

“Global Networking of Information—OCLC’s Strategy for the Future”

by

Jay Jordan, President and Chief Executive Officer

OCLC Online Computer Library Center

7th International Bielefeld Conference

5 February 2004

Good morning! It is an honor for me to participate in the 7th International Bielefeld Conference. I would like to take this opportunity to thank Dr. Norbert Lossau for inviting me to speak at this important gathering of library leaders in Europe. In my remarks I will discuss recent trends in the information environment, OCLC’s evolving strategy to help libraries and other knowledge organizations thrive in this environment, and some of our strategic partnerships in Europe.

The OCLC cooperative now connects more than 45,000 libraries in 84 countries and is actively engaged in the global networking of information. The cooperative is a truly international community. There are some 34,500 libraries in the U.S. that are participating in the OCLC cooperative. There are now approximately 8,800 libraries in 83 countries outside the U.S. that are participating. There are about 3,000 libraries, primarily institutions of higher education, participating in OCLC in Asia Pacific... There are approximately 800 participating institutions in Canada... There are approximately 678 participating institutions in Latin America and the Caribbean. In Europe, the Middle East and Africa, there are approximately 4,300 institutions participating in OCLC PICA. Our cooperative is becoming worldwide, and the global networking of information is both a theoretical and practical issue for us.

Permit me to quickly review what the cooperative does. OCLC is a nonprofit, membership organization. It is a dot.org, not a dot.com. We are organized for the public purposes of furthering access to the world’s information and reducing library costs. We provide services to libraries in these areas: cataloging, resource sharing, reference, and digital collection management and preservation. OCLC is also involved in research.

OCLC.org is today a worldwide cooperative, and the global networking of information is both a theoretical and practical issue for us.

For example, resource sharing is becoming increasingly international. On January 5, the Statsbiblioteket Århus in Denmark generated the 132 millionth online interlibrary loan request on the OCLC Interlibrary Loan service. It is a public library for research and studying and serves as the main library for the university of Aarhus and the University Hospital of Århus. They sent their request to five public libraries in the U.S. for a book on tape. (Unfortunately, it was not filled.) More than 7,000 libraries use the interlibrary loan service, which is observing its 25th anniversary in 2004. Libraries used the service to arrange 9.3 million interlibrary loans last fiscal year.

Last year, the Office of Research observed its 25th anniversary. A silver anniversary retrospective appears in the latest OCLC Annual Report. The Office of Research was established by OCLC Founder Fred Kilgour in 1978. Today, it is a resource for OCLC, for member libraries, and for the wider information community. The mission of the Office of Research is to expand knowledge that advances OCLC’s public purposes of furthering access to the world’s information and reducing library costs. This

mission is achieved through the integration of computer, library, and information sciences into research activities including experimentation, prototyping, standardization advancements, studies, and research collaborations. Let's look at a few of their current activities.

OCLC Researchers also are involved in the Open Archives Initiative, more commonly known as OAI, which promotes interoperability standards to facilitate dissemination of content. Lorcan Dempsey Vice President, OCLC Research, serves on the OAI steering committee. In the OAI space, universities are experimenting with institutional repositories that gather as much of the intellectual output of an institution as possible into a searchable online collection. OCLC is also working with Dspace, the open-source repository software created by MIT and Hewlett Packard. In fact, we created the software that Dspace uses to support its OAI capability. We have two open-source software programs—OAI Cat and OAI Harvester—that implement OAI protocols for data storage and harvesting in support of institutional repositories.

The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. DCMI is a distributed organization. On the screen is a list of the current members of the board of trustees. OCLC Research is the administrative host of the Dublin Core Metadata Initiative. Makx Dekkers is the Managing Director of DCMI, and he is now based in Barcelona. DCMI's activities include consensus-driven working groups, global workshops, conferences, standards liaison, and educational efforts to promote widespread acceptance of metadata standards and practices. OCLC Research Scientist Stuart Weibel was the first executive director of the Dublin Core, and he was instrumental in obtaining the Dublin Core's endorsement as an ISO standard. Under his leadership, the DCMI became an international program, with participants in 50 countries and translations in 20 languages. DCMI participants are leaders in digital library and Internet research activities around the world.

Last August, OCLC Research released a FRBR algorithm as open source software. It is now available at the OCLC Web site. With the algorithm it is possible to write computer programs that will process MARC records and generate sets of records which would be grouped for display as single works corresponding to the IFLA FRBR concept of a work.

OCLC is making this open-source algorithm available to the library community as a public service and will provide updates and other communications on its ongoing development.

Also last August, Die Deutsche Bibliothek, the Library of Congress and OCLC signed a memorandum of understanding that will lead to the development of a Virtual International Authorities File. OCLC will use its matching algorithm software to process bibliographic and personal name authority records from the retrospective files of LC and DDB to generate the initial Virtual International Authorities File. The VIAF will be updated regularly with metadata harvested automatically from the participating national authority files using OAI protocols. This is a landmark in international library cooperation that will closely involve OCLC research scientists.

