



# Five-Year Information Format Trends

*This report presents data and forecasts about information format trends that will likely shape the information landscape of the future.*

## Market Trends & Analysis

Information trends and format innovations that are quickly taking shape have created a complex and challenging new information landscape.

- Over the next five years, librarians will continue to be faced with managing an unprecedented array of content in a rapidly proliferating mix of formats.
- Many library patrons in the developed world are quickly adopting new information dissemination technologies, from PCs, to PDAs, to cell phones, to MP3 players.
- Tools, services and technologies that did not exist ten years ago are shaping expectations of access to information.
- Libraries must balance new patron demands against the reality of tighter budgets and smaller workforces.
- Library resources will continue to be bound by governmental and college/university budgeting priorities and uncertain economic conditions.

**Perhaps the most significant challenge is that the universe of materials that a library must assess, manage and disseminate is not simply shifting to a new set or type of materials, but rather building into a much more complex universe of new and old, commodity and unique, published and unpublished, physical and virtual.**

It should be noted that this report addresses the information landscape of libraries, rather than other cultural heritage organizations, e.g., museums and archives.

Research shows that:

- Traditional information formats (e.g., printed books and serials) are not going away.
- Traditional formats are combining with emerging formats, as print books share space with eBooks and print-on-demand alternatives.

- New formats in the digitization arena are rapidly expanding the information universe.
- The estimated annual production of materials in 'Web-ready formats' (by the year 2007) is projected to be "too large to estimate" by many analysts.

These trends emphasize what libraries already know: they are faced with the challenges of integrating traditional and emerging formats, balancing resource allocation between traditional and emerging technologies, and building new information management processes and procedures.

Looking at information format trends affecting libraries both now and in five years, we analyzed four main areas:

- **Traditional Materials**
- **Scholarly Materials**
- **Digitization Projects**
- **Web Resources**

Traditional and scholarly materials were analyzed to identify emerging information formats and review anticipated shifts in annual production volumes from traditional to new formats.

Emerging trends in digitization were analyzed by reviewing the scope, size and coverage of several of the larger digitization programs currently under way. Calculating an estimate of future worldwide digitization volume proved to be infeasible given the lack of available forecasts.

Although many of the emerging formats will use newer technologies and incorporate Web access, we also reviewed Web resources as a separate category. We analyzed data that looked at the growth potential of the traditional Web as well as the deep, or hidden, Web.

# Traditional Materials

## and their electronic counterparts

### TOP TRENDS

- print books down
- POD growth coming
- eBook adoption modest
- eJournals increasing
- AV media growing

Our analysis looked at books, eBooks, print-on-demand, journals & newspapers, and audiovisual media.

### Books

“Today, the book business stands on the edge of a vast transformation,” according to Jason Epstein, in *Book Business*. (Norton, 2001, p. 2.)

- *Book Sales Decline*—The 2002 edition of *Book Industry Trends* reported that U.S. sales declined in 2001.
- *Book Sales to Slowly Increase*—Total number of books sold is expected to increase only modestly. U.S. book sales are projected to increase 2 percent annually while worldwide book sales continue to grow modestly.

### Print-on-demand

Print-on-demand could very well be an important step toward a digital distribution model for books.

(*Against the Grain*, June 2002.)

- *POD Bears Watching*—Bowker’s Andrew Grabois believes, “The thing to watch is print-on-demand.” (email, Sept. 3, 2002.)
- *POD Growth on the Horizon*—Rick Anderson, University of Nevada, Reno, predicts that in five years books will still be primarily print, but print-on-demand will take hold. Frank Romano, Rochester Institute of Technology, believes “that by 2015, 48% of all books will be digitally printed on-demand.” (*Against the Grain*, June 2002 and Sept. 2002.)

**The average [U.S.] public library materials budget for FY03 will dip 2.1%, the first decline in the past five years.**

*Library Journal*, Jan. 15, 2003.

### eBooks

According to *Book Industry Trends 2002*, e-publishing is undergoing dynamic and promising evolution at a historically familiar pace. No major publisher or agency has ventured an estimate for 2007 eBook output.

