In Focus

Scholarly Communication 2.0: Evolution or Design?

David P. Martinsen*

American Chemical Society, 1155 16th Street, NW, Washington, DC 20036

*Corresponding author, d_martinsen@acs.org.

Hisham Ibrahim, Photodisc, Getty Images

Several sessions at the 223rd ACS National Meeting and Exposition in Chicago focused more or less on what has come to be known as Web 2.0. Although the term is quite common in the communities of web development and publishing, and also in the library community (where it is known as Library 2.0), it may not be as familiar to people outside those areas. Many of you reading this may not be aware of the term and, therefore, would not catch the significance of "2.0", which is being appended to many terms these days, including Scholarly Communication 2.0. As with many terms that slip into our consciousness because of a certain amount of hype, the meaning of Web 2.0 is somewhat ambiguous. Web 1.0 focused on the search and retrieval of documents, albeit a vast number of documents, on the Internet. Web 2.0 involves a more interactive experience with the Internet. Tim O'Reilly, O'Reilly Media, Inc., who coined the term (*1*), recently attempted to redefine it in simpler terms: "Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as a platform, and an attempt to understand the rules for success on that new platform. Chief among those rules is this: Build applications that harness network effects to get better the more people use them. This is what I've elsewhere called 'harnessing collective intelligence'" (*2*). This is slightly different from his earlier definition and is somewhat reminiscent of the phrase coined by John Gage of Sun Microsystems during the 1980s: "The network is the computer", which became Sun's motto for several years. It is perhaps Web 2.0, more than 20 years later, that has made that idea a reality. Word processing documents, spreadsheets, and email, which used to be stored on an individual's computer, can now be stored at one or many sites on the web, where others can view, enhance, and even modify the content. Add to that photos, videos, and songs, coupled with indexing and searching tools that search both the local computer and the web simultaneously, and the result is that the boundaries between the local machine and the network have become blurred. In the extension of Web 2.0 to the scientific communication arena, the relevant technologies are collaborative tools that allow multiple people to pool their collective intelligence to author and tag content, such as blogs, WIKIs, and folksonomies; small interactive software applications to make the web more dynamic and interactive; deeper tagging of content; syndication and microformats, which allow documents or subdocuments from multiple sources to be incorporated into a single composite view; and ubiquitous computing, a proliferation of devices and networks that enable access to the web almost anywhere, at any time.

At the ACS meeting, several sessions offered by the Division of Chemical Information (CINF) and the Division of Chemical Education (CHED) focused on Web 2.0 topics; however, some ran simultaneously, so it was impossible to attend them all. It would also be impossible in this report to capture the breadth of topics and the enthusiasm with which they were presented. Fortunately, some sessions have been posted in PowerPoint, PDF, or audio format (also

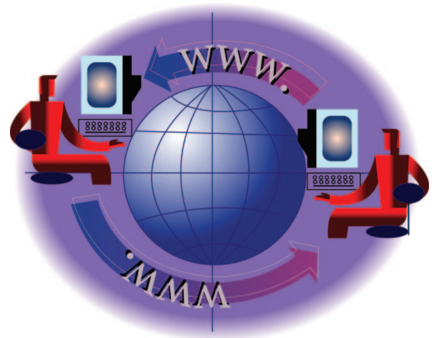
Published online June 15, 2007
10.1021/cb700111w CCC: $37.00

known as podcasts, in Web 2.0 parlance), so it is possible to get the benefit of some of the presentations and discussions, even if you could not attend in person. (The CINF sessions are listed at www.acscinf.org. Click on Meetings, 233 Chicago, and Symposia. Abstracts of all talks are posted, as well as the slides and audio files for speakers who allowed posting. The abstracts of the CHED sessions are at http://divched.chem.wisc.edu. Click on Meetings, and then look under Past Meetings for the material from the ACS meeting.)

The session that is the subject of this report is the Evolving Network of Scientific Communication. It consisted of various papers from the publisher side of the network and highlighted several areas in which publishers have enhanced their publishing models to take advantage of the technological advances in hardware, software, and telecommunications. Many of these started out as, or continue to be, experiments or pilot projects. A number of papers examined this new era of scientific communication from the perspective of scientists themselves. Most of these scientists have at least some specialization in the field of computer science or chemical informatics and therefore have a strong interest in extending the technology surrounding the posting and usage of communications in a digital world. These two groups might be viewed as the design side of the titular question: people who are interested in advancing the infrastructure of the Internet take advantage of the new technologies and thereby increase the efficiency and utility with which scientists locate and integrate that content within their own research environments. Much progress has been made in these efforts.

