

Digital Journals in 2003: Dramatic Growth, New Tools, and Economic Drama

Nancy R. John

Interim University Librarian, University of Illinois at Chicago
nrj@uic.edu

Edward J. Valauskas

Chief Editor, *First Monday*
ejv@uic.edu

The informational deluge in the scientific—and humanistic—disciplines is a very real and palpable problem. I have often compared myself to a man who, although himself not actively participating in scientific research, still tries to keep up with it. He finds himself in a situation of a passenger who not only faces the specter of the train leaving the station without him, but must choose from a vast number of trains, all of which are taking off in different directions. The very fact that I have in my study a pile of unread journals is a sure sign that I have already exceeded the quantitative threshold of information that I can digest.

—Stanislaw Lem¹

The year 2003 was a memorable one for Internet-based and digital journals. In terms of numbers, the quantity of journals available in some sort of electronic format is staggering (although exact numbers are still as elusive as ever). According to a 2003 report commissioned by the Wellcome Trust, there are some 164,000 journals published in print worldwide. Of that total there are some 28,000 journals available in some sort of digital version or 17 percent of all published.² These figures, which cry for an accurate census of digital periodicals, suggest an enormous increase in the past few years. Tenopir and King estimated in 1999 that there were 4,000 digital journals, or 2.5 percent of the total of 157,000 available print journals.³ If, in less than half a decade, the number of digital journals has increased so significantly, there have to be important catalysts driving this trend. These catalysts may be in large part obvious: publishers, readers, and libraries are all driving a trend toward the transformation of periodicals, especially scholarly journals, into various digital forms, parallel to a continued production in paper, or in some cases to periodicals born digital and remaining so.

In an all-consuming paradoxical effort to provide increased access to more information, any time and anywhere, at lower costs, digital journals have been viewed as the ultimate solution. Publishers have responded by making bundles available of various flavors, combining print subscriptions with electronic versions at a variety of price points from free to hundreds of thousands of dollars. Libraries have increasingly shed their passive role by actively debating the merits of differential subscription costs, taking on the roles of publishers, subscription agents, and digital zealots all at once. Scholars too have tossed off their passive roles as creators of content, once all too willing to give their work to the highest-cited journal without question. Within the past few years, but especially in 2003, scholars have revolted (and seem ready to continue to revolt, if they receive the blessing of tenure and performance review committees), shrugging off their roles on editorial boards of profitable journals, winging it by creating new journals

with new partners. Probably the best example of this phoenix-like behavior was demonstrated by Vivian Siegel when she abandoned her editorship of the prestigious journal *Cell* to join the Public Library of Science (PLOS).⁴ The Public Library of Science, a nonprofit organization, released in October 2003 its journal *PLOS Biology*, a monthly, open-access, free, peer-review, Internet-based journal.⁵ Scholars are also fighting for their copyrights, wresting them away from publishers, looking for new ways to recycle their words into classroom assignments, chapters, essays, and even into altogether new books. This revolution will only continue to build momentum in 2004, thanks to further calls to “just say no” to commercial publishers and their journals.⁶

Driven by the demands of users to get full-text articles anywhere, any time, 2003 was the year of the link (as in hyperlink). The OpenURL specification, a standardized syntax for parsing and delivering metadata about an object, took hold as many services for linking full text in article databases to citations appeared on the scene. The wider development of OpenURL-based implementations extended the accessibility of full-text collections of electronic journals. At year’s end the National Information Standards Organization (NISO) was getting ready for the 2004 balloting of the OpenURL Framework.⁷ Nearly every commercial vendor of electronic content, e-journal management, or library services is providing OpenURL support or interoperability. Many noncommercial services, such as arXiv, are also providing support.⁸

Many vendors developed sophisticated link resolvers that leverage the OpenURL syntax to allow users to find a citation in one place and get full text from an appropriate source. Put more simply, a link resolver knows where a library has licensed its full text and will direct the user there rather than to another source where the user would have to pay. As a result, licensors of e-journals could pick and choose the interface from which they accessed journals as the full-text content of a single publisher was accessible from dozens of indexing services. The importance of this new development is revealed by data published by JSTOR, which shows that 11 percent of the uses of the JSTOR article database in 2003 came from external links.⁹

A critical factor, then, in development in 2003 was the rebound of some indexing and abstracting services, which without full text had been on the wane. Now, empowered by linking technologies, these indexing services, which had been typically both broader in coverage and full of features compared with many of the simple front-ends developed by publishers, found their customer bases expanding. Publishers, however, noticed this development, and many of them have begun to improve their interfaces to their content, or in a few cases to undertake significant development of new interfaces that may well rival the traditional indexing and abstracting operations.