- *eBook Sales Modest*—While adoption of eBooks remains modest, new readers could be a catalyst to wider acceptance of eBooks. However, Grabois predicts, eBook sales “will fall short of 2001 by several thousand titles.” (email, Sept. 3, 2002.)
- *No Projections for Future*—Neither analysts nor vendors have ventured to predict the future production of eBooks. However, with a 50 percent new book-to-eBook conversion rate, and a modest non-U.S. and backlist conversion rate, roughly 450 thousand eBooks could be produced in 2007.

**In 2002, U.S. libraries “continued to expand electronic collections [and] vendors ... brought significant content to the market”.**

*Library Journal*, Dec. 15, 2002.

### 2001 U.S. Book Sales

**Total books: down 3.3%**

**Professional books: down 7.6%**

*Book Industry Trends 2002.*

**A monograph is handled at least 25 times  
in today’s distribution system.**

*Charleston Report*, Nov/Dec 2002.

## Journals & Newspapers

While the production of print journals is expected to be steady, online journals are predicted to increase.

- *Active Print/Online Journals Increasing*—The number of active print titles will increase slightly. However, the growth of journals available online is projected to increase substantially.
- *U.S. Daily Newspapers to Decline/Online Steady*—U.S. daily newspapers will decline by about 3 percent (by 2007), according to *Ulrich’s and Editor & Publisher*. The number of electronic newspapers is expected to remain flat.

*The consumer book publishing industry is barely growing and is only modestly profitable.*

*New York Times*, Jan. 23, 2003, C12.

## Audiovisual Media

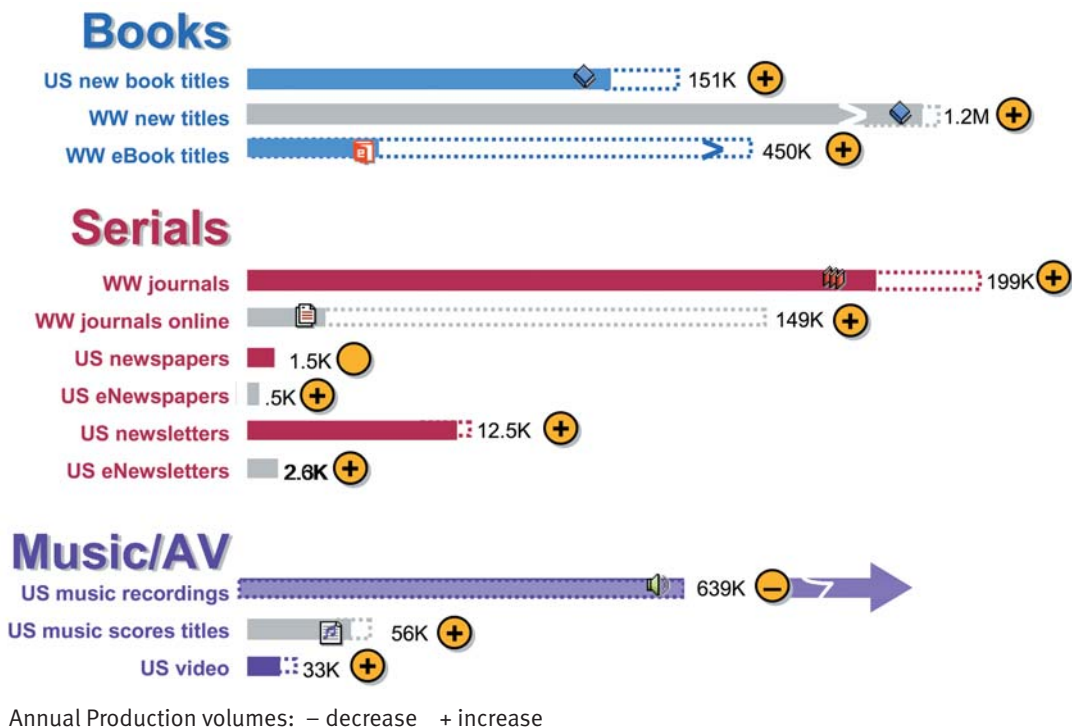
The expectation is that more audiovisual materials will be embedded into technology formats in the future, e.g., e-learning applications and the Web.

