Chemistry on the Internet—the Road to Everywhere and Nowhere[†]

Stephen R. Heller[‡]

U.S. Department of Agriculture, Agricultural Research Service, Building 005, Room 337, Beltsville, Maryland 20705-2350

Received July 26, 1995[⊗]

The ability to connect the information stored in computers around the world is presenting a challenge and an opportunity to chemists, information producers, and information providers. Are the chemistry resources on the Internet just an electronic Potemkin village or is there real substance to the multitude of computer resources now available? This paper will present examples and information which support both extremes, as well as providing one possible excellent use of the Internet.

INTRODUCTION

The use of computers in chemistry has been growing over the past few decades. The applications for using computers have gone from various types of complex and lengthy calculations to the ability to manipulate data and information to assist chemists in doing their work. Computers started out as large pieces of metal with lots of vacuum tubes and small sets of instructions (the computer program). Today they are small pieces of metal and plastic with large and growing sets of instructions. The personal computer (PC) has certainly changed the entire world in a rather short period of time. The reason for this is the low cost of computers and their high value in many areas of work and entertainment.

Today after moving from the industrial revolution to the first stage of the information revolution almost everyone wants to get the benefits of this revolution as quickly as possible. This has lead to enormous amounts of time, energy, and money being expended in the area of information and computer systems (hardware and software). Put in the context of chemistry, information and software are both like entropy—they are difficult to grasp, weigh nothing, and obey the Second Law of Thermodynamics; i.e., they always increase.[1] The recent almost continuous growth in many popular software programs (word processors, spreadsheets, etc.) has given rise to a new computer term—fatware.

For a number of years only the U.S. Government (primarily the Department of Defense and their ARPA (Advanced Research Projects Agency) project and the National Science Foundation[2]) had the foresight to realize the needs of a modern technologically oriented organization. The ARPANet, the forerunner of the Internet,[3] was designed to link different computer systems in different locations. Only in the past decade have scientists begun to use this electronic highway to connect themselves to the powerful computer resources they needed for their work. With the Government funding most features of this system (the people, the computers, and the network), and the system was designed for and used by universities and researchers, this was an ideal environment to work in. Like most examples of King Arthur's Camelot, it was destined to change.[4]

Due to technical, economic, and political changes in the world as well as the need for industry to move into new areas of businesses the relatively small and content scientific community has had a dose of reality injected into its system. What had been relatively quiet and easy times in academia, government research labs, and parts of industry is now changing, and they are being told that they must live by the same criteria as the rest of the world. This has happened and continues to happen due to the lack of money to support a Camelot environment. Large segments of the scientific community have been shaken to their foundations by being asked to become accountable and justify their existence as well as paying part of the expenses for resources they use. The totally free lunch appears to be over for the scientific community.

THE EVOLUTION OF INFORMATION RESOURCES AND THE INTERNET

There was the time we now call BC (before computers) when there appeared to be a finite amount of information one needed to learn in order to get an advanced degree, undertake research, and work in a scientific discipline. Compared to today's world, things took place at a slow pace, research was published in a timely, but not rapid, period. Today we have become overloaded with information and the volume and speed in which we get it.

As more powerful and larger computers were developed, it became possible to store large amounts of information and search this information to meet the particular need of a chemist. In the beginning the costs were low, as the bulk of the revenues from the database producers were from their printed products, and the electronic products were just "extra" revenue. While individuals may have created their own small databases of information there was either no desire or no practical way to deliver this to others.

Between 1970 and the early 1990s there have been a number of technological changes which have altered this traditional situation. Computers have become much more powerful, disk and other storage media have become rather inexpensive, and computer networks have been able to easily and inexpensively connect people, primarily scientists, around the world. Internet, which started in 1969 as ARPAnet, linking four research center computers, has grown

‡ E-mail address: SRHELLER@NALUSDA.GOV.

† This paper is dedicated to Professor Shin-ichi Sasaki on his retirement as President of the Toyohashi University of Technology, Toyohashi, Japan. A warm and generous man and a pioneer in the field of computers in chemistry.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1996.

in size to today where it links tens of thousands of local area networks (LANs) and wide area networks (WANs). In the middle of 1995 it is estimated that there are 6−7 million computers connected to the Internet. This vast and potentially powerful resource which has become available to most scientists is creating a euphoric atmosphere, and it seems everyone wants to get on the bandwagon. The Internet, which had been like a private club, is now on the front pages of newspapers and magazines. It seems like everyone wants to be "on the net", and, regrettably, just about everybody is on the net.

As the U.S. Government, mainly the NSF began to phase out its support of the Internet backbone in the United States, there was a great deal of concern that the "free" Internet would soon start costing a lot of money. What most of those who want the "good old days" of unlimited free access seem to forget that it did cost "real" money to put together and run the networks, computers, and data entry resources and projects. This is because, in most cases, the cost came out of overhead, not from individual budgets.

With the financial stresses occurring today in all areas (academia, government, and industry) supporting research at the level seen in the 1960s−1980s is no longer possible at this time. The reduced NSF support is just one of many such examples. Libraries offering unlimited no cost searches to customers constitute another perk that has virtually disappeared. Financial limitations imposed on libraries are also reducing the hours they are open and the number of books and journals they subscribe to. Thus the notion of "free" information via the Internet is particularly appealing at this time.

However none of this unbridled enthusiasm should really be a surprise if you think about the way in which any new activity is undertaken. First there is unlimited optimism, which turns into deep despair, and, if the activity does survive, then reality sets it.

