LA-UR- 98-2575

Approved for public release; distribution is unlimited.

Title: Integrating the Digital Library Puzzle: The Library Without Walls at Los Alamos

CONF-980861--

Author(s): Richard E. Luce

Submitted to: International Digital Library Summer School, Tilberg, The Netherlands

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Los Alamos
NATIONAL LABORATORY

DISCLAIMER

Integrating the Digital Library Puzzle: The Library Without Walls at Los Alamos

Richard E. Luce
LWW Project Leader and Research Library Director

Abstract

Current efforts at the Research Library, Los Alamos National Laboratory (LANL), to develop digital library services are described. A key principle of LANL's approach to delivering library information is the integration of products into a common interface and the use of the Web as the medium of service provision. Products described include science databases such as the SciSearch at LANL and electronic journals. Project developments described have significant ramifications for delivering library services over the Internet.

Los Alamos National Laboratory (LANL) is world renowned for its state-of-the-art computing facilities. Thus it is only fitting that a significant part of our Research Library would be a digital library, a "*Library Without Walls.*" The *Library Without Walls* allows researchers to connect via the Web and use computer technology to access local and worldwide information resources, anytime, from anywhere.

Many of the digital library efforts in the United States to date have been deep but narrow research projects. In contrast, the LWW project is distinguished by its integration of a broad set of library components that work synergistically to build a digital library in a production environment. Our goal is to integrate the delivery of information on the customers' desktop from digital resources. The long-term goal is to create a network of knowledge systems that facilitate collaboration among researchers.

Background

A variety of factors have propelled digital library developments to the forefront of information science, including: rapidly advancing information technologies coupled with customer expectations that libraries explore every avenue to deliver information effectively and efficiently; plans for a National Information Infrastructure, with its associated requirements for storing and accessing vast amounts of digital data over the information superhighway; and limited financial resources for the acquisition of ever-growing scientific publications, which continue to outpace inflation.

The LANL Research Library is a hybrid between an academic library with strong emphasis on collections and a special library with emphasis on service and customized products. During the last several years, the conceptual paradigm of research libraries has

changed from a focus on the building housing physical collections to that of an information access service, bounded neither by physical location, nor by traditional book and journal collections. In this paradigm, patrons connect remotely and use technology to access local and worldwide library, public, and commercial information providers -- a concept known as the **digital library**.

In late 1993, Los Alamos National Laboratory (LANL) underwent a massive reorganization. The Research Library was placed within the new Computing, Information, and Communications Division of the Laboratory. The reorganization provided the Research Library with an opportunity to raise the visibility of several major strategic initiatives focusing on information management. One of the most significant initiatives is the *Library without Walls*.

Several drivers are accelerating change at LANL. One of the key drivers is the desire to increase the researchers' scientific productivity. In addition to meeting our customers' needs, the continued funding for the LWW project has been contingent on successfully addressing the issue of increasing researcher productivity.

Key Goals

The Research Library mission emphasizes building partnerships with customers. The LWW project embodies that principle by requiring all product development to be based on customer requirements. It embodies the belief that we will become a leader in providing access to global electronic information resources by anticipating and meeting our customers' needs.

Several key principles define the *Library without Walls* initiative:

- Use of the Web. A key early assumption was the use of the World-Wide-Web for access and delivery. The heterogeneous computing environment at LANL places a premium on delivering digital library services across multiple platforms. The exponential growth of the Web, with browsers such as Netscape as a preferred client tool, has significant implications for library access and developing applications.

- The virtual library is not a single entity at Los Alamos but requires the seamless integration of other virtual library resources through technology linkages. The convergence of a multitude of distinct efforts, encompassing a variety of dimensions and each solving different facets of the digital library puzzle will be required.

- The digital library requires convergence of many distinct efforts, encompassing a variety of fields (e.g., information science, networking, computer science, cognitive psychology, sociology, etc.), will be required.

- It will be necessary to incorporate interactive compound documents and digital artifacts that extend beyond the linear capabilities represented by print publications.

- Systems and products will need to be measured by their ability to facilitate new forms of collaboration among users.

The long-term goal is the creation of a network of knowledge systems which facilitate synergy and collaboration between people.

Laying the Foundation

Before embarking on a digital library initiative, a sound infrastructure was required to improve our information technology foundation and provide appropriate tools to support our service requirements. The following three capabilities were integrated to provide that foundation:

(1) Library Automation System Upgrade. In 1992, the Library upgraded its automated library system and online catalog by migrating to the Geac Advance system. Previously, the automated system had been viewed as the centerpiece of library automation and technology efforts. That notion was replaced with a framework in which the new automated system became one of several modules, each of which had to be integrated to meet a multiplicity of information access and delivery requirements.

