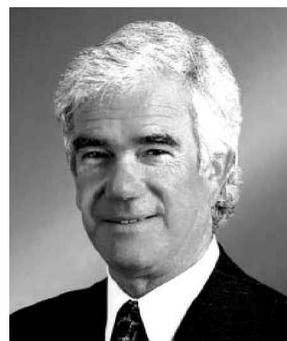


E-content in the OCLC Global Library Co-operative

Based on a paper given at the 26th UKSG Conference, Edinburgh, April 2003

In the next stage of the digital library, co-operation will be more important than ever. OCLC, the world's largest library co-operative, intends to help institutions manage their e-content from e-journals to e-books to emerging collections of digital objects. WorldCat, co-operatively built and maintained by OCLC and its member libraries, is being transformed from a bibliographic database to a knowledge hub that will help people identify and connect diverse content in a networked world. Emerging standards, new linking tools, and new techniques for harvesting metadata are among the topics explored.



JAY JORDAN
President and CEO, OCLC

Introduction

The OCLC (Online Computer Library Center) vision statement is 'OCLC will be the leading worldwide library co-operative, helping libraries serve people by providing economical access to knowledge through innovation and collaboration'. This co-operative is going to provide libraries around the world with flexibility and choice in services and programmes. Ultimately, we want to provide more value to more libraries, to be more diverse, and to be more inclusive.

More than 43,000 libraries are linked through the OCLC network including 8,000 in 85 countries outside the US which represented approximately 22 million US dollars last fiscal year, or 12 percent of total revenues. There are about 3,000 libraries in Asia Pacific, 800 in Canada, and 310 in Latin America and the Caribbean. In Europe, the Middle East and Africa, there are approximately 4,100 institutions participating in OCLC.

Since 1981, 43 UK institutions have become OCLC members. They have cataloged millions of items over the past two decades. Last year, they set holdings for some 867,000 items, and input nearly 141,000 original cataloging records.

WorldCat

WorldCat is the world's foremost bibliographic database. Created and maintained by member libraries and OCLC, it spans over 4,000 years of recorded knowledge, and it grows as our knowledge grows. A new record enters WorldCat every 15 seconds. WorldCat is also the access point for a large quantity of e-content. Last November, the British Library, an OCLC member since 1985, input the 51-millionth record into WorldCat.

WorldCat is the host database for the CONSER program, which began in the 1970s at the Library of Congress as a project to convert manual serial cataloging into machine-readable records and has evolved into an ongoing program to create and maintain high quality bibliographic records for serials. In 1997, CONSER became a bibliographic component of the Program for Co-operative Cataloging. The 41 institutions that are members of CONSER include the National Library of Wales, the National Library of Canada, the Library of Congress and research and university libraries. They input, authenticate, and modify serial cataloging records on OCLC or contribute original records via FTP. This co-operative serials cataloging program has resulted in a database of approximately 978,000 quality serials records. Last

year, CONSER members authenticated some 30,000 records. CONSER is also delving into issues surrounding e-journals and e-resources and is likely to be involved in a new international effort – SUNCAT. At OCLC we hope that this initiative will strengthen and further internationalize the CONSER program for the benefit of researchers and scholars worldwide.

FirstSearch

The OCLC co-operative makes a large amount of e-content available through the FirstSearch service, which provides access to over 60 databases and over 1.2 million full text articles in some 4,400 e-journals. At the heart of FirstSearch is WorldCat. Using FirstSearch, materials in a library's collection are highlighted in results of searches in dozens of leading databases. Users can link from WorldCat to e-books, full text in databases, online public access catalogs and web sites. FirstSearch is now searchable in six languages – English, Spanish, French, Japanese and two versions of Chinese. Librarians can now set up FirstSearch so their users can view Z39.50 catalogs of other libraries and can actually see the shelf-status of an item. You can also link to JSTOR archives and Infotrieve for document delivery. And, library patrons can launch their own interlibrary loan requests from FirstSearch.

NetLibrary & Digital Archive

In January 2002, OCLC acquired the assets of netLibrary, a leading provider of electronic books. Today, the netLibrary division provides more than 50,000 copyrighted titles from 315 publishers, 88 percent of which have been published since 1990. About 7,300 libraries presently use netLibrary eBook content and tools.

The OCLC Digital Archive provides long-term access and preservation for your digital collections and associated metadata. You can manage your resources and choose to share your collections/content with the world or limit access for your own administrative purposes. Objects can be added to the Archive one at a time, using the Digital Archive harvester and preservation metadata tools, or in batches. Current participants include the US Government Printing Office (GPO), and the State Libraries of Ohio, Michigan, and

Connecticut. The Library of Congress, National Library of Medicine and the Smithsonian Institution are participating in a trial of the Digital Archive. The Digital Archive follows the Reference Model for an Open Archival Information System (OAIS), which has been adopted as an ISO standard. Dissemination Information Packages from the Digital Archive conform to the Metadata Encoding and Transmission Standard (METS). To date there are about 1,000 objects in the archive.

Office of Research

This year the OCLC Office of Research is observing its 25th anniversary. OCLC researchers represent libraries in the international standards community and direct Dublin Core. This standard has been translated into 30 languages. Seven national governments, representing 130 million people, have adopted the Dublin Core as the national metadata standard for government resources. It received NISO approval in 2001. Researchers have also been studying the IFLA Functional Requirements for Bibliographic Records to identify the issues involved in applying these standards to large online systems and databases such as OCLC's.