TODAY'S INTERNET

The situation today is mostly in the first stage of development. All the technology is here. People are experimenting with what to do and how to do it. Because the entry cost is so low (as compared to the original set-up costs for commercial systems such as DIALOG, QUESTEL-ORBIT, STN, and so on) most experiments are done by people as a hobby. The first wave of chemists on the Internet were attracted by e-mail, bulletin boards, list servers, and perhaps a few databases which individuals put together as an exercise.

Industry, which has to have an income to offset expenses, was slow in getting started. Security was, and still is, a major issue for most organizations. Security is a very important subject and a critical issue for commercial success of much of the Internet, due to the abilities of computer hackers who have been able to penetrate many computer systems and cause a variety of problems. Believing in the free market system, I am sure that some day in the near future this will be fixed to the extent needed for commerce. As such it is not a matter to discuss in this paper.

For chemists there are a great number of Internet sites to explore. At the time this paper was written over 150 chemistry departments in the USA have put up web sites. In the rest of the world there are well over another 150 sites, with the United Kingdom and Germany having the most sites after the U.S.A. A number of chemical companies (Dupont, Monsanto, Eli Lilly, Shell, Texaco, Unilever, Rohm & Haas, General Electric, and others) have web sites, and their numbers are growing daily. Many computer hardware and software companies have web sites. At times it seems that counting the number of stars in the sky would be easier than counting the number of web sites.

Exploring or surfing the web is similar to the black hole which many chemists succumbed to two decades ago when computers first came to the labs. Chemists, most of whom had little training or aptitude, began writing their own computer programs, in some random computer language. Vast amounts of time and resources were expended in this area. The problem was, assuming the work was done correctly, that rarely could anyone else make use of the computer program or change it to met a slightly different requirement. Thus many people went out to reinvent the wheel.

While considerable time and efforts were spent developing computer programs, at least there was usually a positive result. Today on the Internet, with the hype and euphoria associated with this phenomenon, it often seems that much of the time and resources spent here are, to a very large measure, a major waste of organization resources. The problem, as noted by an article in Time magazine, is that "a good deal of the Internet's content is tasteless, foolish, uninteresting or just plain wrong. And it's habit forming".[5]

FREE INTERNET RESOURCES

Many people perceive the Internet as being a "free" resource, whose cost is paid for in some fashion which need not be the concern of the user. While today there are vast amounts of "free" chemical information on the Internet, there are two main issues which one must consider about such information. They are (i) what exactly is the nature of the information there and (ii) how long will it be there.

As mentioned above, there are a few hundred chemistry web sites available. The exact number is impossible to determine since sites come and go. Most sites have a number of common denominators. They have been put together as a side project or hobby, they have little information, what little information is there is useful to an even more limited number of people, and there are links to many other web sites with the same characteristics. These Internet sites are very much like Bingo: You never know what is next and if it is useful. Finding something on the Internet is like trying to find the bathroom in a house with 250 000 unmarked doors.

EXAMPLES OF FREE CHEMISTRY ON THE INTERNET

Good/Valuable:

Publishing

Membership lists of organizations and groups with telephone and FAX numbers and E-mail addresses

Organization Information - e.g., chemistry department faculty, research areas, publications, courses, degree requirements

Catalogs - from publishers, chemical and lab suppliers, hardware and software companies

Order Status - being able to track the progress of an order for chemicals or supplies using your internal purchase order number

List Servers and Bulletin Boards - e.g., Chemical Education, Computational Chemistry, Chemical Information

Advertising and Product Information - being able to quickly locate the latest product information, specifications, and costs

Demonstration software and databases - being able to see samples of a software program or database

Downloading of software updates - being able to update your software with recent bug-fixes or new versions

Databases - Material Safety Data Sheets (MSDS) Databases which allow one to obtain information on a chemical quickly and easily and Toxicology Databases which allow one to obtain toxicology data and information on chemicals and pesticides

Bad/Little Value

Advertising and Product Information - taking print advertisements and scanning them into a computer with no changes, with no ability to contact the vendor electronically

While it would be nice to have a list of Internet good and interesting web sites, the rapid pace of sites going up and coming down makes this a near impossible and impractical task. By the time this manuscript is published, much of the details of web sites will have changed. However with this caveat I would suggest that the following seven wide web sites are likely to be available when this article is published and are good jumping off points to other current chemistry sites. The first five are university web sites and the last two are commercial publisher web sites.[6] They are as follows:

a. http://hackberry.chem.niu.edu:70/webpage.html
b. http://www.rpi.edu/dept/chem/
c. http://www.osc.edu/chemistry.html
d. http://www.indiana.edu:80/~cheminfo/
e. http://www.chem.ucla.edu/chempointers.html
f. http://www.wiley.com
g. http://www.elsevier.nl

http = hypertext transfer protocol; www = world wide web

FEE BASED INTERNET RESOURCES

Today the majority of the data and information on the Internet costs money to access. CAS, DIALOG, QUESTEL-ORBIT, and the ACS (which includes the ACS Publications products, CAS, and STN), to name just a few information resources, can be accessed easily (and less expensively) via the high speed Internet network lines. However, the need for and cost to use these systems has not changed. The data and information they provide either from their own production or by distributing materials from other producers require time-consuming and are generally labor intensive to produce. These are not free and are never likely to become free.

FUTURE INTERNET RESOURCES IN CHEMISTRY

It is always both stimulating and worrisome to attempt to predict the future in a world where technology is changing so rapidly. However while there is rapid technological change, people evolve and change much more slowly. While some 90% of the population in the U.S.A. may have and use a Video Cassette Recorder (VCR) every week, probably 99% of these same machines probably do not show the correct time, but rather have 12:00 flashing all the time. I interpret this as meaning that people will use things only if they are simple to use and useful. Getting people to learn to set the clock on a VCR and getting people to change their habits on how they look for and obtain information require the same criteria. So with this caveat in mind I will discuss those activities which I believe are most likely to prove useful, easy to use, and viable in the future and those which I believe will not be a viable part of the Internet.

