Trends in Digital University Libraries[*]

Jay Jordan[**]

Resumen

El autor discute cinco tendencias en bibliotecas digitales universitarias: 1) comunicación académica; 2) desarrollo de depósitos institucionales; 3) preservación y almacenamiento digital; 4) metadatos, y 5) hábitos de información de estudiantes. La cooperación será el elemento clave en la evolución de la biblioteca digital. (FRRE)

Palabras clave: bibliotecas digitales, archivos digitales, cooperación bibliotecaria, comunicación académica, depósitos institucionales, metadatos, hábitos de información. (FRRE)

Abstract

The author discusses five trends in digital university libraries: 1) scholarly communication; 2) development of institutional repositories; 3) digital preservation and archiving; 4) metadata; and 5) information habits of university students. As the digital library continues to evolve, cooperation will be a key element.

Keywords: digital libraries, librarian cooperation, digital archives, scholarly communication, institutional repositories, metadata, information habits.

Cooperation in the digital future is a topic of great interest for OCLC and for the university libraries that participate in our international cooperative. Trends in Digital University Libraries is a very large topic because, as we all know, librarianship has been in period of extended and almost continuous change since the advent of the first computers in libraries, now over 30 years ago.
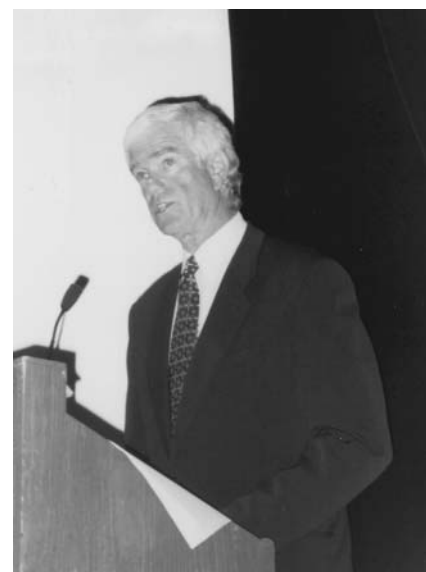
University libraries are going through a period of tremendous change, and we should view this as an exciting time to be in the profession. I would like to share with you a statement written by a distinguished librarian and the founder of OCLC, Frederick G. Kilgour, on the nature of change in librarianship.

"Libraries now find themselves forced to a dynamic state of instability, which should not be confused with insecurity. Since the age of Pericles, the

Trends in Digital
University Libraries

unstable periods in human history have been the most productive periods. Certainly, the period of instability into which libraries are now entering after a long period of changelessness will be one of the intellectually productive periods in librarianship".[1]

Fred Kilgour wrote that before the appearance of the World Wide Web. Indeed, he wrote it in 1966. Nearly 40 years later, we are still in an age of instability, and if we are lucky, it will continue, because although we have made much progress, we still have much to do. While Fred looked at this as another Age of Pericles,

other librarians have looked upon as the ancient Chinese curse, "May you live in interesting times". Either way, we are in the middle of building digital university libraries.

As we build these digital university libraries, the more things change, the more they remain the same. Libraries still collect, organize, provide access, and preserve information. But, as you know, the variety of resources that libraries work with, and of tools available to them, has increased significantly. They are using new electronic tools that until recently were only dreamed of. They find themselves having to collect and organize materials that range from terra cotta cones to e-mail. The library without walls has become a reality. You no longer go to the library, the library comes to you.

In this paper, I will briefly address trends in digital university libraries in five areas: 1) scholarly communication; 2) development of institutional repositories; 3) digital preservation and archiving; 4) metadata; and 5) information habits of university students.

At the outset, I will focus not only on technology, but also on people and institutions. As Michael Lesk has pointed out: "Building digital libraries is not just a question of piling up disk drives; it involves creating an entire organization of machines and people, perhaps even a culture, in which people are able to find information and use it".[2] Let me explore the first aspect of this culture, which is scholarly communication.

1) Scholarly Communication. Digital publishing technologies are forcing a fundamental change in the structure of scholarly communication. This audience is all too familiar with the issues and sometimes conflicting needs of scholars, university libraries and

---

[1] KILGOUR, Frederick G. Implications for the Future of Reference/Information Service. In: *Collected Papers of Frederick G. Kilgour, OCLC Years.* Dublin, Ohio : OCLC, 1984, p. 14.

[2] LESK, Michael. *Practical Digital Libraries: Books, Bytes, and Bucks.* San Francisco, California : M. Kauffman, 1997. p. 5.

publishers. In the existing publishing model, publishers acquire and distribute the scholarly output from academic communities and then resell it to those same communities, with the publisher retaining copyright of the scholarly information. At the same time, costs of scholarly journals and publications have steadily escalated, which in turn reduces library budgets for monographs. Over the past twenty years, the prices of scholarly journals have greatly exceeded ordinary inflation. After almost a decade of data collection, certain trends have become clear. The average percentage of the library materials budget that is spent on electronic materials was 16.25 percent in 2000-01, nearly five times as much as in 1992-93. Almost $132 million was reported spent on electronic resources in 2000-01, by 106 universities. There are some ARL librarians who believe that this figure is actually higher than reported and growing.

