Nancy E. Elkington
OCLC Online Computer Library Center, Inc.

OCLC Research: Past, Present, and Future

Abstract
From its modest beginnings in the mid 1970s as a small research and development unit of Online Computer Library Center, Inc. (OCLC), OCLC Research has evolved to become a significant research organization, a preeminent laboratory for the exploration of innovative uses of library data and information system technology, and a key agent of change, advancing the state of the art within OCLC and in the global library and information science community. OCLC Research is one of the world's leading centers devoted to exploration, innovation, and community building on behalf of libraries, archives, and museums and is dedicated to helping memory institutions more effectively serve users of information, information systems, and cultural heritage collections.

Keywords
OCLC; Research; RLG Partnership; WorldCat; Library & Information Science; Collective Collection; Identifiers; Special Collections; Terminologies; Web Services

ORGANIZATIONAL FOUNDATIONS

Founded in 1967, OCLC is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing the rate of rise of library costs. More than 69,000 libraries in 112 countries and territories around the world use OCLC services to locate, acquire, catalog, lend, and preserve library materials. OCLC and its member libraries cooperatively produce and maintain WorldCat, the OCLC Online Union Catalog.

The Research and Development Group was formed in 1974 under the leadership of James E. Rush. The OCLC Research Department was formally established within OCLC as a distinct administrative unit on July 1, 1977, and expanded over the next several years by manager W. David Penniman Neal Kaske (1981-1983), who championed the name change to the Office of Research and led the staff toward a deeper understanding of online catalogs while also securing project funding from the Council on Library Resources and the National Science Foundation. From 1984 to1986 Michael J. McGill served as Acting Director and put in place many of the outreach efforts that served to more widely share the work of the division with colleagues elsewhere. Martin Dillon followed as Director from 1986 to 1994; during his tenure the Office of Research shifted its research focus more directly onto the issues facing the OCLC membership. Research Scientists Edward T. O'Neill and Thomas B. Hickey were co-acting directors from July 1993 to May 1994. In 1994, Terry Noreault ([Terry Noreault promoted to Vice President, 1999](#)) was named Director, and in 1999 was named Vice President, Office of Research and Special Projects, dedicating the division "to research that both explores the place of the library in the changing technology environment and develops tools that enhance the productivity of libraries and their users." Thom Hickey, Chief Scientist, was acting Director from 2000 until Lorcan Dempsey (Lorcan Dempsey to head office, 2001) took up the reins of leadership in 2001. When RLG, Inc. (formerly the Research Library Group) and OCLC became a single agency in July 2006 (RLG membership approves move, 2006), RLG's programmatic activities, programs-related personnel, and partner relationships with 150+ leading research-oriented libraries, archives, and museums became part of an expanded division at OCLC initially called OCLC Programs and Research, but now renamed OCLC Research.

CURRENT ORGANIZATION

The Vice President of Research (Lorcan Dempsey, also Chief Strategist) reports to the President and directs the OCLC Research division. The division's formal mission statement reads, "The mission of OCLC Research is to expand knowledge that advances OCLC's public purposes of furthering access to the world's information and reducing

library costs" (OCLC, n.d.a). OCLC Research's work serves two primary audiences: (a) the OCLC membership and the global library community, and (b) the RLG Partnership, a transnational group of libraries, archives, and museums supporting research and scholarship (Michalko, 2009). The Vice President, RLG Programs (James Michalko) reports to the Vice President of Research and manages work designed to respond to the needs of the RLG Partnership. The two Vice Presidents serve as part of OCLC's Senior Leadership Team and jointly develop a dynamic and impactful work agenda that is responsive to the needs of the community. The division is composed of nearly 50 staff, including research scientists, program officers, software engineers and architects, a user interface designer, a project manager, and administrative staff, all of whom work in teams in support of the active work agenda.

