VOLUME 37 NUMBER 3

ACCOUNTS of CHEMICAL RESEARCH®

MARCH 2004

Communicating Original Research in Chemistry and Related Sciences

LORRIN R. GARSON
Information Technology, Publications Division, American Chemical Society, 1155 Sixteenth Street, N.W., Washington, District of Columbia 20036

Received August 1, 2003

ABSTRACT
The availability of scientific information in electronic form is the convergence of traditional journal publishing, electronic communications, and the widespread availability of computer technology. This revolution in scientific communication has its roots in developments that started in the mid-19th century and culminated with the extraordinary progress in telecommunications and computer technology in the latter years of the 20th century. Eighty-three percent of scientific journals are now available online. The benefits of electronic journals include rapid publication, instantaneous linking to external information sources, and the capability to deliver new types of information. To date neither electronic-only nor preprint servers have been well received by the chemical sciences community. Continued advances in telecommunications, computer technology, and acquisition of scientific data in structured formats hold promise for even greater advances in communication of scientific information.

> "Despite the limitations which any single discipline must have in providing an over-all view of science, chemistry, because of its fundamental position in the science hierarchy, probably offers the broadest available index to the growth of science and to the trends in the communication of science information." Baker, Tate and Rowlett[1]

The delivery of scientific information via the World Wide Web represents the convergence of two important communication channels—printed scientific journals and electronic communications—plus widespread availability of inexpensive computer technology. The first two printed scientific journals were published in 1665, the French journal *Journal des Scavans* and the British publication *Philosophical Transactions*. Electronic communications and the development of computers have an earlier history than is often apprecited.

- 1843—patent for FAX was awarded to Alexander Bain.
- 1844—Samuel Morse installed a telegraph line between Baltimore, Maryland, and Washington, DC.
- 1876—Alexander Graham Bell patented the telephone.
- 1890—Herman Hollerith was awarded a contract for processing the 1890 U.S. census using punched cards.
- 1924—Hollerith's Tabulating Machine Company becomes IBM.
- 1941—Konrad Zuse developed the first programmable calculator using binary numbers and Boolean logic.
- 1964—IBM released the IBM model 360 mainframe computer.
- 1965—Digital Equipment Corporation (DEC) introduced the PDP-8 minicomputer.
- 1969—Honeywell sold its model H316 "Kitchen Computer" at Nieman Marcus priced at $10 600 ($53 087 in 2003 dollars).
- 1969—U.S. Department of Defense initiated the ARPANet between military installations and universities.
- 1974—Vint Cerf and Bob Kahn proposed connecting networks together to form an "Internet".
- 1977—Apple Computer Company introduced the Apple 1 computer.
- 1979—CompuServe went online.
- 1981—IBM introduced the IBM PC.
- 1992—Tim Berners-Lee spawned the World Wide Web with the release of hypertext markup language (HTML) and hypertext transport protocol (HTTP) specifications.

Lorrin R. Garson was born in Virginia, Minnesota. He received his B.A. from the University of California at Riverside in 1961 and M.S. and Ph.D. from the University of Maine in 1964 and 1967. He was on the faculty in the Medicinal Chemistry Department, College of Pharmacy, at the University of Tennessee in Memphis from 1967 to 1974. In 1974, he joined the staff of the American Chemical Society in the Publications Division. His research interests in the past 30 years have focused on employing technology for the efficient creation of publications and electronic delivery of chemical information.

10.1021/ar0300017 CCC: $27.50
Published on Web 02/10/2004

Electronic communication of scientific information first occurred with secondary publications. As early as 1960, Chemical Abstracts Service produced samples of *Chemical Titles*, which was subsequently published biweekly in 1961 on magnetic tape. In 1965, off-line batch searching of *Chemical Titles* was available commercially. In 1967, *Polymer Science & Technology* (POST) was published in printed form as well as on magnetic tape. *Chemical Abstracts Service Source Index* (CASSI) was made available on magnetic tape in 1970. In 1974, Lockheed Dialog licensed files from Chemical Abstracts Service to make them available online, as did Bibliographic Retrieval Services (BRS) in 1976. In 1980, CAS Online was established with a chemical structure search and display system. Other secondary services such as the National Library of Medicine (NLM), BIOSIS, and Thompson ISI also created electronic delivery systems in the 1970s and 1980s. In December 1983, the creation of Scientific and Technical Information Network (STN) International was announced with the first operational link established in 1984 between Chemical Abstracts Service in Columbus, Ohio, and FIZ (Fachinformationszentrum Energie, Physik, Mathematik, GmbH) in Karlsruhe, Germany.[2] STN International was established as a homogeneous international network long before the public Internet became a reality.

In 1980, the American Chemical Society, in cooperation with Chemical Abstracts Service and BRS, made available a database of 1000 articles from the *Journal of Medicinal Chemistry* to a group of chemists for testing, which was arguably the first full-text scientific journal available online. In 1982, all sixteen of the ACS's journals were made available online through BRS.[3] During 1983−1986, the ACS journals were made available through STN International, as were selected journals from Elsevier, John Wiley & Sons, the Royal Society of Chemistry (RSC) and others. These implementations suffered serious defects that limited their adoption and use. Only ASCII characters were permitted; thus, for example, the α character was represented by the string ".alpha.". Tabular material was unavailable as were graphic data, thus excluding line art, half-tones, and color. Even if graphics had been available, the slow speed of dial-up telecommunications at the time would have made downloading impractical. Despite the lack of success with these endeavors, experience gained in creating these systems was very valuable in traveling the road to the World Wide Web.