Library budgets have been unable to keep up with these price increases. At the same time, library users -university faculty and researchers- demand that scholarly journals be held in libraries as a vital part of the research and communication process.

In 1987, *Ulrich's International Periodicals Directory* began listing e-journals for the first time; that year, there were 1,200 e-journals. In 1996, there were 5,517 e-journals, and in 2002, there were 31,888. This steady increase in the number of e-journals is worldwide. Some 15,500 are published in the U.S.In Latin America, Mexico is the leader with 278 e-journals, followed by Brazil with 133 titles and Argentina with 121.[3]

The academic community has put forth proposals to reassert control over the scholarship produced by the academy. Marshall Keyes has summarized these proposals as follows: "Librarians can spin endless scenarios about the future of electronic scholarly publishing. In the most benign, colleges and universities declare all faculty research to be 'work done for pay,' allowing institutions to claim copyright for written material as they do now for patents. The institution then cooperates with libraries and scholarly associations to make this material available to the world at a low cost. The opposite scenario… has publishers breaking up journals and charging high fees for individual articles".[4]

To date, one of the most ambitious and most well known initiatives in scholarly publishing is an initiative of the Association of Research Libraries called SPARC, the "Scholarly Publishing and Academic Resources Coalition". This alliance of universities, research libraries and other organizations was established in 1998, and now has approximately 200 member institutions in North America, Europe, Asia and Australia. SPARC's agenda focuses on enhancing broad and cost-effective access to peer-reviewed scholarship through: 1) incubation of competitive alternatives to current high-priced commercial journals and digital aggregations; 2) public advocacy of fundamental changes in the system and culture of scholarly communication; and 3) education campaigns.[5]

One of SPARC's publishing partners is BioOne, which began operation last year. BioOne is a web-based aggregation of research in the biological, ecological

---

[3] *Ulrich's International Periodical Directory*. New York, N.Y. : R. R. Bowker, 2001. Irregular Serials & Annuals.

[4] KEYS, Marshall. The Evolving Virtual Library : A Vision, Through a Glass, Darkly. In: *The Evolving Virtual Library II*. Medford, New Jersey : Information Today, 1999. p. 172.

[5] SPARC *: The Scholarly Publishing and Academic Resource Coalition* [online]. <http://www.arl.org/sparc> [August 20, 2002].

and environmental sciences. The BioOne journals are available outside the U.S. and Canada in the OCLC FirstSearch Electronic Collections Online service. OCLC is providing funding for the development of BioOne's database and is also serving as one of the archive sites for BioOne content. There are now 41 libraries subscribing to BioOne in 15 countries: Australia, China, Finland, Israel, Japan, Korea, Mexico, New Zealand, Norway, Saudi Arabia, South Africa, Sweden, Switzerland, Turkey, and the U.K. BioOne added 11 journals its first year, bringing the total to 52. BioOne content is linked to citations in other FirstSearch databases.

JSTOR is another important development in scholarly communication. JSTOR converts paper journal titles to searchable, digital form and makes them available to the scholarly community. Since its founding in 1995 by the Andrew Mellon Foundation, there are now 949 institutions participating in the U.S., and there are 451 institutions participating outside the U.S., including eight in Mexico and 26 in South America. There are 296 journals from 156 publishers in the system. OCLC provides back-up support for the JSTOR system At this writing, JSTOR was exploring various models to support e-archiving of not only print publications, but of publications born digital.[6]

In 2001, the Mellon Foundation launched another initiative in scholarly communication -ARTstor. The rationale for this new initiative is similar to that of JSTOR, in which libraries redundantly store long backruns of journals. An independent not-for-profit organization, ARTstor will develop, store and distribute electronically digital images and related scholarly materials for the study of art, architecture and other fields in the humanities -particularly for teachers, students and scholars at educational and cultural institutions. This shared archive will reduce costs for participating institutions by making it unnecessary for them to create their own core archives. The resulting web image gallery will distribute a number of deep scholarly collections. ARTstor will work with content providers to obtain perpetual, non-exclusive rights to aggregate such materials and distribute them electronically for educational and scholarly purposes and support the mission of institutions that seek to expand access to their own holdings for academic audiences without incurring the financial and administrative burdens of distribution.[7]

Individual research libraries are not only participating in such efforts as SPARC, JSTOR and ARTstor. They are also acting on their own to meet the needs of their constituents.

For example, Sarah Thomas, Carl A. Kroch University Librarian, Cornell University, recently wrote to the faculty members of the university: "One new role for the Library is that of electronic publisher, using our expertise in digital technology to assist in the dissemination of knowledge and ideas".[8] She went on to list some of the activities the library is involved in: digitizing architectural photographs and historic

---

[6] *Journal Storage : The Scholarly Journal Archive* [online]. <http://www.jstor.org> [August 2002].