(2) Modernizing the Staff Workstation. Providing appropriate technology tools for library staff was imperative to enhancing our collective technology knowledge, and ability to respond to customer requests. Three basic workstation goals were established: all library staff would have appropriate workstations on their desks; all workstations must access all library and administrative systems; and all workstations had to be connected to internal and external networks. Implicit in these goals was the requirement to keep these workstations current and up-to-date.

(3) Developing a Robust Local Area Network (LAN). The key criteria for developing the LAN were integration of the desktop tools with access to Internet resources and shared application software to facilitate communication. Initially Ethernet connections were installed, providing a shared 10MB Ethernet backbone. The LAN has grown today to a 100MB backbone with dedicated 10MB pipes to the staff desktop.

First Generation Products

The first generation of LWW products, initiated in late 1994, was designed to meet the key customer requirement of easy access to information supporting scientific research. Strategic assessment led to selection of the emerging World Wide Web for delivery. Four product lines were sequentially developed to respond to those needs:

1. Electronic Journals

Two objectives have been established for this LWW component: (1) Integrate full-

image/full-text journal articles with the library catalog; and (2) integrate full image/full-text journal articles with search and retrieval of citation databases such as *SciSearch® at LANL*, the Energy Database, INSPEC, BIOSIS, Engineering Index, etc. Databases and electronic journal articles have been integrated to allow users to search large databases, click on an icon and read the associated article(s) on the desktop.

Currently over 800 full-text, full-image journal titles are now available for authorized LANL users through network connections. We are collaborating with several publishers to provide electronic journals to the desktop in PDF (Portable Document Format). Extensive partnerships with publishers have been developed to provide this capability. Publishers working with the LWW project include Academic Press, AIP/APS, Elsevier Sciences, HighWire Press, Institute of Physics, Springer-Verlag, and the American Mathematical Society. In addition, members of the New Mexico Library Services Alliance have joined the project and cooperate on consortium purchasing of electronic journals. The LWW is now beginning to develop browsing capabilities to these 800+ electronic journals.

2. Access to Scientific Literature

SciSearch® at LANL

The *SciSearch® at LANL* database contains 16 million bibliographic citations from 1974 through the present delivered via the Web. LANL licensed the ISI Science Citation Index database and developed a Web front-end user interface. SciSearch is an international, multidisciplinary index to the journal literature of science and technology. The database, which covers 5,200 journal titles, contains 16 million citations, spanning 1974 to the present, and 18,000 weekly updates. LANL has linked over 200 million cited references and automatically created over 95 million active hyper-links. Recently over 560,000 electronic journal articles have been hyperlinked to the product.

To help researchers keep pace with the explosion of scientific publications, LWW developed a new service -- customized alerts with individual profiles. Weekly Alerts is an automated service that can be used to track the contents of favorite journals or subjects of interest. Users register and create individual profiles (a customized search strategy), which are then checked against the 18,000 weekly citations added to the database. New items matching the researcher's profile automatically generate an e-mail notification of the new item of interest. New citations to papers can also be tracked using this service. Researchers can be notified when others cite their papers, or they can be notified when an important paper in their field is cited.

Citations of interest can be quickly compared with the Research Library's holdings to determine if the needed volume is in the collection. Researchers can also mark citations they are interested in downloading from their browser, and then download them all at once in various formats. Once a list of citations has been downloaded, it can be printed, saved to a file, or e-mailed to a colleague.

SciSearch® at LANL was the first Web-based product for this large scientific database, released in 1995. *SciSearch®*, one of the first LWW products, was developed and deployed 2 years before the commercial data supplier (ISI) understood how to deliver a Web-based product (Web of Science).

BIOSIS® at LANL

To support Life Sciences research, the *BIOSIS@ LANL* database was put up on the Web for Lab researchers. The 11 million record *BIOSIS at LANL* (Bioscience) database, has been linked to over 300,000 electronic journal articles. The database contains citations, with abstracts, from Biological abstracts and Biological abstracts/RRM (reports, reviews, meetings). BIOSIS indexes approximately 6,500 journals and 2,000 meetings/year, as well as books and other materials. *BIOSIS@ LANL* has customized links to HighWire Press (Stanford University) journals and articles that are unique in the world today. (Stanford University has also contracted for use of this database).

INPSEC® at LANL

LWW developed a Web interface to the 6 million record INSPEC database and has linked over 300,000 electronic journal articles to the database. INSPEC (Information Service in Physics, Electrotechnology and Control) is the leading English-language citation database for the world's literature on all aspects of physics, electronics, and computing. INSPEC scans papers from approximately 4,200 journals, 1000 conferences, and other publications.