During 1989−1995, the ACS Publications Division, Bellcore, Chemical Abstracts Service, Cornell University, and Online Computer Library Center (OCLC) collaborated in what became known as the CORE project.[4,5] This was an effort to create a prototype digital library at Cornell University using the ACS journals as the data source and software from OCLC as the user interface and back-end system. In 1992−1997, the ACS participated with approximately 20 other publishers in the Red Sage Project to create a prototype electronic library at the University of California at San Francisco.[6,7] In the Red Sage Project, RightPages software from Bell Laboratories was used for the user interface and underpinning data structures. Both CORE and Red Sage were reasonably successful; however, the arrival of the World Wide Web made these efforts obsolete before they could be brought to fruition.

The Web, availability of broadband telecommunications, and inexpensive desktop computers had a dramatic enabling effect on electronic publishing. For the American Chemical Society, the following events occurred:

- 1995−Supporting Information for the *Journal of the American Chemical Society* was made available on the Web for both Mosaic and Gopher browsers.
- 1996−*The Journal of Physical Chemistry* was made available on the Web on the occasion of the journal's 100th anniversary.
- 1997−All 26 ACS journals were made available on the Web.
- 2002−The ACS Journal Archives were made available on the Web.

Peter Stang, Editor of the *Journal of the American Chemical Society*, in an editorial wrote, "Electronic publishing and the World Wide Web represent the biggest revolution in publishing and the dissemination of ideas since Johannes Gutenberg invented the modern printing press in 1455."[8] In a recent survey of 275 journal publishers, Cox and Cox[9] have reported that of the scientific, technical, and medical (STM) titles, 83% are available online.

Notable Players in Electronic Publishing

Unquestionably, the age of electronic dissemination of scientific information has arrived and is an integral part of STM publishing. Most STM publishers now deliver both print and Web products and provide Web-based manuscript submission systems for their authors.

Elsevier is the largest commercial publisher in science and offers an increasingly integrated line of products. Elsevier's journals are available through a variety of purchase plans through Science Direct. Science Direct is linked to Elsevier's Scirus, a free "Google-like" search engine for the sciences. Elsevier's ChemWeb is a portal for chemistry and is tightly coupled with Elsevier's preprint server. Elsevier's MDL, a software company that largely focuses on the drug discovery market, is also associated with CrossFire Beilstein.

At the other end of the business model spectrum lies the Public Library of Science (PLoS). This is a venture in which it is proposed authors, rather than journal purchasers, would financially support the publishing enterprise by paying a $1500 per manuscript fee. The production cost per manuscript for ACS journals in 2002 was $1544 exclusive of paper, printing, and distribution,[10] which is remarkably close to PLoS's $1500 fee. PLoS was established in October 2000 as a nonprofit organization and has its roots in a protest movement originating at Stanford University in which scientists were asked to boycott those publishers that would not allow unrestricted free access to their journal articles six months after publication.

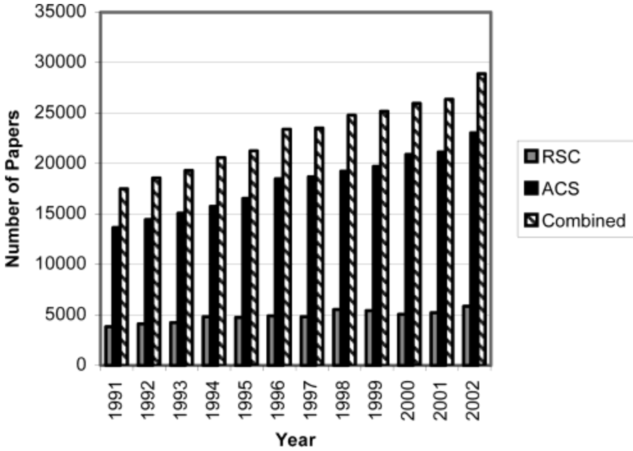

FIGURE 1. Journal Articles Published by the ACS and RSC, 1991–2002.

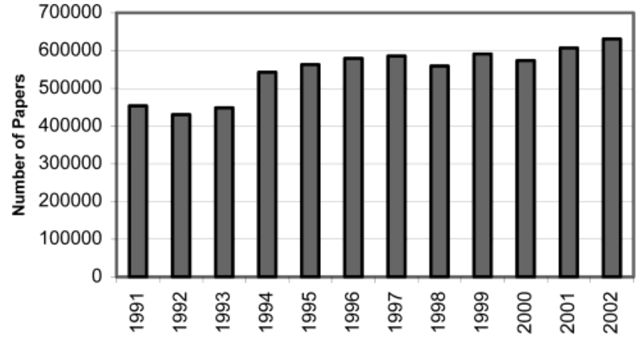

FIGURE 2. Journal Articles Covered in *Chemical Abstracts*, 1991–2002.