[7] The Andrew Mellon Foundation, Press Release, April 5, 2001. The Andrew Mellon Foundation, 140 East 62nd Street, New York, NY 10021, U.S.A.

[8] THOMAS, Sarah E. "The Carl A. Kroch University Librarian" [online]. In: *Gateway to Cornell University Library, March 2002* <http://campusgw.library.cornell.edu/about/message3-02.html> [August 2002].

literature in home economics; creating an online archive of agricultural journals; helping independent publishers of mathematics journals make the transition to online publishing; working with faculty to develop digital course reference materials for online learning.

Another trend worth watching is the development of electronic peer review systems. Among the 25,000 peer-reviewed journals worldwide, an increasing number are abandoning the traditional, mail-based process for reviewing and editing articles, turning instead to electronic systems. About 30 percent of scholarly publishers have adopted online systems. The *Journal of the American College of Cardiology* expects to save 80 percent of its postal budget, $60,000 to $70,000 per year, by switching to an electronic system. Other advantages include convenience and reduced risk of errors and lost mailings. The Nature Publishing Group in London reports that about 90 percent of reviewers have adopted its electronic peer review system for 18 journals. Indeed, there is consensus among publishers that electronic peer review will soon be universal as demand for speedier publication increases.[9]

2) Institutional repositories. In the past year, universities have begun experimenting with super archives, or institutional repositories. These institutions are inviting their professors to load copies of their research papers, data sets and other works into the repositories. The objective is to gather as much of the intellectual output of an institution as possible in a searchable online collection.
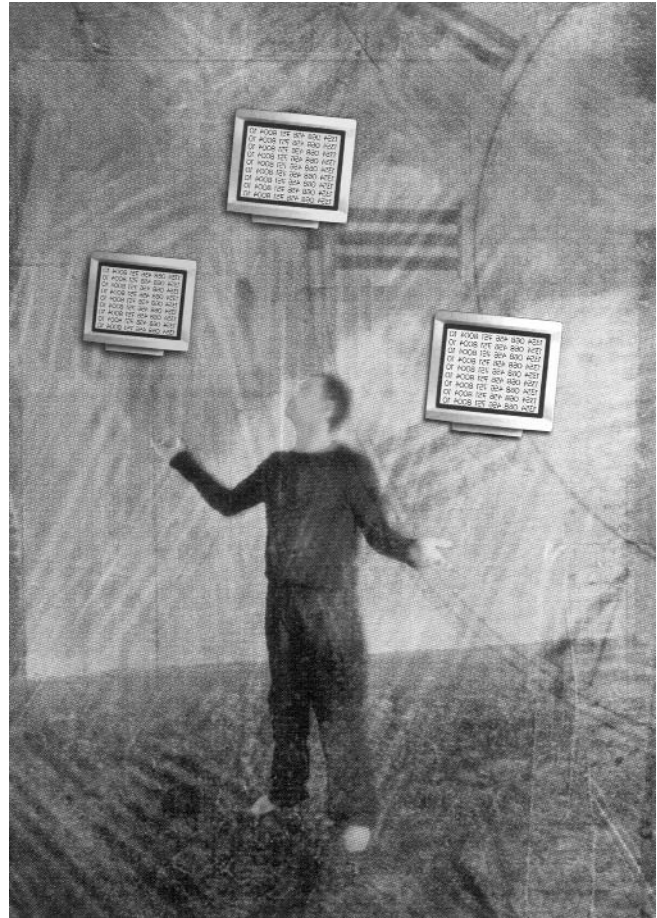
In July, SPARC released a major white paper, "The Case for Institutional Repositories: A SPARC Position Paper," which examines the strategic roles institutional repositories serve for colleges and universities. According to the announcement, the paper asserts that institutional repositories are a natural extension of an academic institution's role as a generator of primary research, and envisions such repositories as critical components in the evolving structure of scholarly communication. The repositories will answer two

---

[9] MILLSTEIN, Sarah. "Scholarly Reviews Through the Web" [online]. *New York Times*, August 12, 2002. <http://www.nytimes.com/2002/08/12/technology/12NECO.html> [August 2002].

challenges currently facing academic institutions. First, institutional repositories will reform scholarly communication by stimulating innovation in a disaggregated publishing structure. Second, they will serve as tangible indicators of an institution's quality, thus increasing its visibility, prestige, and public value.

One such institutional repository is Dspace at the Massachusetts Institute of Technology (MIT). There, a scholar who wants to present a working paper to her/his colleagues around the world no longer has to submit the paper to a print journal, then wait months for it to be published. Instead, the researcher can simply pull up a web journal and instantly submit the paper and data sets online. This virtual intellectual asset sharing is part of a joint project between MIT and Hewlett-Packard to create a long-term, sustainable digital repository. This fall, MIT will open the Dspace archive to all its professors. The project will also release a set of free software tools so that any college or university can create its own online repository.