Good:

Databases - use Internet only as a less expensive network access e.g., CAS/STN, DIALOG, QUESTEL-ORBIT, etc.

Subscriptions to Data and Information - e.g., being able to get ACS supporting information and page images quickly

Bad:

None Yet

PUBLISHING AND THE INTERNET

Of all the examples of what the Internet can do for the chemical community, publishing is the one which will be discussed here in detail. Other areas and examples will follow in future papers. A major Internet issue for the chemical community[7,8] and the scientific community[9,10] is electronic publishing and its potential impact on the way research is reported and made available. For over 200 years, technical competence, reputation, promotion, tenure, and other indications of advancement have been largely based on printed publications in well recognized peer-reviewed journals.[11] As I have noted before,[12] journal articles have been and currently continue to be the most socially acceptable form of communication and reward/promotion. Today, in the middle of the 1990s, the principal electronic involvement of most scientific manuscripts is that they are submitted in electronic form, using one of a number of word processing programs. Essentially all refereeing of scientific manuscripts is accomplished by postal system mail, with some refereeing done by FAX or electronic mail (e-mail).[13]

With the dwindling number of subscriptions to printed scientific journals (where publishers had historically made the bulk of their profits) publishers have good reason to be concerned about the future of printed information. Many publishers of scientific and technical materials are experimenting with ways to enter the electronic information age. At the present time there are many experiments in electronic publishing. A number of journals are now either partially[14] or, in the case of the preprint publishing forum of Ginsparg, completely electronic.[15] One thing to note is that these experiments are just extensions of the way publishers do business today, not new methods and institutionalized ways of publishing and dissemination of information.

One experiment in the area of electronic publishing in which I am involved is an Internet column in *Trends in Analytical Chemistry* (TrAC).[16] Among the innovative procedures used in this activity is that all manuscripts are being submitted electronically. Manuscripts (including figures) are transmitted to the column editor by e-mail or are downloaded by the editor from an Internet web site of the author. Referees are then selected, and an electronic copy of the manuscript is sent to the referee. The resulting comments are then sent back to the author by e-mail. The corrected manuscript is then sent on to the publisher in Amsterdam. This process has helped to speed up publication by a few months at the least. In the first year of this column it has been possible to get a manuscript in press about 3−4 months after the manuscript was first submitted. This time frame appears to be a reasonable goal for other journals to go in the future, but only after enough referees are able to perform this electronic processing of manuscripts. Of course the Internet column manuscripts in TrAC come from experienced computer and Internet users, so the process is viewed as reasonable. Overall, as the use of computers by chemists is low, this new method is likely to take time before this process is considered normal, but there is little doubt about it being the final outcome.

Where is all of this likely to lead? As I once mentioned in a lecture: The paperless office is as practical as the paperless bathroom. Stated another way, it is unlikely that printed copies of books and journals are going to disappear for a long time. (But that does not mean some things in this field will not change substantially.) The computer, with its vast speed, search, and retrieval abilities, is, without a doubt, a very helpful tool. Computers can do many things more quickly, efficiently, and cheaply than people can. Exhaustive database searching is trivial for a correctly programmed computer (e.g., world wide web search engines[17]) but could be a lifelong task for a person. While one can find things quickly in a computerized database, I contend one cannot or will not sit in front of a computer screen and read page after page of scientific articles. One cannot easily bring the computer screen on the bus, train, or sit at a cafe table sipping tea and reading, nor would most people want to.

There are four components in an information system. The first are the scientists who produce the work to be published. Second, there are the publishers who publish and disseminate the scholarly works in a journal or similar publication. Third, there are the abstracting services which take published materials in many languages from all over the world. Lastly, going full circle, there are the scientists who read the journals and abstracts. One point, which is a crucial issue in electronic publishing is that computer systems allow one to do this very differently from that of the world of print. This means one must examine the system from start to finish, rather than talk about how a particular step can be changed or automated. There need to be new processes and procedures, not just modifying existing procedures. "The single biggest error one can make in the area of strategic thinking is to imagine that the future MUST be a logical evolution from the past".[18] What I foresee for the future will be a totally different way of life for publishing in science. Just as the airplane has brought travel throughout the world to a level never dreamed of before the middle of the 20th century, publishing on the Internet will be a radical change from the world of the Gutenberg press.

TODAY'S ECONOMIC ENVIRONMENT

Before going into each of the components, a brief reality check is in order. For the past decade, and with increasing speed in the 1990s, financial pressures have been brought to bear on all organizations. In this discussion the focus is on the organizations associated with publishing scientific manuscripts. The bulk of manuscripts come from universities, government labs, research organizations, and, to a much lesser extent, industry where advancement is measured primarily in terms of journal publications (and increasingly for universities and government labs in the past decade, for patents). Each is undergoing reorganizing, downsizing for the same end result—to get costs down and to keep them down to those below the levels of income (be it from university tuition or sales of products). At the same time organizations are looking for ways to increase their incomes and keep their incomes up. Universities have set up foundations and corporations to handle the business aspects of their faculties' potential commercial efforts. The U.S. Government has established cooperative agreements in many forms to work with industry and, at least for the government scientist, share in the potential financial success of the project (via such mechanisms as sharing royalties on patents). Universities are taking substantial funding from commercial organizations (who find it cheaper than doing in-house research), as government research grants become harder to obtain. Today all universities have "efforts underway to develop new sources of revenue".[19] This crisis is considered a long-term problem by universities who are "facing bigger student financial-aid needs and shrinking government support".[20]