Subsequently, PLoS announced they would become a publisher.[11]

The Budapest Open Access Initiative,[12] PubMed Central,[13] BioMed Central,[14] and the concept of "self-archiving" advocated by Stevan Harnad[15] are endeavors similar to PLoS in that they have a common objective to make information freely available. These efforts should not be confused with the Open Archives Initiative,[16] which is an activity to promote interoperability standards to facilitate efficient dissemination of content. PubMed Central is an archive of life science journal literature operated by the U.S. National Library of Medicine and acts as a distributor for 81 journals plus an additional 57 titles from BioMed Central. Of these 57 journals, 47 published the first paper in 2001, five in 2000, and one in 2002. By September 2001, the combined journals had published 263 papers, rising to 706 by March 2002 and 1713 by July 2003. The amount of material being published is growing in the BioMed Central journals but is a very small fraction of the amount of material published in biology and medicine during this period. PubScience, operated by the U.S. Department of Energy, was to have been the counterpart to PubMed Central focusing on the physical sciences and other energy-related disciplines.[17] However, PubScience was discontinued on November 4, 2002. The viability of such endeavors is dependent upon attracting authors and efficient operations and, in some cases, upon publishers freely making their materials available. Ultimately success is dependent upon the financial support of journal purchasers—unless subsidies are provided from granting agencies or the government.

Professional society publishers, which lie in the middle of the business model spectrum, generally use the same business model as commercial publishers, that is, revenues to support the publishing enterprise are derived from selling subscriptions. However, because society publishers operate on a not-for-profit basis and do not generate profits for shareholders, their prices are generally significantly lower than commercial publishers.[18] The American Chemical Society and the Royal Society of Chemistry are the two largest noncommercial publishers of chemical information. In Figure 1 are shown the numbers of papers published by both organizations[19] from 1991 through 2002. In Figure 2 are shown the numbers of journal articles covered in *Chemical Abstracts* during the same period. During these 12 years, the two societies combined published 4.2% of the journal articles cited in *Chemical Abstracts*. Allowing for another 1–2% of articles that may be published by other society publishers and government agencies, commercial publishers produce approximately 94% of the world's journal literature in chemistry. Clearly the chemical sciences community strongly supports commercial publishing by submitting manuscripts, acting as referees, serving as editors, and encouraging their institutions to purchase the journals in which they publish.

To date electronic-only journals in chemistry have not been well received. The *Internet Journal of Chemistry*[20] (*IJC*) was one of the earliest electronic-only publications in chemistry, established in 1998. However, only 101 articles have been published in this journal. The number of articles published in the *IJC* has steadily decreased since its inception, from 38 published in 1998 to four in 2003 (through July 17th). The *IJC* has been quite innovative in introducing features that could only be possible in an electronic publication. *PhysChemComm*,[21] published by the Royal Society of Chemistry, was also established in 1998. As of May 15, 2003, 284 papers have been published in this journal and the number has been rising steadily from four in 1998 to 177 in 2002. *CrystEngComm*[22] and *Geochemical Transactions*,[23] also published by the RSC, were started in 1999 and 2000 and have published 245 and 38 articles, respectively. The survivability of such journals is problematic. Electronic-only journals are more likely to evolve from current print/electronic journals that will sooner or later cease being printed. Tenopir and King have described many factors pertinent to electronic journals.[24]

Preprint servers have become well established as a means of communication in mathematics,[25,26] physics,[27] and computer science.[26] However, preprint servers in chemistry have not been well accepted, possibly because of greater commercial activities in chemistry. Although the PrePrint Network,[28] operated by the Department of Energy, lists 17 chemistry "collections", Elsevier's ChemWeb appears to be the only substantial preprint server operational at this time. Since its inception on August 21, 2000, 726 papers have been posted on ChemWeb through July

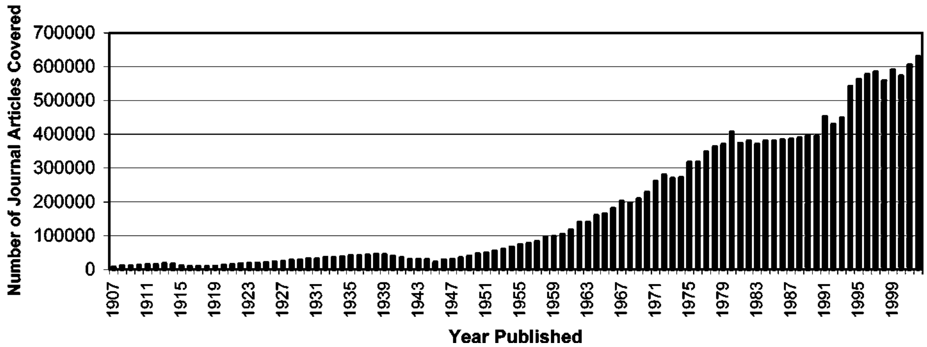

FIGURE 3. Journal Articles Covered in *Chemical Abstracts*, 1907−2002.

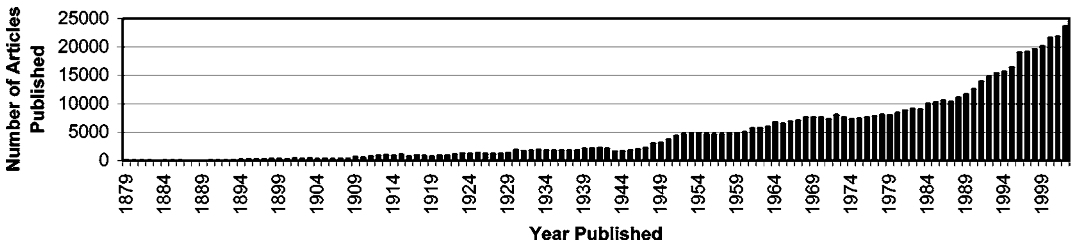

FIGURE 4. Journal Articles Published by the ACS, 1879−2002.