Let me turn now to some of OCLC's activities in promoting the evolution of librarianship through increased advocacy for libraries.

Another way that libraries are meeting the challenge of e-content can be seen in the ePrints UK project in the United Kingdom. We are pleased to be working with a Joint Information Systems Committee (JISC) funded initiative, ePrints UK, as part of its national FAIR (Focus on Access to Institutional Resources) program. We are developing web-accessible applications in knowledge organization and name authority services to enrich metadata records for ePrints UK. Our partners include The Resource Discovery Network (University of Bath) and the University of Southampton. E-prints UK plans to develop a national service through which the higher education community can access the collective output of e-print papers available from compliant Open Archive repositories provided by UK universities and colleges.

OCLC researchers tell me that another growing trend involves “Web services.” These services are a suite of protocols that define how requests and responses between software applications should be encoded and transferred over the Web.

The Metadata Switch shown here enables machines to talk to machines. The one machine says to the other, “Please send me a Dewey number for this record.” The other machine says, “here’s your Dewey number.” While it’s a bit more complicated than that, it is an example of the kind of services that we will be seeing more of as the semantic Web becomes reality.

OCLC Research is also operating an experimental Electronic Theses & Dissertations Project, which uses the Open Archives Initiative’s Protocol for Metadata Harvesting (OAI-PMH) to create a database of metadata for electronic theses and dissertations. These records are available for harvesting as an OAI set. On the screen you can see some of the libraries in Europe that are participating in this experiment—their theses and dissertation records are available for harvesting. OCLC is also a member of the Networked Digital Library of Theses and Dissertations whose mission is to improve graduate education by developing accessible digital libraries of theses and dissertations. This is a major effort to expose a rich set of resources that heretofore have been unavailable to scholars around the world.

Let me turn now to some recent trends in the information environment. In January, we made available on the OCLC Web site a new publication, “The 2003 OCLC Environmental Scan: Pattern Recognition.” As part of our strategic planning process, we took an in-depth look at the environment in which the OCLC cooperative seeks to thrive. The 2003 Environmental Scan includes 100 interviews with information professionals, a review of 250 articles and papers and extensive global research. It was intended to stimulate strategic discussion with our Board of Trustees and Members Council. It met that objective with great success. The Scan is an important resource for our planning in both the near and long terms. We are making the Scan available on the OCLC Web site as an online resource. Print copies are available for a nominal fee.

The report explores trends in a variety of landscapes, including: social, economic, technology, research and learning, and library. It looks at changes that are likely to have the most influence on the lives of information consumers. The landscapes are highly interconnected, and trends in one landscape are shaping the future in others. Let’s take a look a highlights of the social landscape.

In the social landscape, the report explores the rapid move of information consumers to self-service and self-sufficiency. These information consumers are spending more time online doing things for themselves, whether for banking, shopping,

travel, research or entertainment. They are comfortable with Web-based information and content, such as Google. They are generally satisfied with the results that they get, whether from Google or from their other online activities, even though the results may not be as authoritative, reliable and accurate as librarians would like. Today's information consumers, especially young adults, expect seamless access to whatever they want whenever they want it. Let's look at the economic landscape.

This chart from the report shows 75 percent of the world's library spending is concentrated in five countries—the United States, Japan, the United Kingdom, Italy and France.

Public funding for libraries and allied organizations will likely continue to decline or remain at low levels for several years. No matter what the worldwide economy does in 2004, the saga of limited resources versus unlimited needs will continue.

Libraries must reexamine their internal resource allocations in an increasingly digital world. Most important, libraries and allied organizations must be able to do a better job of demonstrating their value more explicitly to their funding agencies. Now, let's go to the library landscape.

The library landscape section of the report examines the people, content, issues, and technology that are affecting libraries today. For example, many librarians and highly trained information professionals will retire within 5-10 years. It will be necessary to fill those positions or reallocate resources to new types of jobs resulting from libraries taking on new roles, such as in e-learning and scholarly communication. There is an increasing need for libraries to interconnect with non-library systems such as Blackboard or WebCT or campus portals. Libraries will increasingly have to manage varied and complex digital content, and there will be increased focus on standards for repositories, preservation, content packaging and exchange, and metadata. Over the next few years, it is likely that most new protocol development will be in a Web services context, which is to say, business processes delivered over the Web based on industry standards. That's just a sample. I urge you to read the report and discuss it with your colleagues.