The symposium papers fell into several categories, not necessarily in the order they were presented. Several of the talks discussed the concept of semantic tagging, the introduction of tagging into data or text to provide context that could subsequently be understood and processed by a computer. Michael Frenkel from the National Institute of Standards and Technology (NIST) in Boulder described the methodology used by NIST, in cooperation with several publishers, to capture richly tagged thermodynamic data from published articles. Tony Hey from Microsoft described procedures to allow scientists to post experimental data on the web and to more seamlessly bridge the gap between data collection, data analysis, and data publication. Nick Day of the University of Cambridge (U.K.) described procedures to harvest data, validate them to some degree, alert scientists of the posting of new data, and provide a capability to view the data in a sophisticated viewing environment. Henry Rzepa of the Imperial College, London, focused on what he called the "golden moment" in the creation of articles, the point at which the author has all of the relevant information at hand and is motivated to create the best article for publication. If the author and publisher working together at that point could enable this semantic tagging, it could be accomplished with greater accuracy and efficiency. Another example of this semantic tagging was presented by Colin Batchelor of the Royal Society of Chemistry (RSC, London) in a separate session on Advanced Mining and Use of Life Science Information. The RSC has introduced semantic tagging within the text of journal articles. The tags can be made visible *via* Web 2.0 technology and enable readers to link to chemical structure records, IUPAC term definitions, and Gene Ontology terms. Semantically rich articles and data repositories can be used for validation, to point out potential problems with data, for discovery, to help users more readily find material of interest, and for linking between related articles.

Another set of talks dealt with the deployment of features to enable more interaction with users on the part of publishers and societies. Dennis Loney from the ACS Member Resources and Technology Department has been involved with collaborative features on



Shakirov, Photodisc, Getty Images

the ACS web site. He gave an overview of some successful community sites on the web and then described the BiotechExchange, ACS's community site for biotechnology. Evelyn Jabri, the Executive Editor of *ACS Chemical Biology*, described some of the community and interactive features of this journal—a WIKI, podcasts, and the rendering of molecules with the Jmol application. Joanna Scott from the Nature Publishing Group (NPG) described the community features in *Nature*, touching on its podcasts and its social bookmarking site, Connotea. Michael Dennis of the Chemical Abstracts Service (CAS) also mentioned some of the community features that CAS is considering. The speakers reported on the rationale behind the development of these electronic offerings, but they also noted that the usage level and the ways in which these features were used did not always correspond with initial expectations. In fact, one of the other speakers in the session noted that he was especially surprised by the popularity of podcasts, which in *Nature*'s case made it to the "Top 100" in iTunes.

The discovery of relevant information from the billions of documents on the web, as well as navigational aids to help readers find relevant information, was the topic of another set of talks. Michael Dennis of CAS talked about the data mining and visualization tool from CAS that helps users make sense of large datasets. Anurag Acharya, the developer of Google Scholar, discussed Google's approach to index all of the world's

Has the new breed of scientists evolved more efficient mechanisms for reading and assimilating content in order to keep up with the data explosion?

literature, in the language of publication, and provide a sensible algorithm to give the reader the best answer to his or her search. Marc Krellenstein, the Chief Technology Officer at Elsevier, and Georgios Papadopoulos, the CEO of Atypon Systems, both discussed the methodology for suggesting documents to readers on the basis of past reading habits and the reading habits of similar readers. Anyone who has used Amazon.com is familiar with these recommendation engines and may view them as extremely helpful, annoyingly obtrusive, or perhaps both at different times. The goal is to make reasonable suggestions based on a reader's real interests.

These last observations take us to the final category of papers, a consideration of the impact of Web 2.0 technologies on real users. Looking at the capabilities available is interesting and exciting to those who are working to make those changes happen, but how is the life of the scientist evolving to make use of those new technologies? As perturbations have been made to the traditional scientific communication environment, what changes have been observed in the behavior of scientists to capitalize on those changes? Is a new type of scientist emerging who is more competitive for survival in this environment? Will more traditional scientists be left by the wayside, or does all of this new technology simply serve as a distraction from the real work of scientific research?

Allen Renear provided some insight into these questions. Renear is at the Graduate School of Information Science and Technology at the University of Illinois, Urbana–Champaign. He presented various statistics that show that changes in behavior are beginning to emerge as a result of what he characterizes as the revolution that was accomplished by Web 1.0. In Web 1.0, virtually all published content for most scientific journals has become available on the web, from volume 1, issue 1. Renear noted that the result of this change is that most scientists have stopped browsing printed journals. However, on average, the time spent looking for articles has nearly doubled since the late 1990s. The number of articles scientists are reading has increased nearly 30%, but the time spent reading is roughly the same over that time period. Has the new breed of scientists evolved more efficient mechanisms for reading and assimilating content in order to keep up with the data explosion? Or, as Renear suggests, are scientists discovering ways to make use of the article, *via* news stories, blogs, Connotea-like sites, and citation management download features, without reading the entire article or without actually reading any of the article? Are articles now being cited without having been read? Or is this a long-standing practice simply magnified by easier access to metadata in today's digital environment? In either case, the change in behavior could have implications for science in the future.