- *Audiovisual Media Growing*—Music sound recordings, sheet music and videos are expected to show steady but modest growth in the U.S., according to industry professionals.

*Music album “sales declined for the second straight year in 2002, down 10.7 percent from the previous year.” Younger buyers are buying fewer albums because they’re finding what they want for free on the Web.*

*Columbus (OH) Dispatch*, Jan. 4, 2003.

### Shifts in Annual Production Volume 2002–2007



# Scholarly Materials

## and their electronic counterparts

### TOP TRENDS

- library book spending down
- research ePapers increasing
- scholarly publishing in crisis
- ePrint archives expanding
- print journals down
- course management materials exploding
- research papers flat

### U.S. University Presses

**Hardcover sales: down 26.8%**  
**Paperback sales: down 9.8%**

**Library purchases: down 12%**

*LJ Academic Newswire*, Nov. 7, 2002.  
*Against the Grain*, Nov. 2002

Our analysis looked at books, journals, scholarly articles, ePrint archives, theses & dissertations, and course management materials.

### Books

When facts are reviewed, many of the same trends emerge for scholarly publishing as for traditional formats. The emergence of the electronic format in both books and print-on-demand is again the significant trend.

- *Library Book Spending Down—The Survey of Academic Libraries, 2002 Edition*, shows that U.S. academic library purchases of print resources are down precipitously: 6 percent from 2000 to 2001, and 8 percent in 2002. (*Against the Grain*, Nov. 2002.)
- *U.S. University Press Sales Down*—Scholarly monograph sales trends are even grimmer. University press hardcover sales fell 26.8 percent in August 2002, and paperback sales dropped 9.8 percent. Library purchases of university press books were down over 12 percent in 2002. (*Library Journal Academic Newswire*, Nov. 7, 2002, and *Against the Grain*, Nov. 2002.)
- *Scholarly Publishing Crisis Emerging*—Stanford University Press has cut its workforce and slashed the number of books published. William Froelich, director of Northeastern University Press, expressed hopes that new delivery systems will become available for scholarly monographs. Cost-effective, short-run printing will [help] keep more books in print. Duke University Press' Ken Wissoker noted that the changed publishing landscape had made book publishing increasingly difficult. A book needs to generate \$32,000–\$42,000 (2,500–3,500 copies). “Most books ... don’t sell 1500 copies,” and course packs adversely affect book sales. (*LJ Academic Newswire*, Oct. 3, 2002.)

### Journals

Rick Anderson, University of Nevada, Reno, believes that five years from now “research journals will be published almost entirely online.” (*Against the Grain*, Sept. 2002.)

- *Large scientific publishers have primary titles online*—In 1994, fewer than 75 peer-reviewed, electronic journals existed. In 1998, 30 percent of titles cited in *Science Citation Index* were online. In 2002, 75 percent of journals cited in *Science Citation Index*, and 63 percent of journals cited in *Social Sciences Citation Index* were available electronically. (*Library Journal*, Apr. 15, 2002.)
- *Researchers Prefer Electronic*—84 percent of faculty and students surveyed at Drexel University prefer eJournals to print, believing that eJournals save time and make more information available. (*Charleston Report*, Nov./Dec. 2002.)

### Electronic Journal Explosion

**1994: fewer than 75 peer-reviewed electronic journals**

**1997: some of the largest STM publishers had no journals online**

**1998: 30% of titles in Science Citation Index online**

**2002: 75% online in SCl;  
 63% online in Social Sciences Citation Index;  
 34% online in Arts & Humanities Citation Index**

*Library Journal*, Apr. 15, 2002.

**Decline in U.S. academic library spending is precipitous:**

**down 6% for 2000–2001      down 8% for 2002**

*Against the Grain*, Nov. 2002.