15, 2003. During this period, records for approximately 1.7 million journal articles were published in *Chemical Abstracts*,[29] and thus the 726 preprints represent 0.04% of the primary literature published. In general, preprint servers rely on postpublication comments of readers for peer review. Such a policy allows for the publication of new discoveries and theories that may be rejected by the traditional review process but also permits the publication of information of dubious quality. The chemical physics preprint database at Brown University[30] seems to be moribund with most of its links broken and some (all?) of its reprints moved to the Los Alamos National Laboratory (LANL) preprint server.[27]

Economics of Publishing

In the past decade, particularly since the advent of electronic publishing, there has been much discussion in the scientific community, in the library community, and among publishers about the cost of journals, who owns the information, who should publish scientific information, and who should pay. I propose that the root cause of this turmoil is largely the exponential growth in scientific information. Figure 3 shows the growth in the number of articles covered in *Chemical Abstracts* (excluding patent information) from 1907 through 2002. Similarly, Figure 4 shows the number of articles published by the American Chemical Society from 1879 through 2002. The growth in scientific information in general likely parallels the growth in chemistry. Mabe[31,32] has shown that the number of journals and articles published has grown at an average rate of 3.46% from 1800 to 2003. He also believes this rate of growth is caused by a similar rate of growth in the number of researchers. Mabe has concluded that, for the twentieth century, the rate of journal growth is self-organizing and in equilibrium.

When electronic publishing first became practical, many expected such publishing would significantly reduce the cost of publishing, and thus subscription prices would fall or at least stabilize. This expectation has not come to fruition. "First copy costs" for scientific journals are about 80% of the total cost for publishing an article regardless of how that article is distributed. Distribution, whether electronic or print, is the remaining 20%. Publishers who have created electronic journals have had to make considerable investments in technology and, because of customer demand, continue to print journals as well. High circulation publications such as *Science* and *Nature* would be expected to have lower than 80% first copy costs because of their higher print, paper, and distribution expenses associated with a large subscriber base.

Ownership of information and who should publish is a complex issue. Historically scientists have been little concerned with these issues as long as their discoveries could be published and widely disseminated. Before the Internet and World Wide Web, the barriers to individuals or their institutions publishing were very high. At this time, there is some movement toward academic institutions taking a more active role in electronic journal publishing.[33−35]

Purchasers of scientific journals frequently complain that journal prices rise at a rate much greater than inflation implying that publishers' profits are unconscionable. The inflation being referred to in this context is monetary inflation and not "inflation" in the amount of information being published. Despite improved efficiencies in journal production, largely the benefit of investments in new technology, these efficiencies do not offset costs associated with publishing the increase in research output. The cost of duplicating research, not knowing the work has already been done, far exceeds the costs of purchasing information. Publication expenses are proportional to the amount of material published, and therefore, prices for subscriptions must inevitably rise to reflect costs due to monetary as well as "submission inflation". Scientists may wish to look in the mirror and utter the immortal words of Pogo, "We have met the enemy and he is us."[36] Parks[37] has expressed this more eloquently in that participants (authors, editors, referees, publishers, librarians, readers, and users) in the publication process have little or no incentive to stop publishing in current journals. Plasmeijer[38] argues that journal price increases are largely the fault of librarians who are willing to pay to build their collections, the mark of a quality library, and from pressure from researchers who do not pay and have no budget constraints for acquiring information. Such a view is simplistic and perhaps a bit cynical. King and Tenopir[39] have described in detail the many factors that have contributed to the extraordinary rise in expenditure for scientific journals—from an estimated $5.05 billion in 1975 ($15.6 billion in 1998 dollars) to $45 billion in 1998.

The notion of an "information explosion" is not new. Chemical Abstracts was established in 1907, in part, because of the explosion in chemical information at the turn of the 20th century. In writing in 1974 on the information explosion, Ernest Eliel[40] quoted Vannevar Bush, writing in 1945, "...publication has been extended far beyond our present ability to make real use of the record." As shown in Figures 3 and 4, the quantity of chemical information published in the past half century has greatly increased, but so have the tools to create and manage this information.

Benefits of Electronic Publishing

There are a number of features that dramatically differentiate electronic and traditional paper publishing. Speed of publication is one obvious difference. Although the speed of peer review does not yet seem to have been noticeably impacted, the submission of manuscripts via the Web, transmission of articles to reviewers, especially between continents, and delivery of galley proofs to authors are an obvious savings in time over traditional mail service or even expensive "overnight" express services. Illustrating this dramatic reduction in time, two articles have been published in *Organic Letters*, on the Web, 8 days after being received from the author, which included peer review.[41,42]

Linking is now ubiquitous in electronic journals, principally in HTML versions. The ACS Web journals have links from bibliographic references directly to the primary source in ACS journals and links to others publishers' articles via ChemPort,[43] Medline/PubMed, and GenBank through the National Center for Biotechnology Information,[44] links to the Protein Data Bank,[45] and links to Supporting Information. Most publishers have similar linking. Efforts to link across publishers' domains have received considerable attention in the last several years. For chemistry and related sciences, ChemPort provides linking to 3723 journals from 197 publishers (as of July 11, 2003). CrossRef, which serves across all scholarly research, provides links to 8156 journals from 221 participating publishers. CrossRef uses the DOI (digital object identifier), which provides a protocol for publishers to create unique keys for digital objects. According to Cox and Cox,[9] CrossRef is the primary mechanism for citation linking with 65% of large publishers and 33% of small publishers participating.