MacKenzie Smith, Associate Director of Technology for MIT Libraries, has made the observation that as university libraries attempt to preserve vast amounts of research in electronic form, technological obsolescence is a major problem. She stated: "It is yet to be proven that publishers who electronically publish this material will succeed in preserving it".[10]

The University of Glasgow runs an e-print service, which provides access to the scholarly work and research of the University, including such items as peer-reviewed journal articles, theses, chapters, conference papers and project reports. The eprints.org archive is a joint operation of the University Library and the Computing Services department. It uses the eprints.org software developed at the University of Southampton. The service is installed and hosted on a server maintained by the Computing Services staff, with the Library taking responsibility for its look and feel, administration and promotion. The software is OAI-compliant and can produce the necessary Dublin Core metadata for harvesting by service providers.[11]

Some of you may know that OCLC's first offices were located in the Main Library of Ohio State University, and adjacent to the campus is another well known international organization, Chemical Abstracts Service. Both Chemical Abstracts and OCLC were pleased to be included in planning for a major institutional repository at Ohio State called the Knowledge Bank. It will provide the academic community with services, resources support, plus communication conduits, and collaborative opportunities for advancing and extending distributive and distance education at The Ohio State University. The Knowledge Bank is part of that project. It is based on the notion that the generation, collection, indexing, and preservation of intellectual content is the most valuable University asset and would be even more valuable if it were maintained in an accessible digital repository.

Under the leadership of Director of Libraries Joe Branin, a group developed a model for the Knowledge Bank that depicts the flow of intellectual content from all areas of the University through submission, storage, and transformation by faculty into technology-enhanced learning modules that can be delivered to students at a distance or in on-campus classrooms. These

[10] *Wired News* (3 Aug 2002) [online]. Lycos, 2002. <http://www.wired.com/news/school/0,1383,54229,00.html> [August 2002].
[11] *(Ariadne (8 Jul 2002)* [online]. <http://www.ariadne.ac.uk/issue32/eprint-archives/> [August 2002].

learning modules are fed back into the Knowledge Bank for others to use and build on. In the broadest sense, the Knowledge Bank is envisioned as an enterprise-wide digital repository of intellectual content, with links to content stored in other locations.[12]

In the United Kingdom, the Joint Information Systems Committee (JISC) has established the FAIR (Focus on Access to Institutional Resources) Programme, which is a national initiative that is trying to encourage a new way of doing things. It will fund projects to support access to and sharing of institutional content. The outcome of FAIR is a set of projects supporting institutional management of e-prints and theses, leveraged by services that harvest the metadata and make it available more widely. OCLC is pleased to be involved in one of these projects, developing prototype knowledge organization services that will analyze and enrich the harvested metadata.[13]

The FAIR programme is inspired by the vision of the Open Archives Initiative (OAI) (http://www.openarchives.org), that digital resources can be shared between organizations based on a simple mechanism allowing metadata about those resources to be harvested into services. In the e-prints community this is realized through data providers who mount the e-prints (and who could be based in institutions, in subject groupings, or in some other way), and who disclose their metadata to a service provider, which again could be based in institutions, or could be subject based, regional, national or international. End-users can either search the particular

data provider of interest, if they know it, or can search the service provider, which will have gathered together the metadata from many data providers. The OAI protocol is one mechanism that can support this model, but there are others. The model can clearly be extended to include other kinds of objects, for example learning objects, images, video clips, finding aids, etc. The vision here is of a complex web of resources built by groups with a long-term stake in the future of those resources, but made available through service providers to the whole community of learning.

In the U.S., the Coalition of Networked Information and the Digital Library Federation are partnering with numerous individual universities, consortia and other organizations on projects involving digital repositories. Perhaps the most visible of these is the aforementioned Open Archive Initiative (OAI). The secretariat is managed by Herbert Van de Sompel and Carl Lagoze at Cornell University. Van de Sompel is well known for his work on the Santa Fe Convention, which defined an interface to permit an e-print server to expose metadata for the papers that it held; this metadata could then be picked up by federated search services or by other repositories that wanted to identify papers for copying .[14] OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. There are currently two types of participants in OAI: Data Providers administer systems that support the

---

[12] The Ohio State University. *Official Proceedings of the One Thousand Three Hundred and Eighty-Second Meeting of the Board of Trustees, Columbus, Ohio*, May 3, 2002.

[13] *Joint Information System Committee* [online]. <http://www.jisc.ac.uk/pub02/c01_02.html> [August 2002].

[14] LYNCH, Clifford A. Metadata Harvesting and the Open Archives Initiative [online]. *ARL Bimonthly Report 217*, August 2001. <http://www.arl.org/newsltr/217/mhp.html> [August 2002].