## Scholarly Articles

Scholarly articles may begin migrating away from traditional scholarly publishing toward ePrint archives and other types of initiatives, according to the British Library. And Rick Anderson, University of Nevada, Reno, forecasts that the “concept of an 'issue' will have started to sound very 20<sup>th</sup>-century. Half of research libraries will have stopped journal check-in.” (*Against the Grain*, Sept. 2002.)

- *Research Paper Production Flat*—A British Library study suggests that while production of research papers is expected to remain flat, there will be a move toward electronic publishing.
- *Two Million Scholarly Articles per Year*—Stevan Harnad, Southampton University, UK, estimates that the refereed research literature consists of 2,000,000 annual articles in 20,000 journals. (eprintblog, Sept. 25, 2002.)

## ePrint Archives

ePrint archives are repositories for electronic versions of papers (self-archived by the authors), made available to the scholarly community prior to publication.

- *ePrint Archives Expanding*—One prominent example, ArXiv, the Physics ePrint archive at Cornell University, is estimated to increase from its present 210,000 preprints to over 385,000 by 2007—an average growth of about 35,000 a year—roughly 135 per weekday. And the ArXiv’s popularity has soared to 15 million downloads per year. (Note: This estimate represents the growth of only one ePrint archive—no industry-wide forecasts are available).

- *Scholarly Articles Accessible*—Stevan Harnad observed that the online medium has now made it possible to provide open-access (toll-free online access) to the entire refereed research literature. (eprintblog, Sept.25, 2002)

## Theses & Dissertations

According to *Dissertation Abstracts International*, a strong trend is emerging toward electronic production and away from creating new paper documents.

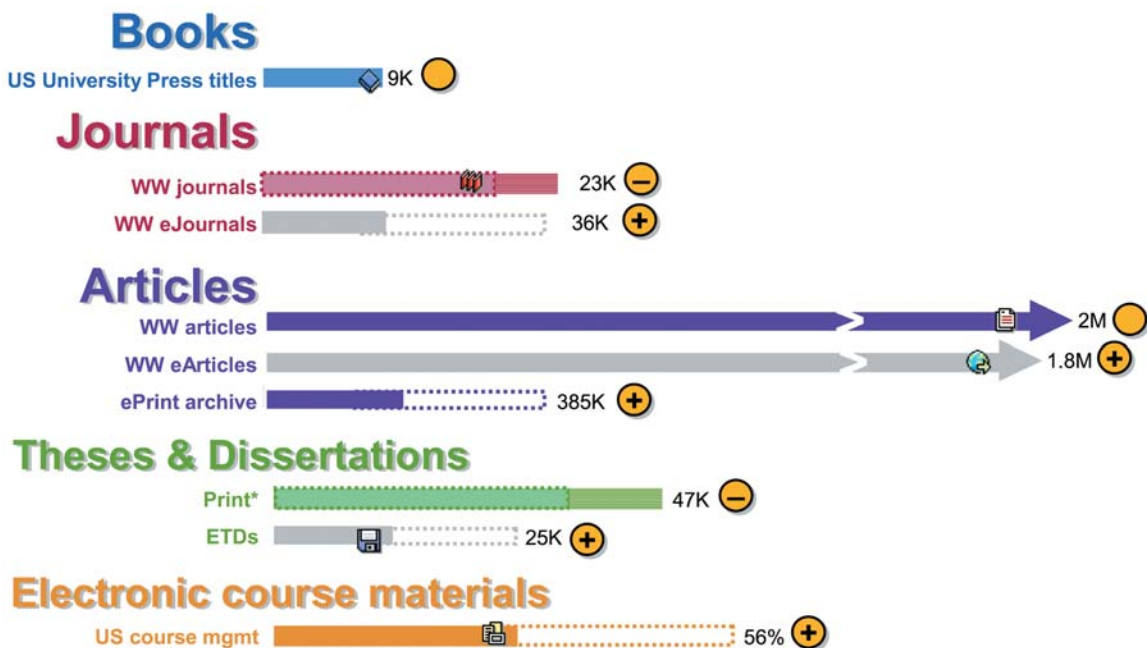
- *Paper Theses & Dissertations Declining*—The number of new paper theses and dissertations is expected to decline moderately.
- *Digital Electronic Theses & Dissertations Increasing*—The British Library study predicts that by 2007 at least 50 percent of all theses and dissertations will be submitted digitally.