In 2000, we shared with libraries our three-year strategic plan to extend the OCLC cooperative and transform WorldCat into a globally networked information resource through new services and a new technological platform. As we begin 2004, I am pleased to report significant progress. We have modified our governance structure to make it more inclusive. We have developed and are now operating the new services that you see listed here on the screen... Digital collection and preservation services, including a digital archive; a 24x7 virtual, cooperative reference service (QuestionPoint); an integrated metadata/cataloging system (OCLC Connexion); and a public access computing portal for public libraries (WebJunction). We are also well along in the transformation of WorldCat, and I will have more to say about that later in my remarks. Now, let's look at some of our new services.

netLibrary continues to make progress since its acquisition by OCLC in January 2002. Its collection now contains more than 60,000 titles, representing copyrighted titles from 400 publishers, with 64 percent of titles published in the last three years. About 8,670 libraries presently use netLibrary eBook content and tools, including 598 institutions in 33 countries outside the U.S. Last week, the Northwest Academic Libraries (NoWAL) consortium of the United Kingdom University and Higher Education Institution libraries has agreed to share access to nearly 16,000 eBook titles from

netLibrary and OCLC PICA. More than 165,000 users in the North West of England will have immediate onscreen access to the collection. This is the first academic consortium purchase of eBooks in the United Kingdom. It forms the largest eBook collection serving the highest FTE number in Europe. Participating libraries are in Cheshire, Cumbria, Greater Manchester, Lancashire and Merseyside.

About 18 months ago, with the Library of Congress, we launched QuestionPoint, a virtual reference desk collaborative service. To date, about 1,000 libraries are using QuestionPoint in 19 countries. The service has logged nearly 100,000 interactions between reference librarians and patrons, including some 25,000 chat sessions. The Global Knowledge Base now contains nearly 4,000 question-and-answer records.

On the screen is a locally branded version of QuestionPoint. It's called Aladin. About 60 public libraries in the Netherlands are now participating, and the Netherlands Public Library Association has created a web site that makes it possible for any public library user to access Aladin/QuestionPoint.. With QuestionPoint, we are developing a new model for collaboration in reference services.

As part of our new digital collection and preservation services, OCLC offers software tools to help libraries be digital publishers. The first is CONTENTdm. You can use this software to add metadata and post Web exhibits of digital materials. You can license it for your own server or use a hosted solution from an OCLC server. More than 100 institutions are now using CONTENTdm to support a variety of digital collections, including the municipal archive of Veenendaal, a city in the Netherlands near Utrecht, which is shown in this slide. The archive is mounting the collection of a local photographer on the Web using CONTENTdm. The digitized collection includes approximately 75,000 unique objects, mainly images, audio, video, text, maps and some newspapers will be searchable on the Web.

There are now more than 1 million objects available on some 100 CONTENTdm sites for potential loading into WorldCat through automatic harvesting of metadata. Several months ago, the Indiana Historical Society became the first institution to register its collections for automatic harvesting, and OCLC is now starting to harvest records for the objects. Users will be able to link from the record back to the object. Clearly, this adds a new dimension to WorldCat.

The second tool is Olive software, which incorporates optical character recognition and automatic XML tagging for improved searching and image files to be viewed by users. The British Library is using Olive to digitize and provide access to The Penny Illustrated Paper, which chronicles Victorian life through 40,000 pages and 500,000 images. The University of Oxford is also using Olive to support its Forced Migration Online Digital Library, which provides instant access to a wide variety of online resources dealing with the situation of forced migrants worldwide. This past year, we also issued a new edition of the Dewey Decimal Classification system.

First published in 1876, the Dewey Decimal Classification system is a general knowledge organization tool that is continuously updated to keep pace with knowledge. It is used in 135 countries and has been translated into 30 languages. In 2003, OCLC published the 22nd edition of the Dewey Decimal Classification. Edition 22 is the first edition of the DDC to be produced in the web environment, which has enabled the editorial team to update the DDC continuously and provide those updates regularly to our

users. Work is underway at Die Deutsche Bibliothek on the first German translation of the DDC. The translation is based on DDC 22, and will be the first outside of the English-language standard edition to be published simultaneously in web and print versions. The Dewey editorial team worked closely with Die Deutsche Bibliothek staff in Germany and colleagues in Switzerland and Austria to improve the geographic tables and history developments for those areas in DDC 22.

In France last year, OCLC PICA implemented its Central Library System for cataloging and interlibrary loan in 400 libraries of all 110 French universities. In Germany, OCLC PICA worked with the Gemeinsamer Bibliotheksverbund (GBV), the Hessischer Bibliotheksverbund (HEBIS) and Die Deutsche Bibliothek (DDB) to bring the OCLC PICA local library and central cataloguing systems to more German libraries. They migrated these partners to CBS4, the new version of the OCLC PICA Central Library System. GBV is also using the OCLC PICA IPORT portal to serve its 28 local systems. IPORT makes it easier for a user to search OPACS and local and international databases. I have already mentioned that Die Deutsche Bibliothek is translating the 22nd edition of the Dewey Decimal Classification into German. It is a great honor to have the National Library also participate in the OCLC Interlibrary Loan service.