The format of digital journals has certainly evolved in the past decade. Ten years ago ASCII or plain text and HTML were the leading formats for electronic content. Now HTML, XML, and the Adobe’s portable document format (PDF) appear as the common containers for content. PDF continues to dominate as an electronic format of choice for display by publishers, while HTML/XML dominates as the format for journals published independently. The PDF format has been criticized for its digital bloat, inflexibility, and inability to easily hyperlink to other documents and files. There are some who also think the PDF is a fossil

format in the making.¹⁰ There has also been some criticism of PDF as being not completely accessible to the blind.¹¹

Yet PDF should not be written off as obsolete or arcane, because it is not a black-and-white choice of one format over another. Instead, it is a matter of offering multiple formats to address the differing needs of authors, scholars, and publishers of all stripes and persuasions. Adobe itself is addressing many of these issues with its Adobe XML architecture and the creation of XMP (XML Data Package) files that transform PDF documents into XML.¹² Many publishers have opted for more-baroque solutions to avoid a PDF/XML dichotomy, creating scripts that automatically convert files (representing specific documents in a journal’s archive) from HTML or XML to PDF on the fly, depending on requests from readers to servers. In any case, Adobe and PDF are not likely to become extinct any time in the near future, especially since 2003 marked the tenth anniversary of Adobe Acrobat.¹³ Sheer numbers are on the side of Adobe. A cursory survey in AltaVista reveals that there are 5.2 million PDF files on the World Wide Web; Google counts 10.8 million PDF files on the Web.

XML is continually making inroads because it provides distinct advantages over HTML (still widely in use in Internet publishing), fundamentally as a means to share structured information.¹⁴ Hence many publishers, scholars, and readers are seeing format issues simply in this way:

XML is all about structure and meaning, and little or nothing to do with appearance; PDF is all about appearance, and has little or nothing to do with structure and meaning.¹⁵

With this seeming abundance of digital journals, in a variety of formats to please almost everyone and accessible by a variety of tools, the obvious question seems to be: Is anyone reading any of these digital journals? A recent study demonstrated the importance of electronic journals in libraries. Scholars, especially scientists, are reading more than ever (for scientists, 216 readings per year, consuming 127 hours per person, compared with 150 readings per year in 1977), and they are reading more journals in digital formats than ever before—41 percent of all readings are in digital formats.¹⁶ This report was reinforced by the findings of a national survey sponsored by the Digital Library Foundation and the Council on Library and Information Resources. This survey found that 75 percent of survey faculty and graduate students use electronic journals.¹⁷ These trends coincide with larger trends in society relative to the use of the Internet. UCLA’s Internet Report *Surveying the Digital Future* found that 70 percent of all Americans who use the Internet regard electronic information as their most important source, taking precedence over television and newspapers.¹⁸

Hence, efforts in understanding the usage of digital resources made progress in 2003, thanks in part to initiatives like COUNTER (Counting Online Usage of Networked Electronic Resources). COUNTER is an attempt to provide utilitarian statistics on the use of digital journals from disparate sources, while providing significant data to publishers on the use of digital content.¹⁹ COUNTER scored a significant endorsement with Elsevier’s agreement in October 2003 to follow COUNTER’s Code of Practice.²⁰

Another factor in the greater use of digital journals was their improving functionality. From availability on handheld devices to document alert services,

improved reference linking within articles, cross-resource searching, ability to export citations into reference management software, and the ability to link articles from more and more different interfaces, there was constant development in interfaces used to locate journal articles. New features were branded with marketing names, e.g., “breadcrumb navigation” and SpringerLink (Springer-Verlag), SmartLinks, and CustomLinks (EBSCO), and Article Linker (SerialsSolutions).²¹

Digital Journal Management

Functionality was linked to another key trend of 2003—digital journal management, that is, helping libraries manage their electronic subscriptions. Services analyzing a library’s electronic holdings emerged. Some of the serials-control vendors began to make use of the data they glean from their clients to help libraries understand the changing landscape, while other parts of the industry helped us all to see how electronic journals are changing and affecting scholarly communication.