THE PAST 10 YEARS: CHALLENGE AND CHANGE

OCLC Research has long focused its attention on addressing complex, often intransigent challenges facing libraries and their users (Godby & Richardson, 2003). In the past decade, many projects have been undertaken and completed and most bear the distinctive hallmarks of innovation: leadership, insight, invention, and action. Four initiatives are singled out because they can be seen as exemplars of the quality of investigation and depth of investment that has been made on behalf of the community at large: Dublin Core, PREMIS, VIAF, and Virtual Reference.

Dublin Core

When the World Wide Web was still in its infancy, one of the earliest challenges facing libraries was how to describe resources that lived only in the Web environment. The first Dublin Core workshop was held in 1995 and was cosponsored by OCLC Research and the University of Illinois' National Center for Supercomputing Applications (NCSA). The goal of the workshop was to agree on a simple, modular, extensible metadata scheme for Web-based resources (Weibel, 2005). The result was a core set of data elements for resource discovery.

Since then, the original element set has been expanded, extended, translated, and profiled in myriad ways, always in keeping with the original intent. The core set is an international standard (International Standards Organization, n.d.) and has been adopted in some form by at least seven national governments, by private industry, and by many libraries and other cultural memory institutions around the world. Of all OCLC Research endeavors, this one has had the most sweeping impact on the information community.

In January 2009, the Dublin Core Metadata Initiative (DCMI) established itself as an independent, public, not-for-profit company limited by guarantee in Singapore. DCMI is now the parent organization under which committees and groups work to develop and "support interoperable metadata standards that support a broad range of purposes and business models" (Dublin Core Metadata Initiative, 2009). The DCMI Web site (www.dublincore.org) is managed by the National Library of Korea (http://www.nl.go.kr/nlmulti/index.php?lang_mode=3); administration of the initiative is

under the direction of the National Library Board Singapore (http://www.nlb.gov.sg/), and the metadata registry is the responsibility of the Resource Center for Knowledge Communities (http://www.kc.tsukuba.ac.jp/en/divisions.html) at the University of Tsukuba (Japan).

PREMIS—Preservation Metadata

In March 2000, OCLC Research and RLG, Inc. launched a joint initiative to identify and support best practices for the long-term retention of digital objects. (RLG, Inc was an independent consortium of research libraries, museums and archives that operated programs, systems and services for its members from 1974 to 2006. In July 2006, OCLC and RLG combined.) This collaboration facilitated consensus-building activity among key stakeholders in digital preservation; the initiative brought together leading international experts from North America, Europe, and Australia to review existing practices, share expertise, and identify common approaches wherever possible (Lavoie, 2001). In 2002, the group published its final report, "A Metadata Framework to Support the Preservation of Digital Objects" (OCLC/RLG Working Group, 2002).

The framework included a set of metadata elements, mapped to the conceptual structure and reflecting the information concepts and requirements articulated in the Open Archival Information System (OAIS; Consultative Committee, 2002) model. The following year, a new working group (also jointly sponsored by OCLC Research and RLG, Inc.) was formed to develop a data dictionary intended for use by those who were actively engaged in preserving digital information. In 2005, the two organizations published the PREMIS Data Dictionary for Preservation Metadata (PREMIS Working Group, 2005). As a mark of recognition for its contribution to the global digital preservation landscape, the publication and its authors were awarded the 2005 Digital Preservation Award by the British Conservation Awards and the 2006 Society of American Archivists' Preservation Publication Award. The working group was retired.

The Library of Congress now sponsors the PREMIS Maintenance Activity and provides a permanent home for the Data Dictionary, XML schema, and related materials. In 2008, version 2.0 of the Data Dictionary was released by the international PREMIS Editorial Committee.

VIAF: The Virtual International Authority File

During the 2003 International Federation of Library Associations and Institutions (IFLA) World Library and Information Congress in Berlin, the Deutsche Nationalbibliothek (DNB), the Library of Congress (LC), and OCLC agreed to develop a Virtual International Authority File (VIAF) for personal names (Bennett, Hengel-Dittrich, O'Neill, & Tillett, 2007). Since then, the DNB, the LC, OCLC, and the Bibliothèque nationale de France (BnF) have jointly conducted a project to match and link the authority records for personal names in the retrospective personal name authority files of the three national library catalogs. The long-term goal of the VIAF project is to link the authoritative names from many national libraries and other authoritative sources into a

shared global authority service for persons, corporate bodies, conferences, places, and more. A prototype (VIAF, n.d.) was developed by OCLC Research and has been undergoing review and improvement for the past several years.