In the UK, we are working with a JISC funded initiative, ePrints UK, as part of the FAIR program to develop web-accessible applications in knowledge organization and name authority services to enrich metadata records. Our partners include The Resource Discovery Network (University of Bath) and the University of Southampton. We are also involved in Dspace at MIT. Dspace is the open source repository software created by MIT and Hewlett Packard. OCLC created the software that Dspace uses to support its OAI capability. In January, MIT Libraries announced initial development of the DSpace Federation with six major research universities: Columbia University, Cornell University, Ohio State University, and the Universities of Rochester, Toronto, and Washington. These institutions are also OCLC members.

Since 1998, the OCLC Office of Research has been analyzing the web. There are currently about 9 million websites, and although growth is slowing, there could be as many as 10.4 million in five years. The surface web is predicted to grow from its current 2 billion documents to over 9

million documents over the next few years. The deep web, which encompasses that portion of the web that is not accessible through open search engines, e.g. library OPACs, reference databases, JSTOR, US Census, etc. could be vastly larger than the visible web.

OCLC has also studied trends in the production of scholarly materials. We looked at books, journals, scholarly articles, e-print archives, theses and dissertations, and course management materials. In general, library book spending is down and print journal publication continues to decline. Only e-materials are going up, and they are going up dramatically. What does all of this mean for libraries?

OCLC researchers tell me that the next big wave is going to be 'web services', which is a suite of protocols that define how requests and responses between software applications should be encoded and transferred over the web. These services are really an extension of the concept of the 'semantic web', that was developed by Tim Berners-Lee .

In the world of the semantic web, computers will find the meaning of semantic data by following hyperlinks to definitions of key terms and rules for reasoning about them logically. The resulting infrastructure will spur the development of automated web services such as highly functional agents. Ordinary users will compose semantic web pages and add new definitions and rules using off-the-shelf software that will assist with semantic markup. One example is the Metadata Switch project which enables machines talking to machines to exchange metadata such as Dewey numbers.

EUCAT

OCLC PICA is developing EUCAT. This is a Pan European Index of Union Catalogs which provides a quality catalog based on metadata, with duplicates identified and grouped and with authority control of authors, subjects and other headings ensuring consistent indexing and recall. The main focus of EUCAT is as a discovery tool, linking to the contributing union catalogs or individual catalogs for services, in particular interlibrary loans and linking to licensed full text services. Document supply services, enriched contents, reviews, abstracts and e-books may be accessed directly from EUCAT or from the

participating union catalogs. OCLC PICA makes EUCAT available through different services, principally PiCarta and PubliekWijzer. It will also be possible to access EUCAT via external interfaces and portals using a search protocol, in particular ZING/SRU or SRW, Z39.50 and openURL

EUCAT currently consists of the Dutch Union Catalogue, which contains holdings of 642 libraries associated with 14.5 million bibliographic records, the libraries of the North German States (Gemeinsamer Bibliotheksverbund GBV) and the Newspaper Index of the Deutsche Bibliothek (Zeitschriftendatenbank ZDB). EUCAT has 20 million GBV records with 37 million holdings from 400 libraries. As a first step in broadening the coverage of EUCAT, the European holdings from WorldCat will be loaded as a mirror copy into EUCAT. OCLC WorldCat currently contains 21 million holdings from 430 European libraries. Arrangements for connecting to various inter-library loan systems are currently being investigated, as is the determination of the business models required to ensure the widest possible co-operative participation in EUCAT.

EUCAT as a centralized index to European library resources provides fast, relevant and comprehensive searching through consistent indexing. The co-operatively built index accommodates the diversity of European cataloging practices, languages and formats. It links to union catalogs, local catalogs and online providers for services. It is also capable of being linked to other services, e.g. e-learning environments, local library web pages, and from abstract and indexing databases for European holdings. EUCAT and WorldCat can provide a gateway to library resources that rivals Google as a gateway to internet resources. Unlike Google, the library resource will lead to an available copy somewhere, the quality of resources retrieved is more consistent and the searching capability yields more precise and complete results.

Conclusion

As we have seen, e-content in the OCLC co-operative begins and ends with WorldCat. This unique database is at the heart of our strategy to extend the OCLC co-operative to the global library community. As we look back across 51 million

records and 32 years, we can see that co-operation has taken us a very long way indeed.

The WorldCat of today is a late twentieth-century knowledge map. It is essentially an electronic version of the card catalog, which itself represented the apex of early twentieth-century knowledge mapping. Now, we are on the verge of a twenty-first-century knowledge map, one that builds on technologies that were only dreamed of when Frederick G. Kilgour, OCLC's founder and first president, hooked up the first terminal to WorldCat in 1971.

Migrating WorldCat to a new platform, based on Oracle technology, will enable it to move beyond bibliography and its current text limitations to include graphics, sound and motion. It will become an even more valuable resource, not only for the professional librarian, but for the library user as well. The new WorldCat will support not only MARC, but also Dublin Core and IFLA's Functional Requirements for Bibliographic Records. And, perhaps most important for a global library co-operative, it will also support Unicode, which will give us the foundation to provide access to information in a number of languages and character sets. It 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. Going forward, 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.

The new WorldCat will include new forms of information and contributors. It will contain not only descriptions of library holdings, but links to digital objects in other knowledge repositories, such as museums, archives, private collections, and historical and professional societies. It will interweave the worldwide web with the physical and electronic media of the world's knowledge institutions.

Jay Jordan

President and CEO

OCLC (Online Computer Library Center)

E-mail: Jordan@oclc.org