## Electronic Course Management Materials

One of the more exciting areas to watch is that of electronic course management materials.

- *More eContent Available*—According to XanEdu, a ProQuest subsidiary, 5.5 billion pages of copyright-cleared articles are currently available for use in course management materials. Even by conservative estimates, this figure is expected to double by 2007.
- *More College Courses Available*—According to results of a U.S. Campus Computing Project survey, as many as 56 percent of U.S. college courses could be available via course management systems by 2007.

### Shifts in Annual Production Volume 2002–2007



Annual Production volumes: – decrease + increase

\*U.S., Canada, Britain & other European countries

# Digitization Projects

## TOP TRENDS

- commercial digitization expanding
- national digitization growing
- state and local projects increasing

Our analysis looked at current digitization endeavors, including commercial, national, state, and local projects.

## Current Activities

The management and preservation of digital content is expanding so rapidly that the total extent of the digitization universe is hard to quantify and its growth hard to predict. While libraries worldwide are building prototypes and digitizing special collections—and many significant and substantial projects are well under way—funding agencies are providing the financial resources fueling the emergence of more digitization initiatives. Current activities could easily account for over 100 million pages.

## Commercial Digitization

Two ambitious examples include: The Gale Group, with over 8 million articles in its *Times Digital Archive* and 20 million pages in its *18th-Century English-Language Books*; and ProQuest, which to date has digitized over 4.4 million pages from *The New York Times*, *Wall Street Journal*, *Christian Science Monitor* and *Washington Post*.

## National Digitization

Typically government funded, several major national digitization projects are currently under way. Two prominent examples include: Gallica 2000 (Bibliothèque nationale de France), which to date has digitized over 15 million pages; and the British Library, which is in the process of digitizing over 100 thousand images for its Digitisation Programme, 'In Place'.

## State & Local Projects

As money is made available for digitization, more U.S. state and local libraries are undertaking digitization projects. Two examples include: the Colorado Digitization Project, with 145 thousand digital objects; and the Everett Public Library (WA) Everett Massacre project, which went online in September 2002 with 215 digital objects, including audio.

## Widespread Involvement in Digitization Projects

### AMICO Digital Library (AMICO)<sup>1</sup>

Not-for-profit organization of 30 institutions collaborating to enable educational use of museum multimedia.

**100,000 items**

### British Library Digitisation Programme: "In Place"<sup>2</sup>

Maps, manuscripts & topographical illustrations, photographs, rare sound recordings, advertisements and music hall songs.

**100,000 images and sounds (2004)**

### Colorado Digitization Project<sup>3</sup>

Provides access to Colorado's history, culture, government and industry. Content may be "born digital" or digitized.

**145,000 objects**

### Gale

*Times Digital Archive, 1785–1985* includes every page as published from 200 years of *The Times* (London)<sup>4</sup>.

**>8,000,000 articles**

### Gale

*18th-Century English-Language Books* includes most books published in English in the 18th century.<sup>5</sup>

**20,000,000 pages (June 2003)**

### Gallica 2000<sup>6</sup>

Bibliothèque nationale de France digital library, with multimedia documents from the Middle Ages to the early 20th century.

**15,000,000 pages**

### Koninklijke Bibliotheek (Netherlands)<sup>7</sup>

Koninklijke Bibliotheek (Netherlands) will convert 400,000 records, making them available online and the materials they represent more accessible worldwide.

**400,000 records planned (2005)**

social history from the antebellum period through reconstruction.