During 2008, the founding partners decided to expand the number of contributing national libraries and invited some two dozen institutions to consider participation. At this writing, a number of national libraries in Europe and the Pacific Rim have indicated their desire to become active members of this project.

The implications of this and related efforts in OCLC Research could be substantial. The techniques and systems that effectively support the management of multiple authority files will be critical to the success of an increasingly international and multilingual, networked community.

Virtual Reference

 "Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-user, and Librarian Perspectives" was undertaken by OCLC Research and Rutgers with grant support from the Institute for Museum and Library Services (IMLS), a U.S. federal agency. The project goals were evaluating the practice, sustainability, and relevance of virtual reference services (VRS). VRS are human-mediated, Internet-based, synchronous library information services. The rapidly increasing use of remotely accessed digital reference resources has increased the demand for librarians to provide reference services online. VRS users, nonusers, and librarians provided data for this research in focus group interviews, online surveys, telephone interviews, and VRS session transcripts. The project provided findings regarding user and librarian preferences, process and usability issues related to technology, and content issues related to information accuracy and query negotiation. In addition, the project results inform the understanding of computer-mediated communication, the impact of virtual relationships, and generational differences in information seeking using live chat technology (Sillipigni Connaway & Radford, 2005-2008; Sillipigni Connaway, Radford, & Dickey, 2008).

CURRENT WORK: GLOBAL IMPACT

OCLC Research works with the community to collaboratively identify problems and opportunities; to model and test solutions; to develop consensus; and to share findings through experimental services, published reports, presentations, and professional interactions. Its vision is global and its impact is profound.

The OCLC Research toolkit includes data, applications, standards, and services; more important, it is composed of an extensive supply of human expertise, ingenuity, creativity, and leadership. Its partners are staff at institutions around the world, who are likewise committed to improving the Web-based information environment for the benefit of users everywhere.

Four areas of focus characterize a wide range of work under way in the division:

- Harnessing the power of terminologies
- Making names work harder
- Mobilizing unique materials
- Understanding and managing the global, collective collection

Harnessing the Power of Terminologies

Controlled vocabularies, such as thesauri and subject heading systems, have been used by librarians for decades to improve the indexing of a wide variety of information resources (including articles, books, movies, music, photographs, and Web sites). The purpose behind this work has been to increase the effectiveness of the end user in locating desired materials.

OCLC Research's Terminology Services (OCLC, n.d.h) prototype uses library and Web standards to make the terms, relationships, descriptions, and other information in controlled vocabularies readily available as resources on the Web. Each vocabulary is fully indexed and searchable and a vocabulary entity can be referenced by its identifier. Data is retrievable in multiple representations including the MAchine-Readable Cataloging (MARC) authority format in (eXtensible Markup Language) XML, used by libraries, and the SKOS (Simple Knowledge Organization System) Core Vocabulary designed for Semantic Web applications.

Librarians and developers of library applications are encouraged to experiment with the vocabularies and services provided by this project. The services may be used in a variety of ways. Some examples:

- as a source of terms for social tagging
- for query refinement in search applications
- to provide context for a search term
- to validate names and subjects in metadata
- to facilitate cross database searching

Making Names Work Harder

OCLC Research has for some time been examining and building experimental services that make data work harder. That is, research scientists have deliberately set out to determine whether there are good ways to make the collective assets of WorldCat deliver even more value because of its very nature and composition. For example, names (personal, corporate, institutional) are a central part of every MARC record.

The idea of WorldCat Identities is simple: create a summary page for every name in WorldCat (Hickey, 2007). Since there are some 125 million records in the database and nearly 25 million names mentioned somewhere, this is a large-scale data mining effort that would have been difficult even a few years ago. In its initial stages, the initiative

focused on personal and corporate names; when viewing the beta version (http://orlabs.oclc.org/Identities) of the service, a page for the Beatles can be found as well as for John, Paul, George, and Ringo.