Various types of information can be delivered through electronic publishing that are impossible or impractical with traditional print publishing.

(1) Mohamed-Kassim and Longmire[46] have published two full-motion videos of image sequences from low-magnification particle image velocimetry (PIV) experiments in studying the drop impact on a liquid−liquid interface.

(2) As a supplement, Smith[47] has published the source code to a collection of Fortran-90 routines used for evaluating the gamma function and related functions using the FM multiple-precision arithmetic package. Having the source code in a machine-readable form is of much greater value than had it been published on paper. Moreover, it would have been impractical to publish this long listing on paper.

(3) Baker, Carfagna, Gao, Dow, Li, Searfoss, and Ryan[48] have published a "mini Web site" in Supporting Information for their paper entitled "Temporal Gene Expression Analysis of Monolayer Cultured Rat Hepatocytes". The Supporting Information consists of 4379 files (18.6 Mb) of a detailed presentation of microarray data on transcripts that demonstrated differential gene expressions. The data can also be acquired in Microsoft Excel format consisting of a file with 4117 rows, which allows facile downloading into databases or applications.

(4) Moorjani, Jia, Jackson, and Handcock[49] have published in Supporting Information a movie showing the left to right movement of a microtubule, bumping into the photoresist wall and then being redirected back into the channel. An analysis of why microtubules do not buckle at higher motor densities is also included in the Supporting Information.

(5) Bond, Feeder, Redman, Teat, and Sanders[50] have published in Supporting Information X-ray crystallographic data as CIF files and diagrams of the molecular units showing displacement ellipsoids and deviations from molecular planes. CIF files can be validated or viewed with freely available software such as enCIFer[51] and Mercury.[52] Publication of X-ray data as CIF files is now a common practice. Also included for this paper are 3D rotatable images of five compounds in PDB format, published as "Web Enhanced Objects".

(6) Hegelund, Bürger, and Pawelke[53] have published extensive analyses of high-resolution IR ν5, ν3, ν6, and ν2 bands of FNO_2.

The digitization of "pre-Web" scientific material is an important opportunity for the scientific community to

12-Acetyldehydroabietonitrile (**1b**)[7].--To a solution of 26.8 g of dehydroabietonitrile (Hercules Inc.) in 200 ml of tetrachloroethane cooled to 0° (dry atmosphere) was added 10 ml of acetyl chloride and 26.5 g of aluminum chloride. The mixture was stirred at 0° for 24 hr, poured into water and extracted with $CHCl_3$. The washed and dried extract was evaporated; the residue was recrystallized from methanol-water, yield 27.5 g of **1b**, mp 157.5-158°, $[\alpha]_D^{25}$ + 97.0° (c, 0.320; 95% ethanol), ir bands at 2247 (nitrile) and 1692 cm^{-1}; nmr signals at 1.18 (C-10 methyl), 1.26d (*J*=6.5 Hz, isopropyl), 1.40 (C-4 methyl), 2.50 (methyl ketone), 7.00 and 7.26 ppm (aromatic protons).

Anal. Calcd for $C_{22}H_{29}NO$: C, 81.69; H, 9.04; N, 4.33. Found: C, 81.61; H, 9.19; N, 4.54.

FIGURE 5. Miniprint from *J. Org. Chem.* electronically magnified 400%. Reproduced from ref 55. Copyright 1974 American Chemical Society.

attain better access to older literature. While there is concern about the long-term storage of digital information, digitization of deteriorating paper journals constitutes a form of preservation in itself, as for example, The British Library's project to digitize the Gutenberg Bible.[54] In this case, not only is preservation accomplished, but also the document is now widely available to scholars with no risk of damage by using the original. In the 1970s, to save composition, paper, printing, and distribution costs, the ACS published experimental details in *The Journal of Organic Chemistry* in "miniprint", which consisted of nine pages of camera-ready material provided by authors reduced to fit on one printed journal page. Interestingly, in the digitization process the compressed material becomes much easier to read by simple electronic magnification as shown in Figure 5.[55]

In 1977, The Royal Society of Chemistry, Société Chimique de France, and Gesellschaft Deutscher Chemiker collaborated to create the *Journal of Chemical Research* in which synopses of papers were published in *J. Chem. Res. (S)* and corresponding fulltext was published concurrently, with the synopsis, in *J. Chem. Res. (M)*. During the same period, the American Chemical Society investigated the feasibility of publishing the *Journal of the American Chemical Society* as a "dual journal", which would consist of summary and archival journals, but concluded there was no mandate for publishing the journal in such a format.[56] In some ways, these efforts were forerunners of the electronic journals we have today.

Most if not all scientific publishers have announced plans to digitize their archival publications. The ACS was among the first publishers to make available its entire collection of about three million pages going back to 1879.[57] The Royal Society of Chemistry has announced plans to complete digitization of its journals by the end of 2003, for the period 1841−1996, and to make this collection available in 2004.[58] The American Institute of Physics made its backfile (back to 1975) available to subscribers in 2003, digitized journals back to 1968 in 2003, and has plans to complete its digitization in 2004 of all its journals back to about 1931.[59] Elsevier's ScienceDirect has been loading its digitized material since January 2001 and expects to complete digitization of its entire journal list in 2003.