OAI-PMH (Protocol for Metadata Harvesting) as a means of exposing metadata; and Service Providers use metadata harvested via the OAI-PMH as a basis for building value-added services.[15]

OCLC has developed open source software to support OAI.[16] It is exploring harvesting metadata from several sources, and it will make some resources available as a service provider. This activity is currently experimental. A bigger question for the future is how OCLC can best support its members as they expose metadata under OAI. We are certainly interested in providing a secure basis upon which to build service provider offerings, and are keen to engage with our members in working towards new service models.

Permit me to mention one more repository. In the U.S., the National Institutes of Health (NIH) operates PubMed Central, which is a web-based repository for barrier-free access to primary reports in the life sciences. It archives, organizes and distributes peer-reviewed reports from journals, as well as reports that have been screened but not formally peer-reviewed. There are almost 50 journals participating in PubMed Central. By storing data from diverse sources in a single repository with a common format, PubMed Central makes the data more accessible and easier to use and opens the door to greater integration with related resources. Full text can be searched and relevant material located efficiently, regardless of its source. Material is presented to the user in a uniform

style while still clearly maintaining the identity of each journal.[17]

3) Digital Preservation and Archiving. One of the more vexing challenges for libraries and other knowledge institutions has been the preservation and archiving of digital resources. A number of initiatives continue to advance our understanding of this vital aspect of the digital university library.

In the United Kingdom, the Cedars project (April 1998-March 2002) explored digital preservation issues. Funded by JISC, and involving CURL institutions (Consortium of University Libraries), Cedars issued some very useful guides to preservation metadata, collection management and intellectual property rights .[18]
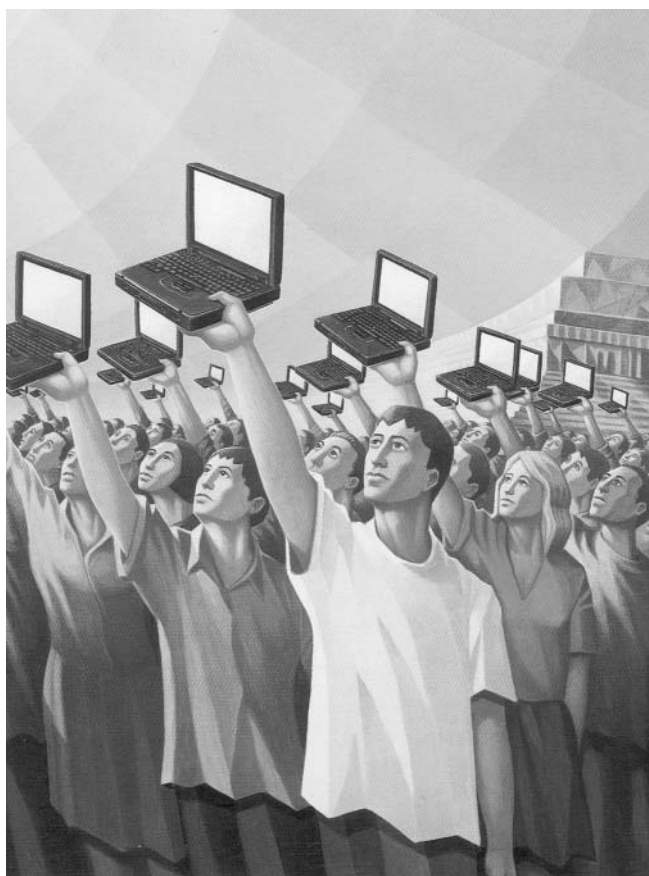
Another interesting research project currently under way is CAMiLEON, which is investigating emulation as a digital preservation strategy. Funded by the U.S. National Science Foundation and JISC (U.K.), the project is a collaborative effort of researchers at the School of Information, University of Michigan (U.S.) and the University of Leeds (UK). CAMiLEON stands for Creative Archiving at Michigan and Leeds: Emulating the Old on the New. Emulation has been proposed as a digital preservation strategy that would make it possible to retrieve, display and reuse digital documents with their original software by running obsolete software on future unknown systems. The project team is

[15] *Open Archives Initiative* [online]. <http://www.openarchives.org/index.html> [August 2002].

[16] OCLC *Research Activities and the Open Archives Initiative* [online].
<http://www.oclc.org/research/projects/oai/index.shtm> [August 2002].

[17] *Frequently Asked Questions* [online]. <http://www.pubmedcentral.gov/about/faq.html> [August 2002].

[18] CEDARS *: Curl Exemplars in Digital Archives* [online]. <http://www.leeds.ac.uk/cedars/index.html> [August 2002].

guidelines for the preservation of access to digital information and to provide a forum for cooperation.[20] In May 2001, the National Library of Australia and the Council on Library and Information Resources (CLIR) provided funding for a project to build a distributed and permanent collection of digital resources from the field of digital preservation selected from the PADI database. This project aims to identify significant resources in digital preservation early in their lifecycle. It also aims to facilitate the cooperative development of a distributed network of 'safekept' material with resource owners, or parties nominated by them, providing long-term access to their material. Materials that have been "safekept" have a visual brand attached to them.[21]