The summary pages are automatically generated and include:

- Most widely held works by the person
- Most widely held works about the person
- Genres of works
- Roles the person plays (e.g., composer, arranger, performer)
- A publication timeline for works both by and about the person
- Languages of publication
- Images of covers (if available)
- Audience level (kids, general, special)
- Related names (linking to WorldCat records and to their own Identities pages)
- Useful links
- Associated subjects

The beta service was tested in 2007 by a group of staff from institutions affiliated with the RLG Partnership (OCLC, n.d.b) and was moved into the production version of WorldCat in 2008. It builds on earlier work in OCLC Research to develop many of the functions behind the page displays. WorldCat Identities is proving to be a valued service for institutions wishing to link to this powerful tool in order to enhance their own users' experiences (Chan, 2008).

Mobilizing Unique Materials

For more than two decades, RLG worked closely with the rare books, archives, and manuscripts community to improve access to unique and special collections (Van Camp, 2003). Moreover, some of the earliest experiments in using digital technologies to effectively disclose special collections were managed by RLG with its members (Erway, 1996).

Today, special collections materials are of increasing interest and importance. As materials from the general stacks become more ubiquitous (through "mass" digitization projects and as institutions move toward joint ownership of books and journals), special collections may become what defines a library collection. With the shift in importance, this is a good time for an examination of the end-to-end process that results in archival and special collections materials being delivered to interested users. The overarching goal of this initiative is to achieve economies and efficiencies that permit these materials to be effectively described, properly disclosed, successfully discovered, and appropriately delivered.

Achieving control over these collections in an economic fashion will mean that current resources can have a broader impact or be invested elsewhere in other activities. The near-term effort is to take a systemwide view that will identify gaps in current

operations—missing evidence, needed changes in practice, cumbersome processes, deficiencies in technology platforms that support description and disclosure, and required supporting services (OCLC, n.d.d).

Understanding and Managing the Global, Collective Collection

For many years, OCLC Research scientists have been mining WorldCat in order to better understand the nature of the global, collective collection. Seminal research has been undertaken that has changed the conversation around the ways library collections intersect and overlap and the implications for the community in its approaches to acquisition, description, storage, access, and digitization (Lavoie, 2006).

Chris Anderson first coined the term "the long tail" in an article in *Wired* magazine in 2004, referring to the extensive and elusive, little-known but highly prized materials at the shallow end of the publishing curve (Anderson, 2004). Since then, OCLC Research staff have posed a number of new questions about the systemwide library book collection (Dempsey, 2006) while also conducting research into several aggregations of holdings and gathering evidence to support new forms of interlibrary collaboration (Lavoie & Waibel, 2008. See also Jackson, Sillipigni Connaway, & Loh, 2007 and O'Neill, Sillipigni Connaway, & Prabha, 2006).

Since 2006, new work to engage with members of the RLG Partnership around collective collection issues of overlap, duplication, and last copies has resulted in a set of initiatives to advance the research library agenda by identifying strategies to understand, prepare for, and help advance libraries, archives, and museums in more profoundly cooperative models of acquiring, managing, and disclosing collections (OCLC, n.d.e). Further, specific attention has been focused on the ways libraries, archives, and museums can more effectively work together (Zorich, Waibel, & Erway, 2008).

COMMUNITIES AND OUTREACH

Throughout its history, OCLC Research has engaged with multiple communities and shared the results of its work in a variety of venues. OCLC Research works with the community and affiliates in the RLG Partnership to:

- collaboratively identify problems and opportunities
- develop prototype and test solutions
- develop consensus
- publish insightful and timely reports
- share findings through presentations and professional interactions

Formal publication (OCLC, n.d.f) in the information and library science literature is a frequent route to sharing findings and new insights. Presentations are often given at professional conferences around the world, contributing to a wider understanding of the work underway in the division.