On January 23 OCLC PICA and Dynix announced a programme of technical cooperation that will focus on seamless interaction between Dynix local systems using the Horizon Information Portal (HIP) and the OCLC PICA Central Bibliographic System (CBS) used in many European countries for the management of bibliographic union catalogues. Both organizations will utilize international standards and will cooperate in joint testing between their systems. The initial priority is to facilitate end-user access to OCLC PICA national and global bibliographic resources including the union catalogues of GBV, HEBIS and DDB, the GGC (Dutch Union Catalogue), and OCLC WorldCat through pre-configuration of these resources in the Horizon Information Portal. Through this programme, both Dynix and OCLC PICA are confirming their support of international standards and open systems, to enable libraries and their end users to increase efficiency and improve access to European resources and scholarship.

Last August, OCLC PICA acquired the LIBPAC V3.Web, the interlibrary loan (ILL) requesting and management system LIBPAC will manage the V3.Web service on behalf of OCLC PICA. V3.Web is used by more than 70 UK public library authorities. It enables interlibrary loan requests to be made nationally (with compliant systems), in Europe and internationally. The V3 database comprises more than 5 million bibliographic records and has location information leading to over 40 million volumes of public and special libraries in the United Kingdom. This system is being migrated to a Leiden platform where it will be linked to the Dutch and German ILL systems and to the global OCLC ILL service through the ISO standard.

Earlier, I stated that a key part of our strategy involves the WorldCat database. It is certainly one of the great examples of the power of library cooperation, and it is a model that has continued to provide value over three decades of continuous technological change. It is today the world's largest bibliographic database. It now contains more than 53 million records, and there are now more than 888 million location listings. Even though WorldCat is one of the great tools of the library world, most of us in this room would agree, however, that it needs to continue to change to remain vital. We are in the midst of a major project to transform WorldCat from a bibliographic database and online

union catalog to a globally networked information resource of text, graphics, sound and motion.

We are now migrating WorldCat to a new platform, based on Oracle technology. We will be off all proprietary databases and the old systems by the third quarter 2005.

The new WorldCat will support not only MARC, but Dublin Core and IFLA's Functional Requirements for Bibliographic Records. And, perhaps most important for a global library cooperative, the new WorldCat will also support Unicode, which will give us the foundation to provide access to information in a number of languages and character sets. WorldCat will also link with other knowledge hubs.

For example, OCLC members now have access to the OCLC PICA Dutch Catalogue GGC. This database contains some 18 million bibliographic and authority records for European publications and Dutch language materials. You can access the Catalogue through the OCLC Connexion browser. You can pull records from the OCLC PICA GGC catalogue and input them into WorldCat. This is the first database outside our computer room in Dublin to be linked through Connexion. Another benefit is that approximately 300 Dutch libraries now qualify for OCLC membership.

As a final example of how we are transforming WorldCat, we are now engaged in the largest, and potentially one of the most significant, test programs in the history of the cooperative—the Open WorldCat pilot. In partnership with the Google search service, we are making a subset of 2 million abbreviated records from WorldCat available on Google, with links to the Web-based catalogs and sites of some 12,000 academic, public and school libraries participating in OCLC. In this pilot, an information seeker who starts a search using Google could end up finding the items needed in a nearby library. We launched the pilot after extensive consultation with leaders in the library community. WorldCat records began to appear in Google search results in mid-December. There are now about 133,000 WorldCat records in Google. Yahoo and AOL are also starting to show WorldCat records on their sites; they get these records directly from Google, which is the only search service authorized to crawl our subset of records.

These records are a different type of content than Google has indexed before, so we're working with them to improve the search results ranking of the WorldCat records, which are labeled "Find in a library." There is widespread interest in and support for this pilot. Indeed, of the 12,000 libraries automatically included in the pilot, only 174 have withdrawn and another 34 libraries -- including several federal and state libraries -- have asked to join. We've already held some market research sessions. Most people we talked to saw value in pilot, especially public librarians who welcomed the opportunity to market themselves in another way. Private academic librarians, however, have expressed concern about being inundated with requests from users that they could not serve. We will conduct research in the United Kingdom in the near future. Throughout the pilot, we will be listening closely to libraries and their users, and in June we will decide whether to proceed with an ongoing service.

As you have seen this morning, OCLC has a very ambitious strategy for the global networking of information. Big changes are ahead. In time, the new WorldCat will link to other knowledge hubs and nodes around the world. There will be a global network of catalogs and metadata, including WorldCat and library collections from around the world.

Our technological platform will rely on open systems architectures and adhere to technical standards that promote the cost-effective, worldwide sharing of information across platforms, scripts, languages and cultural materials.

Going forward, cooperation will be the key.