The Research Libraries Group and OCLC jointly sponsor the Preservation Metadata Working Group. Composed of leading experts from a variety of institutional and geographical backgrounds, the group is examining current practice in the use of preservation metadata, and developing a comprehensive preservation metadata framework applicable to a broad range of digital preservation activities.

testing the feasibility and effectiveness of emulation for preserving the intellectual content, structure, and "look-and-feel" of different types of digital materials.[19]

The National Library of Australia's Preserving Access to Digital Information (PADI) is a subject gateway to digital preservation resources. It is an international effort to facilitate the development of strategies and

Last May, RLG and OCLC released a new report on long-term preservation of and access to research materials in digital form, aimed specifically at those with traditional or legal responsibilities for the preservation of cultural heritage. It defines "trusted repositories" and addresses the seven attributes such repositories must have.[22]

---

[19] CAMILEON : Creating Creative Archiving at Michigan and Leeds : Emulating the Old on the New [online]. <http://www.si.umich.edu/CAMILEON/> [August 2002].

[20] PADI : Preserving Access to Digital Information [online]. <http://www.nla.gov.au/padi/about.html> [August 2002].

[21] BERTHON, Hilary; THOMAS, Susan; and WEBB, Colin. Safekeeping : A Cooperative Approach to Building a Digital Preservation Resource. D-Lib Magazine, January 2002, Volume 8, Number 1.

[22] RLG and OCLC Issue Final Report on Trusted Digital Repositories : Attributes and Responsibilities [online]. <http://www.rlg.org/pr/pr2002-repositories.html> [August 2002].

Trends in Digital
University Libraries

Permit me to say a few words about OCLC's new services in digital and preservation resources. First, in response to requests from the library community, we established a digital co-op as a resource for libraries and institutions seeking to digitize, preserve and publish electronic collections. Our second initiative is Digital Archiving. It facilitates capture of Web documents, creation of preservation metadata for digital objects, ingest of objects into the archive, and the long-term retention of these digital information assets. The University of Edinburgh is participating in the pilot version of this service. Third, we are establishing Digital and Preservation Resource Centers to help libraries convert their collections to digital form, provide online access to them, and preserve them for the future. We are building on the extensive experience of Preservation Resources, an OCLC division based in Bethlehem, Pennsylvania. We recently opened the OCLC Western Digital & Preservation Resources Center in Lacey, Washington. And, we are planning on opening a DPR center in Europe in the coming year.

OCLC is also offering software solutions to help libraries preserve and provide access to valuable resources. The OCLC Digital and Preservation Resources center in Bethlehem, Pennsylvania is now using Olive Software to help libraries provide full online searchable access to their historic newspapers over the web. The software builds an index of every article, photograph and page. A library can mount its digitized collection on its own server or host it on an OCLC server.

And, as this audience knows, another exciting project involving newspapers is under way in Mexico. The National Newspaper Library *(Hemeroteca)* of the National Autonomous University of Mexico (UNAM) is working with Cold North Wind, a Canadian firm, to digitize the entire microfilmed collection of 490 Mexican historical newspapers and periodicals held at the *Hemeroteca* in Mexico City. The project will create the single largest portal of Spanish-language historical newspapers on the Internet and will allow for access to be available to researchers and students in Mexico and around the world. Covering three centuries and over 20 million pages of news, the portal will allow a unique view of Mexican history.[23]

Clearly, preservation and archiving is the last, untamed frontier of the digital university library.

4) Metadata. The advent of digital resources is rapidly changing how metadata is created, modified and used by libraries. Metadata is becoming more oriented to retrieval rather than description because of the relative ease of access that electronic materials provide compared to physical materials. While digital materials present a number of new problems, they also offer the opportunity to develop new approaches to dealing with metadata. Libraries find that they can now provide digital access to repositories, archives and photo collections using new protocols and metadata schemes. If the special collections or at least a description are not digitally accessible, the chances of their use are seen as rapidly declining. The need to generate metadata for these materials, along with the need to merge metadata from many different sources is fueling four trends.[24]

---

[23] *The Hemeroteca of the National University of Mexico and Cold North Wind to Digitize Historical Newspapers and Periodicals* [online]. <http://www.coldnorthwind.com/news/news.html> [August 2002].

[24] JORDAN, J. and HICKEY, Thom. *Metadata Trends*. Paper presented at NIT Conference, Beijing, China, May 27, 2001.

Simplicity. Traditional library cataloging is relatively expensive to create. To be done properly, the relationships the item has with other items need to be identified, persons and institutions need to be ascertained and checked with authority files, and a fairly detailed description of the item needs to be created. Libraries primarily use MARC format, and there have been many attempts at simplification. For example, at the Library of Congress, the Program for Cooperative Cataloging represents an international cooperative effort aimed at expanding access to library collections by providing useful, timely, and cost-effective cataloging that meets mutually-accepted standards of libraries around the world.