In addition to a range of collaborations both formal and informal, standard forms of outreach also include the following:

- partnering with faculty at library schools and libraries on research topics of mutual interest
- visiting institutions where related work is being undertaken in order to exchange approaches and comingle investigations
- collecting impressions and opinions of professionals in order to shape our work
- inviting leaders in the field to share the results of their work in the OCLC Research Distinguished Seminar Series (OCLC, n.d.c)
- hosting RLG Partnership events (OCLC, n.d.g) that provide a venue for institutions to engage with one another around common interests
- managing groups of experts from around the world who come together to solve common problems
- interviewing key players in the field and conducting Webinars on the use of new tools to explore new challenges (OCLC, n.d.g)
- publishing reports and newsletters to more broadly disseminate our work
- working closely with members, partners, and other agencies on the development, maintenance, and deployment of information system-related standards

FUTURE DIRECTIONS

OCLC Research plays a distinctive role in advancing the library agenda and in collaborating with libraries, archives, and museums that seek to better serve their primary constituencies through the exploration and application of networked solutions.

In the near term, OCLC Research's program officers and research scientists will continue to examine, analyze, and address some of the grand challenges facing the global community in partnership with allied institutions, organizations, and individuals:

- Can libraries share the increasing burdens of managing vast collections of printed works?
- Can we do a better job of managing workflows in research assessment?
- Can we work smarter and faster as we invest in projects and programs to disclose and mobilize our unique collections?
- Can we amplify the value of controlled data and at the same time streamline our collective investments for the ultimate benefit of our users?
- Can we assist in the design of services that are optimized to operate in the network cloud as well as those that span personal, institutional and cloud environments? Current information about OCLC Research and its activities is available on its Web site: http://www.oclc.org/research/

REFERENCES

1. Anderson, C. (2004, October) The long tail. *Wired* **12**:10 — Retrieved from http://www.wired.com/wired/archive/12.10/tail.html

2. Bennett, R., Hengel-Dittrich, C., O'Neill, E. T. and Tillett, B. (2007) VIAF (Virtual International Authority File): Linking the Deutsche Nationalbibliothek and Library of Congress name authority files. *International Cataloguing and Bibliographic Control* **36**:1 , pp. 12-18.

3. Chan, S. (2008, 7 December) OPAC: Connecting collections to WorldCat identities. Fresh + New (blog), Powerhouse Museum. — Retrieved from http://www.powerhousemuseum.com/dmsblog/Index.php/2008/12/07/opac-connecting-collections-to-worldcat-identities

4. Consultative Committee for Space Data Systems (2002, January) Reference model for an open archival information system (OAIS). — Retrieved from http://public.ccsds.org/publications/archive/650×0b1.pdf

5. Dempsey, L. (2006) Libraries and the long tail: Some thoughts about libraries in a network age. *D-Lib Magazine* **12**:4

6. Dublin Core Metadata Initiative (DCMI) incorporated in Singapore (2009, January) — Press release. Retrieved from http://dublincore.org/news/press/DCMI_press_20090105_incorporation.pdf

7. Erway, R. (1996) Digital initiatives of the Research Libraries Group. *D-Lib Magazine* — Retrieved from http://www.dlib.org/dlib/december96/rlg/12erway.html

8. Godby, C. J. and Richardson Jr, J. V. Drake, M. A. (ed) (2003) OCLC Office of Research. *Encyclopedia of library and information science* pp. 2231-2237. Marcel Dekker , New York

9. Hickey, T. B. (2007) WorldCat identities: Another view of the catalog. *OCLC NextSpace* **6**

10. International Standards Organization ISO 15836:2003 and American National Standards Institute/National Information Standards Organization. ANSI/NISO Z39.85

11. Jackson, M. E., Sillipigni Connaway, L. and Loh, E. (2007) Changing global book collection patterns in ARL libraries. Report, Global Resources Network. — Retrieved from http://www.arl.org/resources/pubs/grn_global_book.shtml

12. Lavoie, B. F. (2001) Preservation metadata for digital objects: A review of the state of the art (White paper, OCLC/RLG working group on preservation metadata) OCLC

research publication. — Retrieved from
http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