Johan Bollen from the Los Alamos National Laboratory discussed a study that looks at usage data of journal articles, as well as other content, within and across organizations and publications. The goal of the study is to see whether there are ways to model a community based on its reading habits (or perhaps more accurately, its downloading habits; just because an article is downloaded does not mean somebody read it). Through an analysis of usage logs, Bollen can determine what topics a researcher is interested in, what common interests exist across an organization, how organizations differ in interests, and even how the downloading of a specific article propagates through an organization. Bollen wants to calculate as many usage metrics as possible and then study the correlation among those metrics. ISI's Impact Factor, Google PageRank, and the H-index are all used for ranking journal or author status, and each has its weaknesses when applied in a simplistic fashion. With more detailed metrics based on usage, Bollen hopes a more meaningful approach can be found to assess the status of a scientist's work and his or her contribution to the field. As new metrics for gauging a scientist's status emerge, we can expect that scientists will adapt to maximize their rankings by those metrics.

My final comments about the ACS meeting concern Albert Fahrenbach, a student in organic chemistry at Indiana University. He is a blogger who hails from the chemistry department, rather than the School of Informatics, whose faculty and students often participate in CINF programs. In his talk, he discussed collaborative tools for enhancing education in a public setting and research collaboration within a private setting. In discussing the drivers of the new social communities, he acknowledged the time pressure on young scientists, particularly on young faculty members working toward tenure. However, he felt that the advantages of the social networking tools would be compelling for younger scientists. Two days after the symposium, Albert posted his thoughts about the talks on his blog (*3*). He observed that most of the speakers in the session focused on harnessing the power of computers and algorithms to discover information and improve the efficiency of the knowledge transfer process. He suggested in his blog that person-to-person contact, enabled by the web, might in fact be far more efficient. This represents a slight, but significant, shift in our interaction with the web. Are we looking for documents, as most of today's search engines, recommendation tools, and social bookmarking tools assume? The result of a query is always a list of documents. Or are we looking for an answer to a question, which a person is perhaps best equipped to answer?

Two additional items, unrelated to the ACS meeting, are significant. Using Greasemonkey (*4*), a Firefox extension that allows anyone to write scripts that can change the way a web page looks, the Blue Obelisk group, a community of chemists who develop open source applications and databases in chemistry (*5*), has created several such scripts to enable chemistry-related features. One of these tools will insert links to blog stories about journal articles into the tables of contents of any ACS, RSC, Wiley, or NPG journal (*6*). This enhancement to a journal's table of contents is completely independent of the journal publisher. Second, a recent story in *The Scientist* reported that an author who got an idea for an avenue of research from a blog listed the blogger as a coauthor (*7*). Given the competitive world of scientific research, such attribution might not have occurred in the pre-web world. That it did occur in a Web 2.0 world is an indication that the world might indeed be changing.

The new tools associated with Web 2.0 clearly open up a variety of new capabilities for scientific communication. Given the nature of the web, these tools can be developed and deployed by publishers, societies, and individuals. Of course, when a tool is designed, the creators have an expectation for its use and purpose. However, once a program or tool is released "into the wild", often it is used much more, or less, than envisioned. Sometimes it is used in ways that were not even imagined. Tim Berners-Lee, Massachusetts Institute of Technology, pointed out this conundrum and described a cycle of design–release–abuse–redesign in a talk at the 16th International World Wide Web Conference held May 8–12, 2007, in Banff, Alberta, Canada (www2007.org/berners-lee.php). Perhaps the promise of Web 2.0 and, more to our point, Scholarly Communication 2.0 is that the scientific communication network can evolve beyond what a given publisher, society, or scientist designs. This has the potential for bringing benefits to all of science. It also brings with it some dangers. Obvious problems exist, such as spam and misrepresenting another's work as one's own. More subtle issues may not emerge for many years. We chemists, and indeed all physical scientists, need to listen to our colleagues in cognitive sciences to understand the impact of some of the changes we advocate. The ways we acquire and process information and knowledge evolved to work well with the traditional model for the publication of scholarly material. The changes now being made through new technology have the potential to disrupt not only that traditional model of publication, but also the very way in which we humans process the information. We need to monitor the ways in which those changes affect human behavior, and make adjustments for undesired results. Perhaps a model of designed evolution of scientific communication is ultimately the best approach.

REFERENCES

1. O'Reilly, T. What is Web 2.0: design patterns and business models for the next generation of software, www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (accessed May 24, 2007).
2. O'Reilly, T. Web 2.0 compact definition: trying again, http://radar.oreilly.com/archives/2006/12/web_20_compact.html (accessed May 24, 2007).
3. Fahrenbach, A. ACS reflections–the present state of collaboration in science, www.scientificblogging.com/albert/acs_reflections_the_present_state_of_collaboration_in_science (accessed May 24, 2007).
4. Greasespot, the weblog about Greasemonkey, www.greasespot.net (accessed Jun. 7, 2007).
5. Guha, R., Howard, M.T., Hutchison, G.R., Murray-Rust, P., Rzepa, H., Steinbeck, C., Wegner, J., and Willighagen, E.L. (2006) The Blue Obelisk—interoperability in chemical informatics, *J. Chem. Inf. Model 46*, 991–998.
6. Using Javascript and Greasemonkey for Chemistry, http://wiki.cubic.uni-koeln.de/bowiki/index.php/Using_Javascript_and_Greasemonkey_for_Chemistry (accessed May 24, 2007).
7. Secko, D. (2007) Scooped by a blog, *The Scientist 21*, 21–22.