Despite these efforts, as we all know, cataloging and metadata creation has not been getting any simpler. When the material described in the library's catalog is directly accessible to a user from a computer terminal, the cost to view items declines, and less descriptive information is needed for users to make informed decisions before attempting to view it. The user can click on the URL and see if that is what she or he wants.

The Dublin Core Metadata Initiative (DCMI) is the direct result of looking for a simpler way of describing a wide variety of resources by a wide variety of institutions. Stuart Weibel, an OCLC research scientist, is the Executive Director of the Dublin Core Metadata

Initiative. Libraries have been among the most active members of the DCMI work, but others such as museums, archives, and publishers have also been deeply involved. The Dublin Core is based on 15 elements (Figure 1).

One of the guiding thoughts behind the creation of these elements was that Dublin Core metadata was primarily for retrieval, not description. This simplification is one of the aspects of the Dublin Core that has led to its adoption by many diverse groups.

Simplicity has its virtues. It is cheaper and faster. It provides wider coverage. It is good enough for discovery, and good enough to interact with other systems.

Distributed. The uniqueness of special collections is also part of a trend towards distributed metadata. Metadata has always been created locally, but the ease in which information can now be widely shared is making distributed creation and use of metadata a distinct trend. Metadata manipulation is becoming common in so many systems outside of libraries that libraries need to be able to interact with those systems. The Internet and new protocols are making it easier for institutions to collaborate directly and are starting to change the balance of who is responsible for metadata.

OCLC is currently working on two new protocols - Z39.50 next generation and OAI Metadata Harvesting. Z39.50 International: next generation (ZING) is an initiative that is looking at carrying some of the structure of the Z39.50 protocol into some simpler approaches that are more appropriate for the web. Several strands

of activity are under way. This work is in early stages, but is attracting some interest. More can be found out at the Zing page at the Library of Congress. http://www.loc.gov/z3950/agency/zing/zing.html .

As noted above, the Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative Protocol for Metadata Harvesting is being developed to allow locally created metadata in a variety of formats to be shared. The transfer format is Dublin Core in XML. Each institution that wants to share some metadata places it under an OAI server that can respond to requests to send all or part of the file. The protocol allows publication of metadata catalogs that would be useful for special collections and for materials that are outside traditional library cataloging.
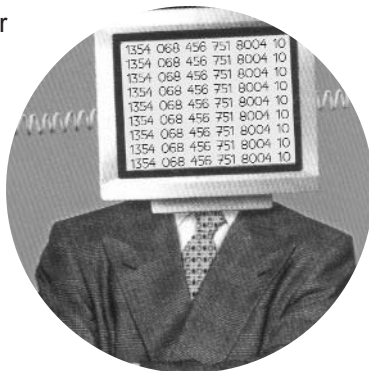
Automatic generation. The large Web search engines depend heavily on completely automatic indexing of Web pages. For much material of interest to libraries there is, however, the need for some manual intervention in the creation of metadata, even for purely digital resources. Automatic generation works best with simpler formats. It can be useful for libraries in extraction of fields, determining algorithms, and matching against known lists. Machine support could, for example, do the following to speed the creation of both simplified and traditional cataloging if the resource is available for harvesting:

- Extract embedded metadata from the resource
- Attempt to determine the language
- Suggest Dewey Decimal Classification numbers

- Perform sophisticated authority matching on names
- Create a skeleton record based on other information in the resource

Future metadata systems will enhance and extend this list, making it easier for humans to guide the metadata creation process while keeping the cost of metadata creation as low as possible.

Standards. Libraries have to deal increasingly with standards developed outside the library community. Standards enable disparate systems to work together. Librarians have always been aware of the importance of standards, but now have to cope more and more with standards developed either outside the library community (such as XML) or in association with other communities (such as Dublin Core).

The emergence of Unicode as a standard covering the vast majority of glyphs needed for various languages and scripts is an example of a standard that will result in new capabilities. Multilingual metadata and authority files are only now emerging as real forces because of it. The fact that the Java programming language handles Unicode strings natively is making it possible for many applications to start supporting arbitrary scripts. Until recently, the barriers for supporting this sort of capability were simply too high. At OCLC, we are very excited that the next generation of WorldCat, the OCLC Online Union Catalog, will support Unicode.

Standardization and codification are areas that libraries understand and where they make a contribution towards knowledge organization. To librarians it is second nature to understand the need and desirability of standard lists of subject headings and authority

files with information distinguishing individuals and institutions. They also understand and cooperate in the creation of rules on the application of these lists and the description of resources.

Those outside the library community are constantly relearning the lessons that librarians learned decades ago about the benefits of consistency and control. This is an area in which libraries can make a substantial contribution towards the development of more intelligent tools, as well as tools that can interact with libraries for the benefit of all.

The final portion of this paper turns to the future of scholarly communication. That future can be found in the classrooms of today's universities -the undergraduate and graduate students who are in the process of becoming the scholars of the future. In the decades ahead, they will inherit and then shape the course of scholarly communication. What are tomorrow's scholars doing today?