With the above reality check on the finances of primarily universities and colleges, consider this scenario for the publishing in the next century. What would happen if universities and the other organizations which create the manuscripts decide they want to receive compensation from publishing. (After all, authors have always received royalties for books and even for just one chapter in a book.) What would happen if a university would require all faculty and staff (all departments) to publish their research in the university or company electronic press (and print press)? In the U.S. Government, for example, the National Library of Medicine (NLM) could become the journal publisher for all of the National Institutes of Health (NIH) research papers. Similarly the National Agriculture Library (NAL) could become the publisher for all of the Agricultural Research Service (ARS) and other U.S. Government agriculture research work. This is similar to their counterparts in industry who have had to turn over their discoveries to their company for patenting. (In those cases where there is joint publishing between organizations, an agreement would have to be reached in advance as to where the manuscripts would be published and how the income would be divided.) Such electronic journals and software programs could become part of the academic, government, and industrial chemist's reward/promotion and tenure system.

Other cost savings from electronic publishing would be the virtual elimination of printing, mailing, marketing and advertising of journals. Today the publishers promote journals by expensive printing, mailing, and advertising. In the electronic journal system of the future, all information about journals and their subject matter would be electronically searchable (using the Lycos engine or a similar software search machine tool[17]), reducing the costs for scholarly publications.

Some academics have said that their "academic freedom" would be trampled upon by forcing them to publish in a given journal. I would expect that a university which offered to share some of the revenues from the publishing activities is likely to win over those faculty members with such concerns, especially those faculty who need funds for research toward their tenure and/or ongoing research activities. For a university to be able to offer faculty the opportunity to publish faster and to establish pay and promotions on the quantity and quality (by measuring the number of times a paper was accessed electronically) is an attractive carrot to be used to move toward this system of the future.

It should also be remembered that until very recently in this century, there were very few scholarly journals. Much of the explosion in the number of journals comes from a combination of new fields of scholarly works, the pressure to "publish or perish", and the egos of scholars who need the recognition of editing a journal. Today there are some 50 000 peer-reviewed scholarly publications throughout the world,[21] with the largest commercial publisher, Elsevier, having revenues in excess of $500 million. In total, the market value of just the commercial publishing companies in the field is some $4 billion. Based on these figures, it has to be very tempting, if not imperative, for universities and other organizations to tap into this gold mine of revenue.

There is also the important financial point that grants are becoming much harder to come by, and any possible revenue source for a researcher is likely to be attractive. Many researchers have been turning to industry for funding as other sources have dried up in the past decade. Many are concerned that such money from industry is corrupting for basic research. Such sources of funding also force a researcher to think hard before pursuing an interesting, but nonpractical research direction, for concern about getting future funding. Internal university funding may become one of the few real sources for basic research in the 21st century.

Adding to the financial issues which electronic journals of the future are likely to help solve is the matter of scientists and scientific journals in less developed countries. At present most of these countries have little funds to pay for the costs of journals and abstracting service products. The August 1995 issue of *Scientific American* contains a telling article about how "many researchers in the developing world feel trapped in a vicious circle of neglect—some say—prejudice by publishing barriers they claim doom good science to oblivion".[22] Balaban has also commented on the financial plight of such countries, focusing on his country, Roumania.[23] He notes that "this then is the paradox: high-quality useful information is being produced, but only a tiny fraction of the potential users can afford to buy it and benefit from it; therefore a larger fraction is compelled to use out-dated information systems, being left out of the information superhighway. Information of high quality exists, and people who wish to use it are there, but one must find a means of mass production and distribution which should be able to (i) reward adequately the producers of information; (ii) to mass-produce and mass-distribute this information as affordable prices for the users, such that (iii) the sum of the many low-margin profits should be able really to lead to the information superhighway". It would seem that the proposed publishing model of the future offered in this paper would address these concerns of scientists from developing countries.

Another factor which needs to concern those who publishing is how a scholarly manuscript will be preserved for future generations. Perhaps even more critical to a researcher than funding is how to guarantee that published results will be readily available for future generations of researchers. Today the professional societies and commercial publishers print the journals. These journals are purchased and stored in libraries. In the paradigm proposed here, I would foresee eliminating the middle-man (i.e., the publisher) and have the library take over the function of the providing the long-term stability and preservation of electronic scholarly works. In fact, I believe that until the libraries will overtly take on this role, the various experiments by in electronic publishing[24] will present majors problems for acceptance on the part of researchers, if a "free" Internet web site journal or pre-print system disappears as a person changes jobs, retires, or the organization no longer agrees to support this electronic activity. Without institutionalization, which the publishers now provide and which the libraries will provide in the future, the bulk of the scientific community will not move away from the traditional printed publishing process (even with slight modifications in the direction of electronic publishing).

COMPONENT ONE: PEER REVIEW, QUALITY CONTROL, COPYRIGHT, AND LANGUAGE

Peer review should not pose any problems for such a system. Today journal editors choose the people who review and referee papers. Most individuals do not review manuscripts for just one journal. Peer review could go essentially as it does today, using peers in the field from other universities or organizations. There is no reason to believe, for example, today or in the future, that a reviewer from one prestigious university in such a system would give a bad review to someone in either a less prestigious organizations or in a competing organization or journal. One advantage of this electronic peer review system is that it is possible (and probable too) to add another step in the beginning of the process by "publishing" initial or "preprint" versions of a paper for a wide audience to examine and comment on (such as is being done by Ginsparg[15] and electronic chemistry conferences such as the Electronic Conference on Trends in Organic Chemistry (ECTOC)[25]).

