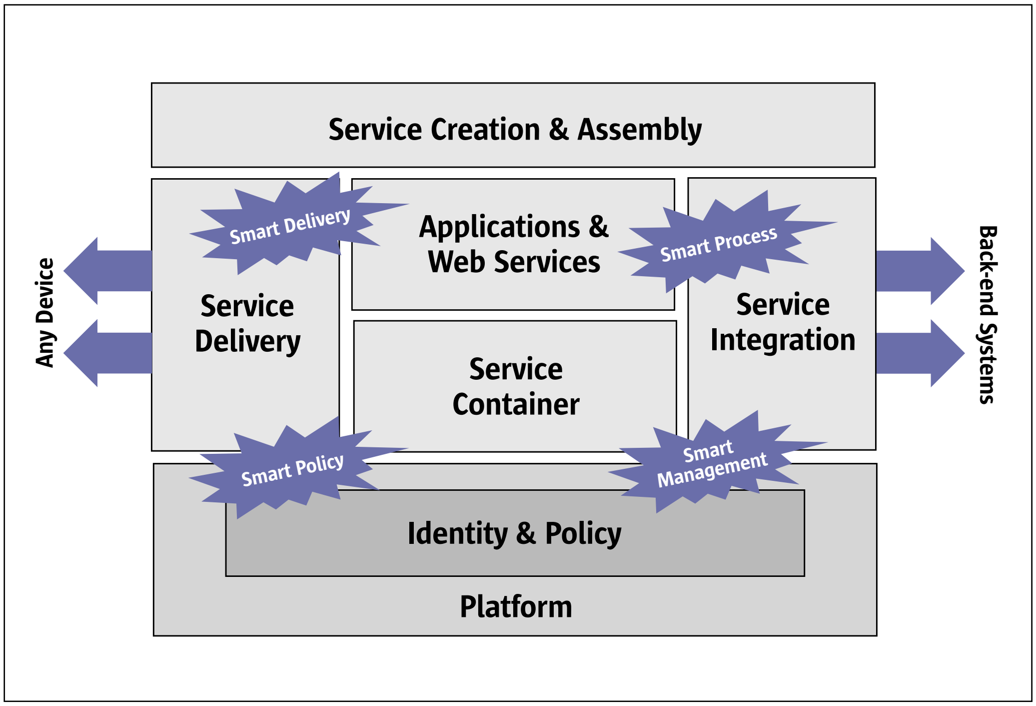

Figure 8. Components of the Sun ONE Smart Web Services Architecture

The Web services architecture includes these elements:

- Service creation and assembly facilities create discrete micro services and assemble them into composite services, or macro services.
- Service delivery facilities provide basic connection, location, discovery, and communication functionality.
- Service integration facilities enable the service to access other Web services and resources such as databases, files, directories, and legacy applications.
- Applications and Web services are the Web services that can be deployed on any type of platform or intelligent device.
- Service container is the runtime environment for the service and provides persistence and state management services, as well as process management facilities that manage service workflow and event processing.
- Identity and policy facilities are deployed to an open directory that manages identity, security, and policy.

The capability of a Web service to understand "context" or the situation it is in allows it to share this information with other services in order to perform as requested. Smart Web services operate within the Sun ONE architecture in four facilities:

- Smart delivery facilities will be able to aggregate, customize, and personalize service results based on context in the digital library of the future.
- Smart process facilities use context to affect business service workflow.
- Smart policy facilities coordinate activities according to policies associated with identity, context, and roles.
- Smart management facilities manage, monitor, and maintain a Web service for interservice registration, pay-per-use and subscription agreements, service provisioning, and management and runtime policies.

The open, modular, standards-based approach of Sun ONE is allowing the educational community to achieve continuous access to its information resources, an integrated infrastructure, and connectivity to disparate campus information systems through a single user sign-on. The Web-based Sun ONE platform extends the life of existing IT investments while making it easy to add new application services through the re-use of components and services.

See the Sun ONE website for more information at www.sun.com/sunone.

WORLDWIDE HEADQUARTERS: SUN MICROSYSTEMS, INC., 4150 Network Circle, Santa Clara, CA 95054 USA
Phone: +1-800-555-9SUN OR +1-650-960-1300 • Internet: www.sun.com

take it to the n[th]

SALES OFFICES

Africa (North, West and Central): +33-13-067-4680 Argentina: +5411-4317-5600 Australia: +61-2-9844-5000 Austria: +43-1-60563-0 Belgium: +32-2-704-8000 Brazil: +55-11-5187-2100 Canada: +905-477-6745 Chile: +56-2 3724500 Colombia: +571-629-2323 ommonwealth of Independent States: +7-502-935-8411 Czech Republic: +420-2-3300-9311 Denmark: +45 4556 5000Egypt +202-570-9442 Estonia: +372-6-308-900 Finland: +358-9-525-561 France: +33-134-03-00-00 Germany: +49-89-46008-0 Greece: +30-1-618-8111 Hungary: +36-1-489-8900 Iceland: +354-563-3010 India: Bangalore: +91-80-2298989/2295454 New Delhi: +91-11-6106000Mumbai: +91-22-697-8111 Ireland: +353-1-8055-666 Israel: +972-9-9710500 Italy: +39-02-641511 Japan: +81-3-5717-5000 Kazakhstan: +7-3272-466774 Korea: +822-2193-5114 Latvia: +371-750-3700 Lithuania: +370-729-8468 Luxembourg: +352-49 11 33 1 Malaysia: +603-21161888 Mexico: +52-5-258-6100 The Netherlands: +00-31-33-45-15-000 New Zealand: Auckland: +64-9-976-6800 Wellington: +64-4-462-0780 Norway: +47 23 36 96 00 People's Republic of China: Beijing: +86-10-6803-5588 Chengdu: +86-28-619-9333 Guangzhou: +86-20-8755-5900 Shanghai: +86-21-6466-1228 Hong Kong: +852-2202-6688 Poland: +48-22-8747800 Portugal: +351-21-4134000 Russia: +7-502-935-8411 Singapore: +65-6438-1888 Slovak Republic: +421-2-4342-94-85 South Africa: +27 11 256-6300 Spain: +34-91-596-9900 Sweden: +46-8-631-10-00 Switzerland: German: 41-1-908-90-00 French: 41-22-999-0444 Taiwan: +886-2-8732-9933 Thailand: +662-344-6888 Turkey: +90-212-335-22-00 United Arab Emirates: +9714-3366333 United Kingdom: +44-1-276-20444 United States: +1-800-555-9SUN OR +1-650-960-1300 Venezuela: +58-2-905-3800 Or Online at sun.com/store

SUN™

Printed in USA 8/02 FE1868-0 /XK