THE 21st CENTURY CHEMISTRY JOURNAL*

Steven M. Bachrach
Department of Chemistry and Biochemistry - Northern Illinois University - DeKalb, IL 60115

Recebido em 3/6/98

Internet publication will radically alter how chemists will publish their research in the next century. In this article, we describe two fundamental changes: enhanced chemical publication which allows chemists to publish materials that cannot be published on paper and end-user customization which allows readers to read articles prepared to meet their specifications. These concepts have been implemented within the *Internet Journal of Chemistry*, a new journal designed to employ the latest technologies for chemical publications.

Keywords: electronic publication; chemical markup language; www.

INTRODUCTION

As the 20th century winds down, it is natural to contemplate the changes that one has seen and experienced during the past 100 years. For a chemist, this reflection brings upon a feeling of wonder and amazement. It is almost impossible to imagine what chemistry was like at the beginning of this century when one ponders the remarkable advances made during this time period. Back in the year 1901, quantum mechanics was still a quarter century away; the neutron was unknown. Characterization of materials was done by color and boiling and melting points. Consider just these few examples of techniques and theories of 20th century chemistry:

- the ability to characterize materials in a non-destructive manner through spectroscopy;
- structural determination via x-ray crystallography and electron diffraction;
- separation technology using various chromatographic methods;
- mechanistic organic and inorganic chemistry;
- quantum chemistry such as molecular orbital and valence bond theories;
- large scale synthesis of complex organic, inorganic, and macroscopic compounds.

Chemistry at the approach of the new millennium is unquestionably as radically altered by advances made this past 100 years as in all of the previous history of mankind combined.

Now this is not at all out of the ordinary. When one ponders the changes made in other disciplines, they too have been remarkably reshaped during this past century. Pertinent to the subject at hand are the changes made in mass communication. The 20th century has seen the development of audio recordings on records, tape, and compact disc. Images can be captured as stills (high-resolution color photography) or as moving images (a continuing development of silent black-and-white movies to color high-fidelity video). These images and sounds are transmitted to the public through radio and television. High- speed color offset printing and photocopiers have altered the way in which books and articles are manufactured. Perhaps most important as we look to the future has been the development of the computer and their networks, significantly altering how we produce materials for distribution. Even the delivery system is clearly now in flux as the Internet becomes a primary medium. For example, electronic mail is supplanting the fax, which supplanted express mail delivery which supplanted the postal system as the mechanism for transmitting information to friends and colleagues.

What is most amazing is when you consider the cross section of these two disciplines. In other words, how have the changes in the communications arena affected chemists? The answer is: in terms of how chemists disseminate their primary research information, not at all! Chemists continue to communicate in a form that has not change in over 300 years. Chemists continue to communicate principally through the scholarly journal article, printed on paper, and distributed in bound collections as journals and books.

In this article, we will discuss how this situation is ripe for change. The advent of the Internet provides a new mechanism for researh information distribution. This mechanism is so compellingly better than the traditional print model that we will see a new publication paradigm in chemistry that will completely revolutionize how we as chemists will communicate in the not so distant future. We will describe the new opportunities for communication both in terms of what authors can publish and how readers will be able to manipulate the information. Lastly, we will briefly describe a new journal, the *Internet Journal of Chemistry*, which implements many of the ideas described here.

THE INTERNET AS A CHEMISTRY COMMUNICATIONS MEDIA

The global computer network known as the Internet establishes the mechanism for communication between disparate computer systems located anywhere on the planet[1]. The major utilization of this network has been, until very recently, the practice of sending and receiving electronic mail. Email is a very cheap medium; messages can be composed and read on virtually any type of system, the necessary software can be obtained for free, and the cost for sending or receiving a message is measured in the fractions of a penny. Scientists have long exploited this mechanism. Email lists have become a major mechanism for chemical communications. For example the Computational Chemistry List boast well over 3000 members and the Chemical Information List (CHEMINF-L) has nearly 2000 subscribers.

Email traffic has now been supplanted by the world-wide web traffic. The web is a intertwined connection of documents located on any computer on the Internet. While most documents are text-based, any file can be delivered via the web. This is

* Conferência proferida na 21ª Reunião Anual da SBQ, Poços de Caldas - MG, Maio/98

precisely the reason for the web supplanting paper-based print as the publication media for chemists. In the next section, we will describe how authors can take advantage of the ability to publish any type of data or file to significantly enhance the value of their publication. The section after that will describe how readers can modify the information content to suit their needs, a process impossible with paper-based publication.