There are at least two additional side benefits to electronic publishing. The first is that, as noted below, more data can be presented in a manuscript (since space in no longer an issue). The ability to present more data would also have the side effect of reducing error and fraud, as the data would be there for others to examine and analyze. The second benefit is that more innovative research is likely to be published, again due to relative lack of space problems that exist now in printed journals.

In the genome biology community, genetic sequences are being "prepublished" in computer database systems, such as GenBank,[26] and when corrections are made to the sequence these are incorporated into the database for all to see and review. GenBank also serves as a way to "publish" all genome sequences, as over 3 billion base pairs (or any large subset) would be impossible to publish in print.

Quality is closely related to peer review, or at least it should be. Peer review is the way to filter out incomplete research, trivial findings, and other things that do not deserve publication in the scientific literature. In general the more highly prestigious journals and magazines, concerned with their reputations, do a fairly good job of publishing good work. However there are many known instances of fraud, or incomplete experimental procedures, so that some published scholarly works are not reliable. In general, from a technical point of view, such problems are resolved by redoing experiments. (Of course reviewers generally cannot do this, so they rely on what is in the manuscript.) Over the past 30 years I have been publishing papers; I and many of my colleagues[27] have estimated that perhaps half of the scientific literature (in chemistry) is questionable. One recent publication[28] which I contributed to showed the lack of reliable solubility data for widely used commercially available pesticides. I would assert that the quality of the scientific literature will improve in the electronic age much faster than it has in the purely print age, due to prepublication versions of manuscripts, which allow for examination and improvement before a manuscript is frozen and finalized.

In summary, peer review is more likely to be improved than compromised in the future. For those "vanity press" journals that are likely to continue or the new ones which are likely to spring up considering the low start-up cost of an electronic journal, it is a price to pay for a free society and should have no real bearing on mainstream science.

After peer review and quality, the issue of copyright is a crucial issue. Copyright has a long and distinguished history in the U.S.A.[29] and elsewhere. At present when authors decide to publish a manuscript they are required to turn over their copyright to the publisher AT NO COST. Only book authors or (often) chapter authors in books receive any royalties from their works. The future depicted here would allow for the authors (or their respective employers) to receive compensation for their scholarly works. Actually with the entire system of the future being electronic, keeping automatic tracking of royalties and payments of royalties should be greatly simplified.

Language is an important issue which must be dealt with in any electronic publishing system of the future. There are two main choices here. Either a manuscript is published in English, which has become the universal language of science, or it is in another language. I would expect that scientists could still publish in their native languages but that such publications would be (when technology allows) translated manually into English (as is done by the German journal *Angewante Chemie*) or eventually by automated means. With 80% or more of the high-quality journals being in English at this time, it is hard to believe there is a first-rate scientist who cannot read and write English.

COMPONENT TWO: THE PUBLISHERS AND THE JOURNALS OF THE FUTURE

The second component of the current system is represented by the publishers. The commercial publishers have been a critical and vital part of the system during the past decades (and, for some professional societies, even much longer). Publishers performed an important service of collecting and distributing scholarly works. (Abstracting services have performed a useful service in the past, and their future will be discussed later.) However, times and needs change. In the U.S.A. (and elsewhere in the world) there used to be many newspapers published every day. There were morning newspapers, afternoon papers, and evening editions. In 1920 there were 14 daily newspapers publishing in New York City. By 1960 there were seven New York newspapers. In mid-1995 there are three which are still publishing.[30] What happened is that news did not go away, but rather the function of disseminating the news was taken over, first by radio and then by television (with many of the same companies owing radio and televisions stations). Will the publishers of scientific information disappear or evolve?

With electronic manuscripts, there likely will no longer be a need to have hundreds of different journals in libraries around the world but perhaps just a few very large (and using economy of scale less expensive to create, print, and distribute) printed repositories. One could create a computer-based organization where papers are indexed, stored, and retrieved in their appropriate subject discipline or cross disciplines if appropriate. Properly linked together electronically over the Internet, all the intellectual information would become part of an organization's resources as well as its own information being an asset to an organization.

While some universities, government, nonprofit, and industry organizations might be too small to undertake such ventures, one could see consortiums being established to have smaller organizations work together. Some organizations might even contract, for a fee, with a publisher to handle the process. (Already in this decade some universities are licensing their sport team logos for millions of dollars a year[31]).

Without a doubt such a structure would have to be staffed, in a manner similar to the way publishing companies are currently staffed. Thus there is no absolute need to eliminate publishing companies.[32] As for the ownership and copyright of the materials, as stated above, that would reside with the university or organization which generated the intellectual property. One should recall that the growth of private publishing companies is a relatively new phenomena, with the real growth coming only in the past 50 years. There are no laws that require scientists to use such publishers of technical and scientific works as we know them today. Hence they may very well go the way of the slide rule if they do not learn to adapt. As scientists and librarians have seen in the past decade, most publishers seem capable of primarily raising prices and not planning for a different future, so there appears to be opportunity here for entrepreneurs.

As to the matter of finances, it would seem reasonable to expect that a good university (with first rate faculty and first rate research) could charge more for access to their store of knowledge than a less prominent organization. (Remember, not all published manuscripts have the same copyright fee.) Fixed fees or subscriptions would cover a good deal of the cost. For those who would actually want to see and read articles, there would be fees for the downloading/printing of articles. (In this way there would be hard-copy, but it would be created only on-demand. This would save paper and be environmentally correct.) Computer system accounting programs would be developed to electronically "count" all the uses and accesses of information and by whom and handle the overall bookkeeping. The existing repositories of computer readable abstracts and journals could be connected to the system so that all the historical knowledge, which the current societies and private publishers own, would neither be lost nor unavailable.