New services included providing “table of contents” views for journals included in aggregator databases, thereby allowing the reader to view an entire issue and browse it. E-journal management services began to provide overlap studies that enabled a library to see how much duplicate electronic content it had acquired from various sources. Digital journal management provided raw data for one of the first important studies of library holdings of electronic journals. An article by Peter McCracken, founder of SerialsSolutions, published in December 2003 in *Libri*, revealed some data about the percentage of library journal holdings that are electronic, ranging from a low of 39.5 percent for doctorate-granting institutions to a high of 86.5 percent for two-year colleges.²²

Increased access and improved functionality perhaps have a dark side as well. In December 2002 illegal access to online journals became an important issue as JSTOR made suppliers aware of the problem of open proxies.²³ Open proxies are computers that allow others the ability to use that computer to gain access to services. In the case of a university, an open proxy would allow anyone to connect from anywhere on the Internet and to gain access to anything for which that Internet address was authorized, e.g., the campus subscriptions to electronic journals and databases. For JSTOR, more than 51,000 articles from 11 journals were downloaded illegally. Efforts to improve Internet protocol (IP)-restricted access were continuing in 2003, and in particular, the protocol known as Shibboleth was showing great promise of offering a reasonable alternative for more reliable restrictive access.²⁴

Economics Show Stress

The economics of digital-journal publishing continued to show stress. The most prevalent models, based on subscriptions to paper with a surcharge for electronic, discourage licensors from going “all electronic” because savings are small and risk is still perceived to be high. Some publishers began to consider escrow deposit of databases to overcome fear of loss of access if libraries move to all-digital formats.

But the biggest news in 2003 was that many institutions, faced with constricting budgets, told publishers that the prices were too high. The argument that carried high pricing several years ago—the high development costs for making electronic products—no longer was a convincing one, in part because some libraries have become digital publishers as well. Many publishers had robust electronic publishing systems in 2003, and were driven to expand content to justify the expenditures they made in their infrastructure. The result was that publishers—big and small, commercial as well as professional associations—acquired publications of others (those who were not yet online and those who couldn’t afford to keep up with advances in interfaces).

Reports on Wall Street of hefty profit margins—approaching 30 percent in the case of a major European publisher—while the budgets of libraries were being reduced (along with the budgets of their parent organizations) only made the situation worse. Debate raged about whether the deals that sweetened large purchases—referred to as the “big bundle” or the “big deal”—and gave e-journal users access to more than the e-versions of the paper titles owned by their libraries, were worth the price. The largest research libraries in the United States saw a great deal of press coverage as they walked away from their big deals. Harvard University cancelled access to a group of 100 titles from Elsevier’s ScienceDirect. Cornell University Library worked with its University Senate to codify a strategy that includes reducing subscriptions and redirecting publications of the Cornell faculty into the more responsibly priced e-journals.²⁵ The University of California also echoed these activities as it launched an aggressive program of combating the rising costs of journals—paper and online. The Santa Cruz campus in particular actually suggested its faculty cut its ties with Elsevier and stop submitting articles to these journals because of the unfair model used to determine price of ScienceDirect.²⁶

The continuing evolution of digital journals promises to be exciting in 2004, as economics and technologies play an ever-increasing role, while usage continues to climb with greater and greater access to new titles born digital as well as enhanced back files. The battles bringing scholars and their institutions together on one side and publishers on the other will only become more dramatic, with little room for spectators.

Notes

1. Peter Swirski, 1997. *A Stanislaw Lem Reader*. Northwestern University Press, p. 51.
2. Wellcome Trust, 2003. “Economic Analysis of Scientific Research Publishing” at http://www.wellcome.ac.uk/en/images/SciResPublishing2_7445.pdf, p. 2.
3. Carol Tenopir and Donald W. King, 2000. *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*.” Special Libraries Association, p. 344, quoted in Amy Friedlander and Randi S. Bessette, 2003. *The Implications of Information Technology for Scientific Journal Publishing: A Literature Review*. National Science Foundation, Division of Science Resource Statistics, Directorate for Social, Behavioral, and Economic Studies, p. 15.
4. Betsy Mason, 2003. “Cell Editor Joins PLoS,” *The Scientist* (January 13) at <http://www.biomedcentral.com/news/20030113/05>.