Energy Science and Technology Database

The Energy Science and Technology database (1974-1997) was indexed and made available to LANL researchers via the Web. The database had previously only been available through commercial vendors and is not available via the Web from any other source. The DOE Energy Science and Technology Database is a multidisciplinary database of 3.6 million records, containing worldwide references to basic and applied scientific and technical research literature. The database includes references to publications provided by the U.S. Department of Energy, its contractors, and other government agencies. Also included are information from the International Energy Agency's Energy Technology Data Exchange (ETDE) and the International Atomic Energy Agency's International Nuclear Information System (INIS). Approximately half of the references are from sources outside the United States.

3. Digitized Los Alamos technical reports

To fully understand the issues involved in electronic publishing, the LWW decided to digitize the formal Los Alamos reports. This effort required scanning technical reports dating back to 1943. Once the reports were scanned, they were OCR'd to make the documents word searchable. After much experimentation, LWW finally settled on using the Adobe suite of products to OCR and create PDF documents. A final step in the conversion process was connecting the PDF document to the online catalog for retrieval via the Web.

Today, over 12,000 LANL technical research reports have been digitized and made available to researchers worldwide via the Web. During this phase, LWW was able to completely process 8,000 pages daily. Once we reached this point, we felt we understood the technology well enough to outsource the production work to a local business with expertise in scanning and document conversion.

4. Web interface to the Library's public catalog

In 1995-96 the library community was filled with great hope and expectation for Z39.50 products to provide search and retrieval capabilities on heterogeneous systems. The Z39.50 software applications from commercial supplies had not yet reached maturity however. At this point LANL obtained a license to OCLC's WebZ product, which we then customized to provide the capabilities we required to interface to our Geac Advance system. The LWW team subsequently developed the first production Web interface for the Geac Advance library catalog. This development spurred Geac to develop a commercial version, which they now market.

Some General Lessons to Date

Digital library tools seem to have a common set of requirements to be successfully used in the research environment. Among the key requirements, the following are essential:

- Usability – DL tools must be easy to use and adapt to the task at hand
- Support browse and serendipitous discovery
- Powerful and scalable
- Interact in and integrate with the larger environment. Hence tools cannot be product or vendor specific since researchers need data from other sites
- Collections will become distributed, thus search engines and location of data get separated
- Support compound documents, rapid prototyping, user feedback
- Allow the user to manipulate the presentation of data and the results of interaction with the toolset(s).

Next Development Steps

In late 1997 the development of the second generation of LWW products was initiated. Those products include:

1. Building a "mega-database", which will attempt to combine all our scientific databases into one virtual database. The impetus for this effort is the result of customer feedback that researchers either don't know (or don't care) about the subtle distinctions between database products, or they do not enjoy spending time performing multiple searches across several large databases to find what they are looking for.

2. Custom Alerts integrated in all LWW products. The capabilities developed for SciSearch at LANL will be spread across all LWW databases, allowing the user to establish profiles and receive weekly updates that match that profile.

3. Customized Interface, which will recognize individual researchers and remember both their user-interface preferences and what patterns they like to use to perform their library research.

4. Distributed Collections, which will incorporate distributed collections owned by other institutions into our digital library framework. Experimentally, through a partnership with the Naval Research Laboratory, during the first stage metadata will be exchanged between institutions and we will access content from remote servers. We believe this will allow us to further existing standards for identifying content on remote servers in a transparent fashion.

Growing a New Customer Base

The Library Without Walls products have attracted interest from beyond the Laboratory. The Department of Energy has designated LWW as a unique resource, which has allowed the Project to partner with external customers in delivering digital library services outside LANL and provides a medium for future technology exchange. LWW is now providing access to *SciSearch® at LANL* to the following external customers:

- Stanford University
- University of New Mexico
- Lawrence Livermore National Laboratory
- Sandia National Laboratories
- Phillips Laboratory and the Air Force Research Laboratories
- New Mexico State University
- New Mexico Institute of Mining and Technology

In addition, the LWW project has formed partnerships with:

- Journal publishers
- Libraries in New Mexico and the United States
- Electronic database producers
- Software companies whose applications facilitate integrating this broad set of digital information together.

This stems from recognition that no one institution is or will be able to build a successful digital library alone.

Summary

This paper has outlined LANL's LWW project approach to developing integrative information technologies with broad access capabilities to provide virtual library services to its customers. A common theme of all the *Library without Walls* projects described in this paper is the use of the latest non-proprietary client tools preferred by the research

community which are available via the Internet. The Research Library has taken significant steps to not only deliver digital information to the Laboratory, but also to the worldwide scientific community and the general public via the Internet. These efforts have provided a platform for us to learn and explore new avenues of delivering information to our customers.