Appearance and Content

Experience has shown users have a strong preference for articles in PDF format over HTML. This has been verified by Davis and Solla[60] who have analyzed user behavior at Cornell University and have commented, "individuals are using the system like a networked photocopier." Considering the importance users place on the linking function, which is absent from the PDF format, this is rather surprising. Examination of the first major book printed in the west (the Gutenberg Bible, Mainz, 1454−1455[54]) shows the format and text mimicking books scribes had previously created manually. At that time, this is what 15th century readers expected; today 21st century readers expect journal articles on the Web to look like they have in print for the past 300+ years. However, there is no compelling reason to believe the linear, top-down format of the printed journal article is the optimum for electronic journals. Little effort appears to have been made to optimize the readability of HTML, and this format is far less readable than the corresponding material in PDF. There is also anecdotal evidence that users prefer to read a scientific article on paper rather than on a computer display. But what of the future when the electronic version is significantly richer in content than the printed journal or when a journal is no longer printed? Clearly efforts need to be made to improve the appearance and readability of electronic documents.

Challenges in Scientific Publishing

Displaying special characters has been one of the challenges in creating HTML files for electronic delivery. In STM publishing of electronic journals, there are usually hundreds if not thousands of small graphic images that are created and displayed for special and accented characters. This is an inefficient process and does not permit scaling these graphic characters. Several publishers are collaborating on the STIX project[61] to create a comprehensive set of fonts (7866 glyphs) that would meet the needs for both print and electronic publishing. These font sets will be available under a royalty-free license. The situation is similar in the display of mathematics. For publications in chemistry, most mathematics is displayed as graphic images derived from the composition process. The World Wide Web Consortium (W3C) has published specifications for MathML,[62] which provides real-time rendering of mathematics and provides scalable rather than static images. Our experimentation has shown MathML to be rendered well with material from the ACS journals, but the current state of browser technology does not yet render mathematics in publication quality comparable to print.

Although accurate and good-looking rendering of data for display purposes is important, the opportunity now exists for the chemical community to acquire scientific data in well-defined, standardized formats. To date, electronic publishing has largely been focused on identifying data associated with format (i.e., identification of authors' names and affiliations, titles, abstracts, references, paragraphs of text, table titles, figure captions, etc.). We are seeing some examples of true scientific data collection in the creation of external databases, particularly in thermodynamics and biological data, but not yet as part of the journal publishing enterprise itself. Capturing scientific data in prescribed formats at the time an author prepares a manuscript for publication would potentially significantly enrich the value of such papers and make the creation of secondary databases more efficient and complete. Some progress has been made to encode chemical information in extensible markup language (XML) as CML (chemical markup language) by Murray-Rust and Rzepa.[63–68] However, there is much more work to be done, and authoring software tools are yet to be developed.

Efforts are also being made to develop data format standards in XML for analytical instruments. The standards group ASTM E13.15 is attempting to develop a specification for a common core of elements in XML format that would address data interchange and archiving issues that could be used across all analytical techniques (MS, NMR, IR, gas chromatography, etc.). This core specification is known as AnIML (analytical information markup language). Subsequently instrument-specific specifications will be developed. In time, these XML specifications are likely to replace the current Joint Committee on Atomic and Molecular Physical Data (JCAMP) standards.[69]

The Future

Considering that electronic journal publishing is less than 10 years old and the remarkable progress that has been made in this short time, we can expect many exciting advances to come in the years ahead. The costs for publishing large data sets and extensive experimental details by traditional means are greatly reduced by electronic publishing, although restraint on the part of authors and editors will be increasingly important. We can look forward to the acquisition of scientific data in evolving standard formats that could significantly enhance our knowledge and understanding of chemistry and related sciences. Heretofore there were high barriers to enter the publishing business, including publishing technology and distribution channels, which are now largely removed by electronic publishing. Nevertheless, editorial talent, capital, and business acumen remain significant barriers to establishing new, successful publishing operations. In particular, sales and marketing offer the greatest challenge to would-be publishers and are important strengths of established publishers. Whether revenues essential for operations (and profits in the case of commercial publishers) are generated from subscription sales, authors' publication fees, or subsidies from governments or granting agencies remains uncertain. Baring significant changes in copyright law or the unlikely widespread abandonment of current publishing practices by the scientific community, established publishers are likely to be dominant for the foreseeable future.

It is a pleasure to acknowledge my colleagues Robert Bovenschulte, David Martinsen, Jack Ochs, and Mary Scanlan for many hours of interesting discussion of publishing issues and their critique of this manuscript.