5. *PLoS Biology* (<http://www.plosbiology.org>).
6. Christopher A. Reed, 2004. "Just Say No to Exploitative Publishers of Science Journals," *Chronicle of Higher Education*, vol. 50, no. 24, section B (February 20), p. B16.
7. National Information Standards Organization (NISO), 2003. "The OpenURL Framework for Context-Sensitive Services, Part 1: ContextObject and Transport Mechanisms," Draft, at <http://library.caltech.edu/openurl/PubComDocs/StdDocs/Part1-PC-20030311.pdf>; NISO, 2004. "NISO Standards Currently Balloting," updated February 9, 2004, at <http://www.niso.org/standards/balloting.html>.
8. See <http://arxiv.org/help/openurl>.
9. JSTOR, 2003. "Article-Level Linking into JSTOR Reveals Interesting Trends," *JSTORNEWS*, no. 7, issue 2 (June), at <http://www.jstor.org/news/2003.06/linking.html>.
10. See, for example, Peter B. Boyce, 1999. "Electronic Scholarly Journals," at <http://cdsweb.u-strasbg.fr/misc/epub-talk/7.html>, talk at Université Louis Pasteur, Strasbourg, France (29 January).
11. Janina Sajka and Joe Roeder, 2002. "PDF and Public Documents: A White Paper," American Foundation for the Blind, at http://www.afb.org/info_document_view.asp?documentid=1706.
12. Adobe, 2003. "Electronic Forms Solutions using XML and PDF: A Technology Primer for Adobe XML Architecture," Adobe White Paper, at http://www.adobe.com/enterprise/pdfs/xml_whitepaper.pdf.
13. Adobe, 2003. "Adobe Marks Technology Milestone with 10th Anniversary of Adobe PDF," press release of June 23, at <http://www.adobe.com/aboutadobe/pressroom/pressreleases/pdfs/200306/062303PDFANN.pdf>.
14. As described by Dan Connolly, 1998. "The XML Revolution," *Nature* (Web Matters, October 1), at <http://www.nature.com/nature/webmatters/xml/xml.html>.
15. Impressions Book and Journal Services, Inc., 2000. "XML and PDF: Why We Need Both," at http://www.impressions.com/resources_pgs/SGML_pgs/XML_PDF.pdf.
16. Donald W. King, Carol Tenopir, Carol Hansen Montgomery, and Sarah E. Aerni, 2003. "Patterns of Journal Use by Faculty at Three Diverse Universities," *D-Lib Magazine*, vol. 9, no. 10 (October), at <http://www.dlib.org/dlib/october03/king/10king.html>.
17. Summary in Deanna B. Marcum and Gerald George, 2003. "Who Uses What? Report on a National Survey of Information Users in Colleges and Universities," *D-Lib Magazine*, vol. 9, no. 10 (October), at <http://www.dlib.org/dlib/october03/george/10george.html>.
18. UCLA Center for Communication Policy, 2003. "Surveying the Digital Future: Year Three," at <http://ccp.ucla.edu/pages/internet-report.asp>.
19. COUNTER, 2003. "About COUNTER," at <http://www.projectcounter.org/about.html>.
20. "COUNTER Code of Practice adopted by Elsevier ScienceDirect," press release of October 27, at http://www.projectcounter.org/newsrelease_oct03.pdf.
21. Bonnie Lida Rogers and Barbara Chaparro, 2003. "Breadcrumb Navigation: Further Investigation of Usage," *Usability News*, vol. 5, no. 2, at <http://psychology.wichita.edu/surl/usabilitynews/52/breadcrumb.htm>.
22. Peter McCracken, 2003. "A Comparison of Print and Electronic Journal Holdings in Academic and Public Libraries," *Libri*, vol. 53, no. 4 (December), pp. 237–241.
23. JSTOR, 2003. "Frequently Asked Questions About Open Proxies and JSTOR," at http://www.jstor.org/about/open_proxy_faqs.html.
24. <http://shibboleth.internet2.edu>.
25. Cornell University, Faculty Library Advisory Board, 2003. "Resolution Regarding the University Library's Policies on Serials Acquisitions, with Special Reference to Negotiations with Elsevier," submitted at the December 17, 2003, meeting of the University Faculty Senate, at <http://web.cornell.edu/UniversityFaculty/FacSen/031217SenateMtg/LibraryResolution.pdf>.
26. Jondi Gumz, 2003. "UCSC Faculty Threaten to Boycott Publisher Over Journal Subscriptions' Costs," Santa Cruz (California) *Sentinel* (October 31), at <http://www.santacruzsentinel.com/archive/2003/October/31/local/stories/05local.htm>.