Before we discuss these significant advances in chemistical publication, we will mention a few generic advantages to electronic publication. Instead of references which indicate the location of a related article, references in electronic articles can be hyperlinks which will take you directly to those articles. E-publishing offers generally more rapid publication since the printing step is completely eliminated. Also, since an "issue" with a set number of pages needs not be created, publication of an article is not held up while enough articles to create an issue are obtained. As an example, the American Chemical Society is able to publish articles on the Internet from 2 to 10 weeks prior to print publication[2]. Related to this aspect, the cost of publication on the net is less than on paper[3]. Since the publication and shipping stages can be eliminated, and this is a significant fraction of the publication budget, e-publications will eventually be cheaper than paper. Currently, publishers are attempting to cover the start-up costs associated with conversion over to a new medium by maintaining or increasing the costs of e-journals. This trend should reverse in the future. Lastly, due to the incredible spread of the Internet into all avenues of life, e-publications will be obtainable to a great percentage of scientists, conveniently available at home or office.

JOURNAL ENHANCEMENTS

We now turn our attention to the scientific justifications for conversion from print media, which has served us for so long, to electronic media. The principle goal of scientific publication is to disseminate knowledge to as wide an audience as possible. The publication speed, cost, and access issues discussed above argue favorably in this regard — these factors lead to a greater readership of science by lowering barriers to access of information. The web provides an additional benefit: the ability to publish information that cannot effectively (or economically) be delivered on paper or that cannot be put on a static page of paper at all. We will be able to communicate to our peers information and data that typically remains locked in our laboratory drawers and file cabinets because we cannot include them in our articles. "Chemical enhanced publication", as we call it, is the major driving force behind our efforts in the area of electronic journals[4-11].

Chemical Enhanced Publication

The best way to describe what is meant by "chemical enhanced publication"[8,10] is by a couple of examples. Many experiments produce three- and four-dimensional data, such as electron density maps and NMR spectra. A very convenient method for displaying multidimensional data is by using color to define one dimension. Publication of color images in paper journals is often prohibitively expensive and therefore usually avoided. However, it is no more expensive to publish on electronic media a color image versus a black-and-white image. No longer must these effective plots be excluded purely on the basis of cost.

Molecular dynamics simulations produce a trajectory, a map of the motion of molecules during a period of time. This data is best represented as an animation or movie. Movies cannot be published on paper; at best what can be done is to present a number of snapshots from different times. Publication of a movie file is again no more difficult or expensive than publication of a text file. (While movie files tend to be much larger that text files, the cost of disk drives and their increasing size, coupled with the inherent distribution of information across many sites on the Internet means that their size poses little barrier to publication on the Internet). Readers can download the movie file and play it over and over again, with slow motion or freeze frame capabilities for detail analysis.

For our last example, molecules are three-dimensional objects. When we publish a structure on paper, the author is forced to decide on a single static orientation to present to the audience. Recognize that a reader may wish to view the back of the molecule, or zoom in on a particular region (say the active site of an enzyme), and paper publication precludes any changes or interactivity with the structure. Using the Internet, an author can actually publish the structure itself, not an image of the molecule. The structure can be downloaded into a reader's molecular visualization program. Now the reader can interactively rotate the molecule to view it from any orientation, zoom in or out, measure distance and angles. The information content provided in this manner is significantly greater that the dead image found on a sheet of paper.

This last case serves just as a single example of the real power of enhancement publication — the ability to allow the reader to manipulate the data. This can be done with, for example, spectra, whereby readers can display a full spectrum, scale it, peak select. Instead of a table of numbers or a single plot, readers could obtain the full data set within a spreadsheet and perform their own numerical analysis. Their are no limits to the interactive nature of the data authors could provide.

The key enabling technology here is that all files transferred across the web are labeled with a MIME type (Multipurpose Internet Mail Extension)[12]. An example of this implementation will disclose the inherent power. A user makes a request to a web server for a particular file, for example a file containing the molecular structure of the HIV protein. This structure has been stored in the Protein Data Bank (PDB) format. The web server answers this request by sending the file along with the MIME type, in this case *chemical/x-pdb*. The browser reads the MIME type, and then decides what to do with this file. At some prior time, the user had configured the browser to direct all pdb files to be displayed with the visualization program *Rasmol*[13]. The browser therefore launches *Rasmol* and directs the downloaded pdb to the program as input. The reader then sees a new window opening up (the *Rasmol* program) with the structure displayed and ready for manipulation. Recognize that the entire request was simply a mouse click on a hyperlink for the HIV protein, the rest was accomplished behind the scenes with no user intervention.