5) Information habits of university students. At the same time that scholarly communication and librarianship are experiencing instability, the ways students and scholars use the library, or do not use it, are also changing. Indeed, after the above review of the significant changes in scholarly communication and the way we run our libraries, it is tempting to call this final portion of my remarks, "meanwhile, back at the ranch."

Today, I would like to share with you the high-level results of a study of the information habits of college students in the U.S. that OCLC commissioned.[25] The study was done by an independent firm, Harris

---

[25] OCLC *White Paper on the Information Habits of College Students,* June 2002.

Trends in Digital
University Libraries

Interactive. It concentrated on the web-based information habits of college students and their use of campus library websites. The study surveyed 1,050 college students, ages 18-24, across the U.S. Here are some of the findings:

· About 80 percent of the surveyed students use Internet search engines for "every" assignment or "most" assignments
· About 50 percent use their library Web portals or their class Web sites
· 88 percent are less likely to use a site that charges a fee for information
· 58 percent stated that there is no difference in the reliability of information on websites with advertising and ad-free websites
· 20 percent said that ad-free websites have more reliable information
· 80 percent are more likely to seek help in person, compared to 50 percent who ask for help online or by telephone
· 80 percent use the library for web access
· 34 percent prefer electronic to print copies
· 50 percent said information on the web was not sufficient for their assignments

About 70 percent of students use the campus library website for at least some of their assignments, and 20 percent use it for most assignments.

During students' most recent electronic visit to the library's website, most used full texts of journal articles (67 percent), the library's catalog (57 percent), databases and journal indexes (51 percent) and electronic books (21 percent). About 89 percent of students use the campus library's print resources, including books, journals, articles and encyclopedia. In addition to using the library's print resources, they make photocopies from print resources and even print copies of electronic resources.

Students perceived the following barriers, whether genuine or not, to their successful use of library resources:

· Inability to access databases remotely due to password requirements and/or license restrictions
· Difficulty searching and navigating within the library and its website
· Costs of copying and printing at the library
· Shortage of knowledgeable librarians
· Lack of the customer orientation they have come to expect as consumers

In summary, college students are confident knowledgeable users of electronic information resources. They value access to accurate, up-to-date information with easily identifiable authors. They are aware of the shortcomings of information available from the web and of their needs for assistance in finding information in electronic or paper formats. At the end of the paper is a series of questions based on the finding for further exploration and discussion by librarians. Here is one example:

"Web search engines and portals provide single-point search access. Can libraries deliver websites that directly present and connect to relevant information from many databases and resources, including faculty, curriculum, and administrative websites?"

It would be very interesting to conduct a similar survey of college students in Mexico and other parts of the world. OCLC is going to use the data from the above study as a baseline. Future studies will compare

academic libraries progress on meeting students' information needs. In particular, progress in usage of campus library websites compared with usage for popular search engines and web portals will demonstrate how well libraries integrate with generally available web resources. Further studies will also analyze how successful librarians are at promotion, instruction and accessibility.

Conclusion

As we contemplate the progress we are making with our digital university libraries, I cannot help but think of how far we have come since Vannevar Bush first set forth a vision of what research and scholarship would be like in the future. Most of you will recall that Bush was a distinguished scientist who was among the first to articulate what would become known as the Information Explosion. In 1945, his seminal essay, "As We May Think" appeared in the *Atlantic Monthly*.

In that article he outlined a photo-electro-mechanical machine that he called the *Memex*. It was a desktop machine that would display text and pictures (from a microfilm library) at the press of a button. Users would create trails of knowledge along storable links, much like today's Web surfers. Bush called it an enlarged intimate supplement to his memory. That article has been read and reread by the people who have shaped the digital age we now live in, including the developer of the World Wide Web Tim Berners Lee. The current version of the *Memex* is the web. There will be more versions, more generations.

Clifford Lynch had some interesting observations on the current state of our digital libraries: "Digital libraries are as rich as our visions about how we can use and reuse digital information. They're as rich as the conversations we can imagine between books. They're as powerful as the linkages we can imagine creating as we amass material and in the digital world it becomes greater than the sum of its parts".[26]

I agree with Cliff. Libraries are on the verge of realizing some long-held dreams about providing access to information to people when and where they want it. With each new advance in selecting, in ordering, in cataloging, in resource sharing, and in reference, we come closer to realizing that shared mutual vision.

Or, as Vannevar Bush said, "There is a new profession of trailblazers, those who delight in the task of establishing useful trails through the enormous mass of the common record."

That profession of trailblazers is in this room, doing important work for research, scholarship and education, not just for today, but for future generations. ❧

Contributor	Format	Rights
Coverage	Identifier	Source
Creator	Language	Subject
Date	Publisher	Title
Description	Relation	Type

Figure 1: Dublin Core Elements

---

[26] LYNCH, Clifford. *Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information* [online]. Presented at Web-Wise 2000 Conference at Johns Hopkins University, March 20, 2002. <http://www.firstmonday.org/issues/issue7_5/lynch/> [August 2002].