**1.5 million images**

### National Newspaper Library (Hemeroteca)

National Newspaper Library (Hemeroteca) of the National Autonomous University of Mexico (UNAM) has contracted digitization of its entire microfilmed collection of Mexican historical newspapers and periodicals.<sup>10</sup>

**>20,000,000 pages planned**

### NewspaperArchive.com

January 2003, signed 5-year deal with EBSCO Publishing to market its online archive to 30,000 libraries worldwide.<sup>11</sup>

**5,500,000 pages**

### Paper of Record

Cold North Wind's portal to its newspaper archive, with content from Australia, Canada, Mexico, United Kingdom and US.<sup>12</sup>

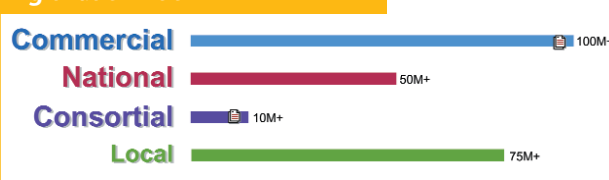
**>3,600,000 pages**

### ProQuest Historical Newspapers

*The New York Times*, *The Wall Street Journal*, *The Christian Science Monitor*, *The Washington Post*, and other newspapers.<sup>13</sup>

**7,800,000 pages**

### Digitization 2002



### Library of Congress

#### American Memory Project<sup>8</sup>

Primary source materials relating to the history and culture of the United States from more than 100 historical collections.

**>7,000,000 digital items**

#### Making of America (MOA)<sup>9</sup>

Digital library of primary sources in American

# Web Resources

## TOP TRENDS

- information seekers' expectations increasing
- Web site growth slowing
- surface Web documents increasing
- deep Web expanding

Our analysis looked at recent research into Web use, as well as estimates of size and growth for both the surface Web and the deep Web.

## Recent Research

Recent research indicates that information seekers begin with the Web.

- *Online Before Print*—Students and faculty turn to online library materials before print and use the library “significantly less” than they did two years ago, according to a study commissioned by the Digital Library Federation. < [www.clir.org/pubs/abstract/pub110abst.html](http://www.clir.org/pubs/abstract/pub110abst.html) >
- *Internet First Choice*—According to the Pew Internet and American Life Project, most Internet users and some nonusers expect to find reliable information online; most find the information they seek; and many now turn to the Internet first. < [www.pewinternet.org/reports/toc.asp?Report=80](http://www.pewinternet.org/reports/toc.asp?Report=80) >

## Surface Web

The surface Web includes Web sites readily available and accessible through open search engines. The significant increase in new Web formats, discussed in previous sections of the report, suggests that it's important to monitor and analyze trends in surface Web site development.

- *Web Site Growth Slowing*—According to the OCLC Office of Research, there are currently about 9 million Web sites, and although growth is slowing, there could be as many as 10.4 million in five years.

**About 98% of U.S. schools had access to the Internet in 2000.**  
*Digest of Education Statistics, 2001.*

***“I love the fact that students now have access to the riches of online content... The question is whether their online searching skills are limited to finding MP3 files and pictures of Britney Spears....”***

*Mary Ellen Bates, Econtent, June 2002.*

- *Surface Web Documents Growing*—International Data Corporation (IDC) predicts that the number of surface Web documents will grow from the current 2 billion to 13 billion within three years.
- *Web Resources Cataloged*—To date, approximately 700 thousand Web resources have been cataloged in WorldCat.

***Seven of ten U.S. college students say they are successful at finding what they seek most of the time.***

***Their first-choice web resources are search engines, web portals and course-specific websites.***

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

## Deep Web

The deep Web encompasses the portion of the Web that is not accessible through open search engines (e.g., library OPACs, reference databases, JSTOR, U.S. Census, etc.)

- The invisible Web could be 50 times larger than the visible Web—According to Isidro F. Aguillo, CINDOC-CSIC, “Measuring informal scientific publication in the Web” \* (2002), the invisible Web is from 2 to 50 times larger than the visible web.
- *Library Catalogs*—There are 40,000 Web OPACs.
- *Other Databases*—Other bibliographic databases, encyclopedias, numeric and full-text data could number 250,000.
- *Non-textual Web Pages*—Numeric and other non-textual Web pages could equal more than 300 million.
- *Document Repositories and eJournals*—There are more than 10,000 eJournals.