References

(1) Baker, D. B.; Tate, F. A.; Rowlett, R. J. Changing Patterns in the International Communication of Chemical Research and Technology. *J. Chem. Doc.* **1971**, *11*, 90–98.
(2) Baker, D. B. Chemical Information Flow Across International Borders: Problems and Solutions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 55–59.
(3) Garson, L. R.; Terrant, S. W.; Cohen, S. M. Presented at the 187th National Meeting of the American Chemical Society, St. Louis, Missouri, April 9, 1984.
(4) Entlich, R.; Olsen, J.; Garson, L. R.; Lesk, M.; Normore, L.; Stuart, W. Making a Digital Library: The Contents of the CORE Project. *ACM Trans. Inf. Syst.* **1997**, *15*, 103–123. http://doi.acm.org/10.1145/248625.248627.
(5) Entlich, R.; Garson, L. R.; Lesk, M.; Normore, L.; Olsen, J.; Weibel, S. Testing a Digital Library: User Response to the Core Project. *Libr. Hi Tech* **1996**, *14*, 99–118.
(6) Lucier, R. E.; Brantley, P. The Red Sage Project. An Experimental Digital Journal Library for the Health Sciences. *d-Lib Magazine* **1995**, *1*, August 1995. http://www.dlib.org/dlib/august95/lucier/08lucier.html.
(7) Red Sage Final Report. http://www.springer-ny.com/press/redsage/.
(8) Stang, P. J. 124 Years of Publishing Original and Primary Chemical Research: 135,149 Publications, 573,453 Pages, and a Century of Excellence. *J. Am. Chem. Soc.* **2003**, *125*, 1–8.
(9) Cox, J.; Cox, L. *Scholarly Publishing Practice: The ALPSP Report on Academic Journal Publishers' Policies and Practices in Online Publishing*; Association of Learned and Professional Society Publishers: 2003.
(10) Bovenschulte, R. D. Costs of Publication. Presented at The National Academies Symposium on Electronic Scientific, Technical, and Medical Journal Publishing and its Implications, Washington, DC, May 19–20, 2003.
(11) Brower, V. Public Library of Science Shifts Gears. *EMBO Rep.* **2001**, *2*, 72–973. http://www.nature.com/cgi-taf/DynaPage.taf?file=/embor/journal/v2/n11/full/embor282.html.
(12) Budapest Open Access Initiative. http://www.soros.org/openaccess/.
(13) PubMed Central, an Archive of Life Science Journals. http://www.pubmedcentral.nih.gov/.
(14) BioMed Central. http://www.biomedcentral.com/.
(15) Okerson, A.; O'Donnell, J. *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*; Association of Research Libraries: Washington, DC, 1995.
(16) Open Archives Initiative. http://www.openarchives.org/.
(17) PubSCIENCE a New Web-Based Tool. http://www.pnl.gov/energyscience/11–99/art2.htm.
(18) University of Wisconsin Chemistry Journal Cost Study. http://www.library.wisc.edu/libraries/Chemistry/cost.htm.
(19) Parker, R., Royal Society of Chemistry, Personal Communication, June 19, 2003.
(20) *Internet Journal of Chemistry*. http://www.ijc.com/.
(21) *PhysChemComm*. http://www.rsc.org/is/journals/current/PhysChemComm/pccpub.htm.
(22) *CrystEngComm*. http://www.rsc.org/is/journals/current/crystengcomm/cecpub.htm.
(23) *Geochemical Transactions*. http://www.rsc.org/is/journals/current/geochem/geopub.htm.
(24) Tenopir, C.; King, D. W. *Towards Electronic Journals. Realities for Scientists, Librarians and Publishers*; Special Libraries Association: Washington, DC, 2000.
(25) Directory of Mathematics Preprint and e-Print Servers. http://www.ams.org/global-preprints/.
(26) The Computing Research Repository (CoRR). http://arxiv.org/archive/cs/intro.html.
(27) arXiv.org e-print archives. http://xxx.arxiv.cornell.edu/.
(28) PrePrint Network. http://www.osti.gov/preprints/.
(29) CAS Statistical Summary, 1907–2002. http://www.cas.org/EO/casstats.pdf.
(30) Brown University Chemical Physics Preprint Database. http://www.chem.brown.edu/chem-ph.html.
(31) Mabe, M. The Growth and Number of Journals. *Serials* **2003**, *16*, 191–197.