One interesting measure of quality and accountability which could come out of such an electronic system would

```
(c) 1995 American Chemical Society. All rts. reserv.

  73009229    CA: 73(2)9229w    JOURNAL
  NMR spectra of sulfur compounds
  AUTHOR(S): Heller, Stephen R.
  LOCATION: Div. of Comput. Res. and Technol., Nat. Inst. of
Health,
Bethesda, Md.
  JOURNAL: Intra-Sci. Chem. Rep.  DATE: 1968  VOLUME: 2  NUMBER:
4  PAGES:
321-440  CODEN: ISCRB6  LANGUAGE: English
  SECTION:
CA873000 Spectra and Other Optical Properties
  IDENTIFIERS: NMR sulfur compds bibliography
  DESCRIPTORS:
Sulfur...
    nuclear magnetic resonance of
Nuclear magnetic resonance...
    of silver compds.
```

Figure 1. CA Abstract of a NMR article on sulfur compounds.

# Items	Index term
148	CONFIRMATION
1	CONFIRMMATION
1	CONFIRUGATION
1	CONFIRURATIONS
1	CONFORMATIION
1	CONFORMATIN
6635	CONFORMATION
3	CONFORMATON
1	CONFORMTIONS
2	CONFROMATION
1	CONFROMATIONAL

Figure 2. Spelling of conformation/confirmation in the Chemical Abstracts database from 1967−1971 (from Dialog searches on File309).

# Items	Index term
375	CONFIRMATION
1	CONFIRUGATION
1	CONFIRUGATIONS
1	CONFIRURATION
1	CONFIRURATIONS
1	CONFOMRATION
1	CONFORATIONAL
5	CONFORMATIION
2	CONFORAMATION
17595	CONFORMATION
5	CONFORMTION
8	CONFROMATION
1	CONFRONTATION

Figure 3. Spelling of conformation/confirmation in the Chemical Abstracts database from 1977−1981 (from Dialog searches on File310).

be the number of times a manuscript was accessed and the number of times it was downloaded and printed. This could be one possible measure of the value of a manuscript—to be used for reward and promotion. This would be similar to, but different from, the Institute for Scientific Information (ISI) citation index.

COMPONENT THREE: THE ABSTRACTING SERVICES

With the vast amount of scientific literature to read, it has become virtually impossible for an individual to physically keep abreast with the literature by just scanning and reading journals. Multidisciplinary areas have increased the number of publications relevant to much of today's research. As the scientific literature grew rapidly after World War II, abstracting services became more important and useful to the end user—generally the working scientist. Abstracting services provide a mechanism for scanning a large amount of the literature and disseminating this information in a well organized fashion.

Most abstracting services assert they provide a certain amount of quality control.[33] One of my first publications, on the sulfur NMR, shown in Figure 1, was abstracted as "silver NMR",[34] which led me to be skeptical at an early age.

When abstracting services became totally electronic it would seem that quality should increase. A number of years ago I gave a talk at an ACS meeting in Toronto[35] on quality control and used the lack of spell checking of abstracts and

# Items	Index term
494	CONFIRMATION
1	CONFORMAION
1	CONFORMAITONN
2	CONFORMATIN
2	CONFORMATINOAL
1	CONFORMATIOAL
31865	CONFORMATION
1	CONFIRMATON
3	CONFORMTION
1	CONFORRMATIONAL
1	CONFRMATION
3	CONFROMATION

Figure 4. Spelling of conformation/confirmation in the Chemical Abstracts database from 1987−1991 (from Dialog searches on File312).

index terms as examples of areas in which abstracting services should be able to improve their quality. As noted in Figures 2−4, there has been some improvement on checking for the spelling of "conformation" and "confirmation", from 1967 to 1991, but there is still a long way to go.[36]

These services also provide the same function as publishers of collecting and disseminating information, which in this instance are primarily abstracts, not scholarly works. Abstracting and indexing is, at present, a very labor intensive process (i.e., costly and getting more so). Recently some publishers have started to charge abstracting services for the abstracts used in abstracting publications. What is happening here is the classic financial squeeze. Profits at publishing houses are going down due to increased costs and fewer customers due to higher prices. Who is left to squeeze more money from? The abstracting services. This plan, if it works, is clearly the calm before the storm. One day the publishers (singly or jointly) may decide to take all the author abstracts and create their own abstracting service.[37] This is becoming easier as time goes by, since more and more manuscripts are computer readable almost from the beginning of the publishing process. Certainly users would be happier to get an abstract months sooner than they currently do. Publishers would have additional revenue without much additional cost, as they already have the abstracts as part of their journal publications. The abstracting services are in financial trouble. After years of telling the publishers they are doing them a favor by "advertising their publications" and telling the users that the information their scholarly fields are too big and spreadout, the National Federation of Abstracting and Indexing Services (NFAIS) has recently proposed the various parties work together.[38,39] One must give NFAIS credit for at least realizing they are likely on the deck of the Titanic. Besides rearranging the deck chairs, it is not clear they wish to do otherwise than find some way to continue their existence when that may not be useful, necessary, cost-effective, or possible. As Peter Hyams has previously noted "It is hard to see how they (scholars) will require third-party 'publishing' organizations to assist the academics 'brainwork'".[40]

While the need for most abstracting services are likely to diminish in the coming years, there are some secondary publishers which may survive the changes suggested above. Factual data is very different and more complex than abstracting and indexing keywords. It is also used much less frequently and requires higher access costs than the complimentary abstracting services. One such service is the Beilstein database. In this instance the data extraction and quality control performed by organic chemists cannot, either at present or for the foreseeable future, be replaced by an automated system. One hopes the Beilstein Institute will be able to continue to afford to keep its high standards for data evaluation as the costs for such work increases.