### ***Google Queries Per Day***

1999: 3 million      2000: 20 million  
2001: 100 million      2002: 150 million

<http://www.google.com/press/milestones.html>  
<http://searchenginewatch.com/reports/perday.html>

# Summary Observations

The information format landscape is changing across the developed world. Many of the 2007 estimates presented in this report are not as precise as the authors would have preferred. Consistency of trends across all four information format categories suggests that both a shift and an expansion of information formats are well under way and will likely accelerate in the next several years.

- A shift in new unit production of traditional and scholarly materials from paper-only to a combination of paper, print-on-demand (POD) and electronic versions is evident in the developed world. Annual production rates of electronic-only formats are growing faster than their paper-only counterparts. Additionally, the 2007 forecasts suggest even stronger shifts.
- On the more distant horizon, print-on-demand is likely to increase and bears watching.

- Traditional information formats are not going away.
- New Web sites are not likely to continue to be created at current or recent growth rates. But, the number of objects (both text and graphic) expected to populate current sites is predicted to grow exponentially.
- Digitization, although the most difficult area to quantify with respect to total worldwide production volume, may emerge as the most significant new format trend by 2007. New funding, imaging technology, marketing and patron appeal will likely fuel expansion faster than other categories.

## Methodology

The OCLC Library & Information Center compiled data comparing current output, or current activity, with projections for 2007. The data was drawn from a wide range of sources principally from the developed world: traditional paper-based resources, full-text journal articles, respected Web sites, and personal correspondence with industry experts. In some cases, where no directly relevant source existed, the librarians conducted primary research and arrived at estimates. These estimates are based on data from published projections, historical trends and professional expertise.

The goal was to analyze worldwide trends and review the overarching information trends that would likely impact all library communities—academic, public, corporate, etc. For some format categories (particularly the traditional ones), international data and 2007 predictions were available. For many of the new information formats, worldwide data was unavailable, inconsistent, or incomplete. Available academic data significantly outweighed available data for public and other library communities, even in the U.S.

For more information, please contact:

### OCLC Library & Information Center

1-800-848-5878

Larry Olszewski, Manager, x6293

Ginny Browne, Knowledge Management Librarian, x5217

Terry Butterworth, Serials Librarian, x4300

Leslie Dillon, Information Services Specialist, x6133

Paula Julien, Acquisitions, Interlibrary Loan, and Circulation, x5100

## References

- 1 <http://www.amico.org/>
- 2 <http://www.tmg.co.uk/what/news.cfm?newsID=14>
- 3 Liz Bishoff, Colorado Digitization Project, email 8/20/02.
- 4 <http://www.galegroup.com/Times/explore.htm>
- 5 *Information Today* 19, no. 7 (July/August 2002) pp. 1,55.
- 6 Gallica 2000 “will soon be expanded further to include some 15000 new out-of-copyright digitised titles. The current three-year plan will be adding *Voyages en France* (5000 titles), *Utopies* and *Voyages en Afrique*.” [http://www.bnf.fr/site\\_bnf\\_eng/connaitrgb/gallicagb.htm](http://www.bnf.fr/site_bnf_eng/connaitrgb/gallicagb.htm)
- 7 <http://www.oclc.org/oclc/press/20020906.shtm>
- 8 <http://memory.loc.gov/ammem/amhome.html>
- 9 [http://www.hti.umich.edu/m/moa.new/moa\\_collection.html](http://www.hti.umich.edu/m/moa.new/moa_collection.html)
- 10 Cold North Wind Press Release, Nov. 14, 2002.
- 11 [http://www.newspaperarchive.com/news/display\\_news.aspx?ContentID=51](http://www.newspaperarchive.com/news/display_news.aspx?ContentID=51)
- 12 <http://www.paperofrecord.com/>
- 13 <http://www.il.proquest.com/proquest/histdemo/default.shtml>

[www.oclc.org](http://www.oclc.org)

The following OCLC product and service names are trademarks or service marks of OCLC Online Computer Library Center, Inc.: OCLC. Third-party product and service names are trademarks or service marks of their respective companies. In addition, the OCLC symbol is a service mark of OCLC. OCLC grants permission to photocopy this publication as needed.