The concept of chemical MIME types was developed by Rzepa and coworkers[4,7,14]. Standards for a variety of different formats have been established and are in widespread use. A journal article in the future will likely contain some text and a number of different interactive components that allow the reader to become an active participant in the understanding (and perhaps even discovery phase) of the research presented therein.

End-User Customization

While many people have noted and implemented chemical enhanced presentations and the opportunities provided, this emphasis of the advantages of electronic publications have focused principally on the author. Certainly the reader is just as important a component in the publication process! We have recently presented a procedure by which the reader as well can benefit from electronic publication[15].

Traditional print publication provides the exact same immutable content to all readers. Some of the design criteria here are of stylistic value only, such as the font face, point size, page size, number of columns, etc. Some of the decisions are arbitrary yet bear on the usability for the reader. For example, a

journal decides what units are appropriate for publication, say kJ/mol for energy. A reader may prefer kcal/mol and must convert every energy measurement to obtain the units she desires.

We have proposed the concept of an end-user customized chemistry journal article. Each reader can individually determine how information should be delivered to their screen. For example, the reader may opt to have all energy measurements reported in units of kcal/mol, and that will be all she ever reads. Customization options can include layout choices, what interactive tools to invoke, sizes of images, automatic conversion of chemical formula into 2D structures, etc.

There are currently two mechanisms for providing end-user customization. Murray-Rust has proposed Chemical Markup Language[9,16] (CML), which is an Extensible Markup Language[17] (XML) specific to chemistry. In CML, chemical documents are tagged (marked-up) similar to how text is tagged to create HTML (Hypertext Markup Language) documents which are used on the web. Instead of indicating bold text and headings as is done in HTML, atoms, molecules, units, bonds and the like are indicated within a document. This CML document is then decoded and interpreted by a browser which understands CML, and which presents the chemical information to the user. The CML solution is very elegant and powerful, however, implementation has been slow (only a prototype CML browser is available[18]) and full development awaits standardization of XML by the major software manufacturers.

The alternative is to create a CML-like document which is not sent directly to the browser but rather is interpreted by the server itself. The server uses a user profile for deciding on how to customize the presentation to the specification of the user. The specifications are incorporated into the document as standard HTML commands and then sent on to the browser. We have implemented such a procedure in the *Internet Journal of Chemistry*. Regardless of how this concept is implemented in the future, either as CML with browser side interpretation, or server side implementation, or something else, future journals will allow the reader to customize the presentation of articles to meet their needs, utilizing a myriad of interactive tools to extract information that could never be included within the print tradition. At one moment in the future, ten readers will be viewing the same journal article and it will appear differently to each one, and each user will be able to extract content in a fashion best suited to their particular needs and styles.

PROTOTYPE: THE INTERNET JOURNAL OF CHEMISTRY

We launched a new chemistry journal in January 1998 to explore the opportunities for significantly improved chemistry publication[11]. The *Internet Journal of Chemistry* (IJC)[19] offers authors the opportunity to enhance their articles by fully incorporating multimedia, large data sets, Java applets, color images and interactive tools. Readers can customize the article presentation through a form interface.

The journal is fully committed to the electronic medium. All publication will be on electronic media only. Authors submit their articles in HTML with associated files. Referees are contacting by email, given a URL to access the article, and return their review by a web-based form. Authors are sent the referee reports by email and submit their revised articles electronically. Outside of the copyright licensing agreement, which must be signed on paper, there is no use of paper at all within the publication process.

We fully believe that all disciplines of chemistry can be better presented using electronic media than on paper. The journal is therefore open to articles in all areas. The journal is abstracted by *Chemical Abstracts* and can easily be cited within an article.

Instructions for authors are fairly unrestrictive, except that the articles must be in HTML. There are no special requirements upon authors to enable the end-user customization. This enabling technology is handled on the publication end through the use of a sophisticated article parser and a customized web server.

FUTURE OUTLOOK: THE JOURNAL AS DATABASE

We envisage a dramatically changed journal environment in the 21st century. Journals will appear on the Internet as electronic documents. Print will slowly disappear, lingering on as an archival resource only. The limitations of print as a means for communicating scientific content will become more and more obvious and severe.