COMPONENT FOUR: THE CUSTOMER

We now come full circle back to the people who started the process—the scientists who perform the research. They create the work, but neither they nor their employers are compensated for their efforts. In fact, it is just the opposite. Scientists need to disseminate their information, to get recognition for their good works, and they need to keep abreast of what is going on throughout the world in their field. In the not so recent past this was easy, and the pace of life a lot less hectic (at least for those who did not grow up in New York City). Today the pace of things are often very hectic. This has caused some scientists into thinking how their work could be made easier and more effective, rather than just working extra hours every day. (Given there are only 24 h in a day, this solution is clearly has its limitations.) The result of this thinking has been a number of experiments[14−16] which show the feasibility of the scenario discussed above.

In every organization there is an administrative or support staff to carry out many basic and important functions in the organization. In the past, the technology of the world required scientists to publish their papers in the most highly recognized scholarly publications, read papers in journals (mostly) in the library or in the lab, and go to the library for access to the abstracting publications. What exists today is a system which has served its community quite well. The problem with the current system is that technology has provided the tools to make drastic changes in the manner in which scholarly works are reviewed and disseminated. Once upon a time, BC (before computers), manuscripts were typed by secretaries and figures were drawn by art departments. Now computers do this and these support staff are either gone or, in the case of secretaries, doing other higher level work. Now that manuscripts are computer readable, it would not take very much of an effort to begin to build a system which eliminates unnecessary pieces of the system and replaces them, in some cases, with the new technology and new way of doing things.

When all is said and done, of all the parties which are critical to the system, the scientist is the first. The second are the organizations whose function is to obtain and disseminate the results; these functions will not disappear in the future but probably will have their work taken over by other organizations. Among the support organizations which can play a role in this paradigm are libraries. In the above publishing scenario an organization will need an office to collect manuscripts, track the review process, and make the information available. This is certainly a new role for a library, but with the changes in technology, it may be one of the more valuable roles libraries can productively play in the future.

SUMMARY

The Internet is a potentially powerful technological tool which is likely to have an impact of chemists and other scientists throughout the world. One major change that is likely to result is a massive restructuring in the way in which scholarly works are created, reviewed, and disseminated. Many, and probably most, publishers and abstracting services are likely to be replaced by new technology. This technology will perform their current functions but in a faster and cheaper manner and provide more useful functions that the current systems these organizations provide today.

ACKNOWLEDGMENT

I would to particularly thank the following colleagues for their comments and input into this manuscript: Steve Bachrach, Stu Borman, Harry Collier, Ray Dessy, Bill Milne, Henry Rzepa, Peter Shepherd, Wendy Warr, Todd Wipke, and Chezi Wolman.

REFERENCES AND NOTES

(1) Augustine's Law XVI, Norman R. Augustine, "Augustines Laws", ISBN No. 0-14-009446-6, Penguin Books: 1987.
(2) Kern, W. C. Computer Applications for Scientific Applications: A Status Report. Chapter 9, In *Computer Applications in Chemistry*; Heller, S. R., Potenzone, R., Jr., Eds.; Proceedings of the 6th ICCCRE, ISBN No. 0-444-42210-2, Elsevier: 1983; pp 103−110.
(3) For this paper I will use the phrase "the Internet" as meaning the computer networks and sources of information residing on computers connected to these networks.
(4) "The only thing new in the world is the history you don't know". Harry Truman.
(5) *Time Magazine*, Special Spring 1995 Issue; p 8.
(6) (a) Dr. Steven Bachrach (smb@smb.chem.niu.edu) - Northern Illinois University. (b) Dr. Joseph Warden(wardej@rpi.edu) - Rensselaer Polytechnic Institute. (c) Dr. Jan Labanowski (jkl@osu.edu) - Ohio State University. (d) Dr. Gary Wiggins (wiggins@indiana.edu), Indiana University. (e) Dr. Max Kopelvich (mik@chem.ucla.edu), University of California - Los Angeles. (f) John Wiley & Sons. (g) Elsevier Science.
(7) Rzepa, H. (15 Jan 1994), Stumpe, R. (1 Feb 1994), Pierce, T. T. (2 Feb 1994), E-mail messages on the Computational Chemistry listserver (http://www.osc.edu/ccl/cca/html).
(8) Wiggins, G. (8 June 1995), E-mail message on the Chemical Information listserver (http://www.indiana.edu:80/∼cheminfo/list-serv.htm).
(9) Hymans, P. *Monitor* June 1995, *172*, 2−3.
(10) Winograd, S.; Zare, R. N. "Wired" Science or Whither the Printed Page? Editorial, *Science* **1995**, *269*, 615.
(11) "With credit from one's peers comes access to all important grant funding, easier publication in leading journals, and a steady supply of grad students and postdocs who make the lab run". Cohen, J. The Culture of Credit. *Science* June 23, **1995**, *268*, 1706−1711.
(12) Heller, S. R. The Future of Chemical Information Activities. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 284−291.
(13) In discussions with a number of journal editors I would estimate that less than 1% of the refereeing is handled electronically at this time.
(14) (a) *Current Clinical* trials is considered by most scientists to be the first real electronic journal. See: *Science* **1991**, *253*, 1480. (b) Applied Physics Letters - Online. This is a complete and electronically searchable version of the printed publication - Applied Physics Letters. The American Chemical Society have their supporting journal information available in electronic form (http://pubs.acs.org/supmat/supmat.html). (c) *J. Molecular Modeling*. For further details point your web browser to: http://www.organik.uni-erlangen.de/indo/JMOLMOD/jmolinfo.html. (d) Borman, S. see: http://pubs.acs.org/hotartcl/index.html and look at articles from the March 23, 1995 issue of *C&E News*, pp 42−49. (e) For the electronic version of the *J. Biol. Chem*. point your web browser to http://www-jbc.stanford.edu/jbc. (f) Details of the Elsevier Tulip project can be found at http://www.elsevier.nl/info/projects/tulip.htm. (g) Details of the Springer-Verlag/John Wiley & Sons Red Sage project can be found at http://library.ucsf.edu/lib/gen/redsage.html.
(15) E-print was initiated by Ginsparg, P. Los Alamos National Labs, in 1991. To view it, point your web browser at http://xxx.lanl.gov. Also, see: Taubes, G. Networks - Publication by Electronic Mail takes Physics by Storm. *Science* **1993**, *259*, (Feb. 26), 1246−1248.
(16) *Trends in Analytical Chemistry* (TrAC); Elsevier: Amsterdam. To read those TrAC articles published to date, point your browser to http://www.elsevier.nl:80/section/chemical/trac/menu.html.
(17) There are a number of "search engines" on the Internet. Among the more widely used are the following: (a) The Lycos search engine at Carnegie Mellon University: http://lycos.cs.cmu.edu. (b) InfoSeek Search from InfoSeek Corporation: http://www.inforseek.com. (c) W3 Search Engines from the University of Geneva: http://cuiwww.unige.ch/meta-index.html. (d) CUSI (Configurable Unified Search Interface) from Nexor, UK: http://pubweb.nexor.co.uk/public/cusi/cusi.ht-