(32) Mabe, M.; Amin, M. Growth Dynamics of Scholarly and Scientific Journals. *Scientometrics* **2001**, *51*, 147−162.
(33) University of California eScholarship Repository. http://repositories.cdlib.org/escholarship/.
(34) UC to Launch Open-Access Journals, *The Scientist*, http://www.biomedcentral.com/news/20030616/03.
(35) DSpace Federation. http://www.dspace.org/.
(36) Kelly, W.; Crouch, B., Jr. *The Best of Pogo*; Simon & Schuster: New York, 1982.
(37) Parks, R. P. The Faustian Grip of Academic Publishing. *J. Econ. Methodol.* **2002**, *9*, 317−335.
(38) Plasmeijer, H. W. Pricing the Serials Library: in Defense of a Market Economy. *J. Econ. Methodol.* **2002**, *9*, 337−357.
(39) King, D. W.; Tenopir, C. Scholarly Journal and Digital Database Pricing: Threat or Opportunity? In *Bits and Bucks: Economics and Use of Digital Libraries*; Lougee, W., Mackie-Mason, J., Eds.; MIT Press: Cambridge, MA, 2003, in press. http://web.utk.edu/~tenopir/eprints/database_pricing.pdf.
(40) Eliel, E. L. Aspects of the Information Explosion. *J. Phys. Chem.* **1974**, *78*, 1339−1343.
(41) Harmata, M.; Rashatasakhon, P. Intramolecular 4 + 3 Cycloadditions. Aspects of Stereocontrol in the Synthesis of Cyclooctanoids. A Synthesis of (+)-Dactylol. *Org. Lett.* **2000**, *2*, 2913−2915.
(42) Cheng, D.; Kreethadumrongdat, T.; Cohen, T. Allylic Lithium Oxyanionic Directed and Facilitated Simmons-Smith Cyclopropanation: Stereoselective Synthesis of (±)-cis-Sabinene Hydrate and a Novel Ring Expansion. *Org. Lett.* **2001**, *3*, 2121−2123.
(43) ChemPort. http://www.chemport.org/.
(44) National Center for Biotechnology Information (NCBI). http://www.ncbi.nlm.nih.gov/.
(45) Protein Data Bank. http://www.rcsb.org/pdb/.
(46) Mohamed-Kassim, Z.; Longmire, E. K. Drop Impact on a Liquid−Liquid Interface. *Phys. Fluids*, *15*, 3263−3273. http://ojps.aip.org/dbt/dbt.jsp?KEY=PHFLE6&Volume=15&Issue=11.
(47) Smith, D. M. Algorithm 814: Fortran 90 Software for Floating-Point Multiple Precision Arithmetic, Gamma and Related Functions. *ACM Trans. Math. Software (TOMS)* **2001**, *27*, 377−387. http://portal.acm.org/citation.cfm?doid=504210.504211.
(48) Baker, T. K.; Carfagna, M. A.; Gao, H.; Dow, E. R.; Li, Q.; Searfoss, G. H.; Ryan, T. P. Temporal Gene Expression Analysis of Monolayer Cultured Rat Hepatocytes. *Chem. Res. Toxicol.* **2001**, *14*, 1218−1231.
(49) Moorjani, S. G.; Jia, L.; Jackson, T. N.; Hancock, W. O. Lithographically Patterned Channels Spatially Segregate Kinesin Motor Activity and Effectively Guide Microtubule Movements. *Nano Lett.* **2003**, *3*, 633−637.
(50) Bond, A. D.; Feeder, N.; Redman, J. E.; Teat, S. J.; Sanders, J. K. M. Molecular Conformation and Intermolecular Interactions in the Crystal Structures of Free-Base 5,15-Diarylporphyrins. *Cryst. Growth Des.* **2002**, *2*, 27−39.
(51) enCIFer from the Cambridge Crystallographic Data Centre. http://www.ccdc.cam.ac.uk/prods/encifer/.
(52) Cambridge Crystallographic Data Centre. http://www.ccdc.cam.ac.uk/.
(53) Hegelund, F.; Bürger, H.; Pawelke, G. The High-Resolution Infrared Spectrum of FNO_2 between 500 and 900 cm^{-1}: The $\nu5$, $\nu3$, $\nu6$, $\nu2$ Coriolis Interacting Tetrad. *J. Mol. Spectrosc.* **1997**, *184*, 350−361. http://dx.doi.org/10.1006/jmsp.1997.7337.
(54) The British Library, Gutenberg Bible. http://prodigi.bl.uk/gutenbg/background.asp.
(55) Herz, W.; White, D. H. Resin Acids. XXIV. Intramolecular Functionalizations of 11-Oxygenerated Abietanes and Podocarpanes. *J. Org. Chem.* **1974**, *39*, 1−11.
(56) Terrant, S. W.; Garson, L. R. Evaluation of a Dual Journal Concept. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 61−67. http://pubs.acs.org/cgi-bin/archive.cgi/jcisd8/1977/17/i02/pdf/ci60010a002.pdf.
(57) ACS Journal Archives. http://pubs.acs.org/archives/.
(58) Royal Society of Chemistry Retrodigitization. http://www.rsc.org/is/journals/retrodigitisation.htm.
(59) Ingoldsby, T. C., American Institute of Physics, Personal Communication, July 15, 2003.
(60) Davis, P. M.; Solla, L. R. An IP-Level Analysis of Usage Statistics for Electronic Journals in Chemistry: Making Inferences About User Behavior, in press. DOI: 10.1002/asi.10302.
(61) Scientific and Technical Information Exchange (STIX). http://www.stixfonts.org/.
(62) World Wide Web Consortium (W3C), MathML. http://www.w3.org/Math/.
(63) Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML, and the Worldwide Web. 1. Basic Principles. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 928−942.
(64) Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML and the World-Wide Web. 2. Information Objects and the CMLDOM. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1113−1123.
(65) Gkoutos, G. V.; Murray-Rust, P.; Rzepa, H. S.; Wright, M. Chemical Markup, XML, and the World-Wide Web. 3. Toward a Signed Semantic Chemical Web of Trust. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1124−1130.
(66) Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML, and the World Wide Web. 4. CML Schema. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 757−772.
(67) Gkoutos, G. V.; Murray-Rust, P.; Rzepa, H. S.; Viravaidyaa, C.; Wright, M. The Application of XML Languages for Integrating Molecular Resources. *Internet J. Chem.* **2001**, *4*, Article 12.
(68) Gkoutos, G. V.; Kenway, P. R.; Murray-Rust, P.; Rzepa, H. S.; Wright, M. A Resource for Transforming HTML and Molfile Documents to XML Compliant Form. *Internet J. Chem.* **2001**, *4*, Article 5.
(69) Joint Committee on Atomic and Molecular Physical Data (JCAMP). http://www.jcamp.org/.

AR0300017