Furthermore, the function of the journal itself will evolve. In the current print environment, journals contain the primary chemical literature. Access into the literature is provided through third-party databases, such as *Chemical Abstracts* and *Beilstein*. Once the journals appear as electronic documents, publishers will begin to provide powerful search engines. We are already seeing some of this concept on the Internet with the advent of Alta Vista and HotBot search engines. These are text only and therefore have limited utility for chemists. The new electronic journals will incorporate chemical search engines that can recognize chemical structures and substructures, spectroscopic content, protein sequences, etc. These search tools will transform the journal from simply literature to a full-fledged database.

Over the longer term, one can imagine that these search tools cross over journal boundaries so that the entire collection of chemical information can be accessed through common interfaces. Further extrapolation can take this chemistry information database and couple it with other disciplines to create a global knowledge library.

This global database still remains a dream for the future. Nevertheless, a fundamental change in chemistry literature is starting. All of the traditional publishers, both commercial and the societies, are producing electronic versions of their journals. A small number of electronic only journals, such as *IJC* and the *Journal of Molecular Modeling*[20] are pushing the technology into new areas and guiding us to a brave new world.

ACKNOWLEDGEMENTS

Are made to the Editorial Board of the *Internet Journal of Chemistry* for assistance and guidance in developing this project. This work was supported initially by the Camille and Henry Dreyfus Foundation and currently by InfoTrust, Ltd.

REFERENCES

1. For a general introduction to the internet and its use, see *The Internet: A Guide for Chemists*; Bachrach, S. M., Ed.; ACS Publications: Washington, 1996.
2. See the ACS Publications web pages, URL: http://acs.pubs.org/
3. For further discussions of the economics of electronic publishing, see Odlyzko, A. M. *First Monday* **1997**, *2*, URL: http://www.firstmonday.dk/issues/issue2_8/odlyzko/index.html.
4. Rzepa, H. S.; Whitaker, B. J.; Winter, M. J.; *J. Chem. Soc., Chem. Commun.* **1994**, 1907.
5. Casher, O.; Chandarmohan, G. K.; Hargreaves, M. J.; Leach, C; Murray-Rust, P.; Rzepa, H. S.; Sayle, R.; Whitaker, B. J.; *J. Chem. Soc., Perkin Trans. 2*, **1995**, 7.
6. Bachrach, S. M.; In *Proceedings of the 1996 International Chemical Information Conference*; H. Collier, Ed.; Infonortics: Calne, England, 1996; pp. 135-140.
7. Rzepa, H. S.; Casher, O.; Whitaker, B. J.; In *Proceedings of the 1996 International Chemical Information Conference*; H. Collier, Ed.; Infonortics: Calne, England, 1996; pp. 141-148.

8. Bachrach, S. M.; Murray-Rust, P.; Rzepa, H. S. Whitaker, B. J.; *Network Science* **1996**, 2, URL: http://www.netsci.org/Science/Special/feature07.html
9. Murray-Rust, P.; Rzepa, H. S.; Whitaker, B. J.; *Chem. Soc. Rev.* **1997**, 1-10; URL: http://www.rsc.org/is/journals/current/chsocrev/csr398.htm
10. Bachrach, S. M.; URL: http://www.chem.niu.edu/IJC/Text/
11. Bachrach, S. M.; Burleigh, D. C.; Krassavine, A.; *Issues Sci. Technology Librarianship* **1998**, *17*, URL: http://www.library.ucsb.edu/istl/98-winter/article1.html
12. Borenstein, N.; Freed, N. "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies", Internet RFC 1521, Bellcore, Innosoft, September 1993.
13. Sayles, R.; *RasMol*, Glaxo-Wellcome, URL: http://www.glaxowellcome.co.uk/netscape/software/down.html
14. Rzepa, H. S.; Murray-Rust, P.; Whitaker, B. J.; *J. Chem. Inf. Comp. Sci.* **1998**, in press.
15. Bachrach, S. M.; Krassavine, A.; Burleigh, D. C.; *J. Chem. Inf. Comp. Sci.* **1998**, submitted.
16. Murray-Rust, P.; URL: http://www.venus.co.uk/omf/cml/doc/index.html
17. *XML, Principles Tools and Techniques*; Connolly, D., Ed.; O'Reilly: Sebastopol CA, 1997
18. Murray-Rust, P.; *JUMBO*; URL: http://www.venus.co.uk/omf/cml/download.html
19. *Internet Journal of Chemistry* URL: http://www.ijc.com/
20. *Journal of Molecular Modeling* URL: http://www.organik.uni-erlangen.de/info/JMOLMOD/