ml. (e) WebCrawler from the University of Washington: http://webcrawler.com/.
(18) Collier, H. (Infornortics Ltd.), e-mail message, 23 June 1995.
(19) El-Khawas, E. American Council on Education, Research Briefs, 5, #8, p 1 (1994).
(20) Arenson, K. W. *Universities Taking on Risks To Overcome Financial Squeeze*. *New York Times*, **1995**, July 24, p 1.
(21) *Business Week*, **1995**, June 26, p 44.
(22) Gibbs, W. W. Lost Science in the Third World. *Sci. Am.* **1995**, *273*, 92−99.
(23) Balaban, A. T. Mozart, Ford, and the Information Superhighway Paradox. *Am. Sci.* Submitted for publication.
(24) Rzepa, H. A Personal View of the Future of Electronic Journals in Chemistry. *Trends Anal. Chem.* In press. The article is also available in electronic form at http://www.elsevier.nl:80/section/chemical/trac/menu.html.
(25) Rzepa, H. Electronic Chemistry Conferences. *Trends Anal. Chem*. **1995**, *14*, 240. The article is also available in electronic form at http://www.elsevier.nl/locate/trac.
(26) Benson, D.; Lipman, D. J.; Ostell, J. GenBank. *Nucleic Acids Res.* **1993**, *21*, 2963−2965. To access this database via the Internet point your web browser to http://www.ncbi.nlm.nih.gov/.
(27) For example: (a) Dessy, R. private communication, e-mail message, 23 June 1995 indicated a serious lack of trust in scientific ethics at large institutions and of man scientific investigators. (b) Roy, R. "the scientific literature is littered with erroneous reports, most of which are not corrected, and only some of which have been retracted", *C&E News*, **1995**, May 22, p 26.
(28) Heller, S. R.; Scott, K.; Bigwood, D. W. The Need for Data Evaluation of Physical and Chemical Properties of Pesticides. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 159−162.
(29) Section 8.8 of the U.S. Constitution states "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writing and discoveries".
(30) *New York Times* **1995**, Saturday, July 15, p 26.
(31) *Business Week* **1995**, May 22, p 6.
(32) In ref 14d Borman writes that "electronic publishing can be so inexpensive that some scientists and librarians are already beginning to view traditional publishers as obsolete middlemen.
(33) CAS staff, private communication.
(34) *Chem. Abstr.* 1970, Vol. 73, Abstract 9229w, p 412. In the 8th Collective CA Subject Index 91967-1971. Page 21195S refers to this article as the NMR "of silver compds., 73: R 9229w.
(35) Heller, S. R. Better Living Through Computerization. Presented at the Computer Secretariat Symposium on How the Computer has Affected the Practice of Chemical Information, 3rd Chemical Congress of North America and *Abstracts of Papers*, 195th American Chemical Society National Meeting, Toronto, Canada, June 9, 1988; American Chemical Society: Washington, DC, 1988; Vol. 195, p 35.
(36) The new ACS product SciFinder does not yet have this capability, but it is hoped to incorporate it in the near future. Macko, J. L. CAS, private communication.
(37) Elsevier is also starting to do this on a modest scale.
(38) Kaiser, D. Federation Calls for Scholarly Communication Summit. National Federation of Abstracting & Information Services (NFAIS) Press Release, May 2, 1995. NFAIS, 1518 Walnut Street, Suite 307, Philadelphia, PA 19102.
(39) "Primary/Secondary Publishers: A Complementary Relationship", May 1995. NFAIS, 1518 Walnut Street, Suite 307, Philadelphia, PA 19102.
(40) Hyams, P. *Monitor* **1995**, *172*, 2−3.

CI950081W