13. Lavoie, B. F. (2006) Books without boundaries: A brief tour of the system-wide print book collection. With Schonfeld, R. C.. *Journal of Electronic Publishing* **9**:2

14. Lavoie, B. F. and Waibel, G. (2008) An 'artful' analysis of the rich holdings of four museum libraries. *NextSpace* **8**:February

15. Lorcan Dempsey to head office of research (May 21, 2001) OCLC Press Release. OCLC Digital Archive. —
http://worldcat.org/arcviewer/1/OCC/2006/07/19/0000023368/viewer/file28.html

16. Michalko, J. The RLG Partnership. *Journal of Library Administration* — in press

17. OCLC Online Computer Library Center About research. — Retrieved, from
http://www.oclc.org/research/about

18. OCLC Online Computer Library Center Betatest WorldCat Identities project. — Retrieved from
http://www.oclc.org/programs/ourwork/renovating/leveragevocab/identities.htm

19. OCLC Online Computer Library Center Distinguished seminar series. — Retrieved from http://www.oclc.org/programsandresearch/dss/default.htm

20. OCLC Online Computer Library Center Effectively disclose archives and special collections program. — Retrieved from
http://www.oclc.org/programs/ourwork/collectivecoll/archives/default.htm

21. OCLC Online Computer Library Center Managing the collective collection. — Retrieved from http://www.oclc.org/programs/ourwork/collectivecoll/default.htm

22. OCLC Online Computer Library Center Publications 2000-2009. — Retrieved from http://www.oclc.org/research/publications/default.htm

23. OCLC Online Computer Library Center RLG Programs events. — Retrieved from http://www.oclc.org/programs/events/default.htm

24. OCLC Online Computer Library Center Terminology services. — Retrieved from http://www.oclc.org/research/projects/termservices/default.htm

25. OCLC/RLG Working Group on Preservation Metadata (2002, June) Preservation metadata and the OAIS information model: A metadata framework to support the preservation of digital objects. — Retrieved from
http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

26. O'Neill, E., Sillipigni Connaway, L. and Prabha, C. (2006) Last copies: What's at risk?. *College and Research Libraries* **67**:4 , pp. 370-379.

27. PREMIS Editorial Committee (2008, March) PREMIS data dictionary for preservation metadata, version 2.0. — Retrieved from http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

28. PREMIS Working Group (2005, May) Data dictionary for preservation metadata. — Retrieved from http://www.oclc.org/research/projects/pmwg/premis-final.pdf

29. RLG membership approves move to combine with OCLC (2006) — http://worldcat.org/arcviewer/1/OCC/2007/01/09/0000057113/viewer/file17.html

30. Sillipigni Connaway, L. and Radford, M. L. (2005-2008) Seeking synchronicity: Evaluating virtual reference service from user, non-user, and librarian perspectives. Final project report, Institute of Museum and Library Services

31. Sillipigni Connaway, L., Radford, M. L. and Dickey, T. J. (2008) On the trail of the elusive non-user: What research in virtual reference environments reveals. *Bulletin of the American Society for Information Science and Technology* **34**:2 , pp. 25-28.

32. Terry Noreault promoted to vice president (January 8, 1999) OCLC Press Release. OCLC Digital Archive. — http://worldcat.org/arcviewer/1/OCC/2003/07/08/0000003706/viewer/file77.html

33. Van Camp, A. (2003) RLG, where museums, libraries, and archives intersect. *LIBER Quarterly* **13**:3/4

34. VIAF: The Virtual International Authority File — Retrieved from http://orlabs.oclc.org/viaf/

35. Weibel, S. L. (2005) Border crossings: Reflections on a decade of metadata consensus building. *D-Lib Magazine* **11**:7 — [July/8/August]. Retrieved from http://www.dlib.org/dlib/july05/weibel/07weibel.html

36. Zorich, D., Waibel, G. and Erway, R. (2008) Beyond the silos of the LAMs: Collaboration among libraries, archives, and museums. OCLC Research , Dublin, OH