Communicating chemistry

Theresa Velden and Carl Lagoze

New web-based models of scholarly communication have made a significant impact in some scientific disciplines, but chemistry is not one of them. What has prevented the widespread adoption of these developments by chemists — and what are the prospects for adoption over time?

The World Wide Web, the semantic web, and the social networking tools of Web 2.0, together with the digitization of content and the increasing processing power of computers are revolutionizing the ways in which researchers communicate with one another, disseminate results, collaborate, and share data and knowledge[1]. In a number of scientific fields, new models of science communication have emerged that are based on these technologies. Many of these have found wide community participation. Examples include the public dissemination of non-refereed, author-submitted manuscripts of research articles through the preprint server arXiv in physics, mathematics and quantitative biology[2], the sharing of research data through public databases such as GenBank in genomics, the integration of this data through PubMed Central with open-access journal literature in biomedicine[3] and community-based open peer review and discussion in interactive online journals in geosciences[4].

The success of these new web-based and social-network models in disciplines neighbouring to and at the periphery of chemistry (such as drug design) stands in contrast to the lack of comparable success stories in chemistry itself. Why do similar initiatives in chemistry fail to gain critical mass and widespread usage? This question and the opportunities offered by such initiatives were the focus of a workshop titled New Models for Scholarly Communication in Chemistry, held in Washington DC on 23–24 October 2008. The participants in the workshop included experts from the chemistry, publishing, information-science and information-policy communities.

We summarize here the results of that workshop, which are described fully in a white paper — *The Value of New Scientific Communication Models for Chemistry* — that is publically available (under a Creative Commons Attribution 3.0 License) online[5]. Discussions leading up to the final publication of the white paper showed the controversy surrounding the questions of whether and how chemistry might benefit from new models for science communication, and about what the characteristics of chemistry research and scholarship are that will shape its future communication system. As a result, the white paper reflects some of the aspects of this controversy and presents a perspective on it, rather than a consensus among all of the workshop participants.

> For any transformation of the science-communication system, incorporating the legacy data is not just a feature that is nice to have, but essential.

The workshop and subsequent research revealed that the barriers to transformation of communication models for chemistry are, by and large, not technical. In fact, there exist at present a number of technical developments that pave the way for the exchange of chemical information on the web. Examples of these initiatives include a standardized chemical mark-up language, a computable identifier for organic molecules (the IUPAC international chemical identifier, or InChI), open-source tools for the manipulation and management of chemical information[6], and the use of free, hosted Web 2.0 services to support 'open-notebook science' (http://tinyurl.com/65t45d). The pioneers of these developments promote a vision where the entire lifecycle of research — from planning of research projects and experiments, followed by conducting experiments in the lab, analysing results, and finally to dissemination and publication — is supported by capture, storage, and interlinking of the underlying (raw and derived) research data[7].

The expected benefits of such approaches include increased transparency of the research process, improved verifiability of research results and their reproducibility, increased efficiency in local management of data in research teams, and increased efficiency of global research through the ability to re-use data. In turn, these developments offer opportunities for new forms of research by processing and mining large aggregated data sets that may facilitate the distributed and open research of under-researched areas (such as tropical disease) that lack commercial incentives, but which would benefit from 'crowd sourcing'. Such efforts, however, entail the mobilization of a large number of independently contributing volunteers, although cases that demonstrate its potential (and success) do exist. In astronomy, for example, the Galaxy Zoo project is used to classify galaxies in a large survey of the sky[8].

These and other prototype projects demonstrate technical and conceptual feasibility. So far, however, these initiatives have failed to reach critical mass and have not become an integral part of the science-communication system in chemistry. Outside of specific subfields such as cheminformatics or crystallography, few chemists seem to perceive these developments as opportunities to enhance science-communication practices[9].

Perhaps this is only an issue of latency. Maybe we will eventually see substantial changes to the present models of science communication in chemistry, if only delayed in comparison with other disciplines. However, such simplistic assumptions of 'technical determinism' or inevitability have fallen into disfavour with those who study science-communication systems[10]. Instead, present thinking on the transformation of scholarly practices emphasizes their historical contingency and notes that they are social as much as technical arrangements, where the social and the technical aspects mutually shape one another[11,12]. We agree with this broader perspective and note that our use of the term 'new models of scholarly communication' indicates a complex socio-technical arrangement rather than just a technical system.

Concerning this broader perspective, we pose a number of fundamental questions about the adoption of new models of scholarly communication in chemistry. Should the historical resistance to new models be taken as an indication of an extreme mismatch between the opportunities offered by new information and communication technologies and the research practices and values of the chemistry community? Or, are there particular barriers that could be sensibly addressed? Or, does the existing system in chemistry already provide the best possible match to research needs, meaning that no fundamental changes are expected from new technological capabilities apart from incremental improvement?

Also, is it misleading to suggest that there is a deficiency of the present science-communication system in chemistry, and to imply that other disciplines are more advanced in their adoption of web-based communication models? For instance, one could argue that there is no significant difference in the way chemistry has taken up web-based information and communication technologies compared to other disciplines. On closer examination, we see that innovations like arXiv or PubMed Central have been adopted in specific fields in physics and the life sciences respectively, but not uniformly across the entire disciplines. So how would one sensibly compare developments in different disciplines and learn from such comparisons?

The present reward system in science inhibits any research-driven change in science communication that focuses on its communicative function.

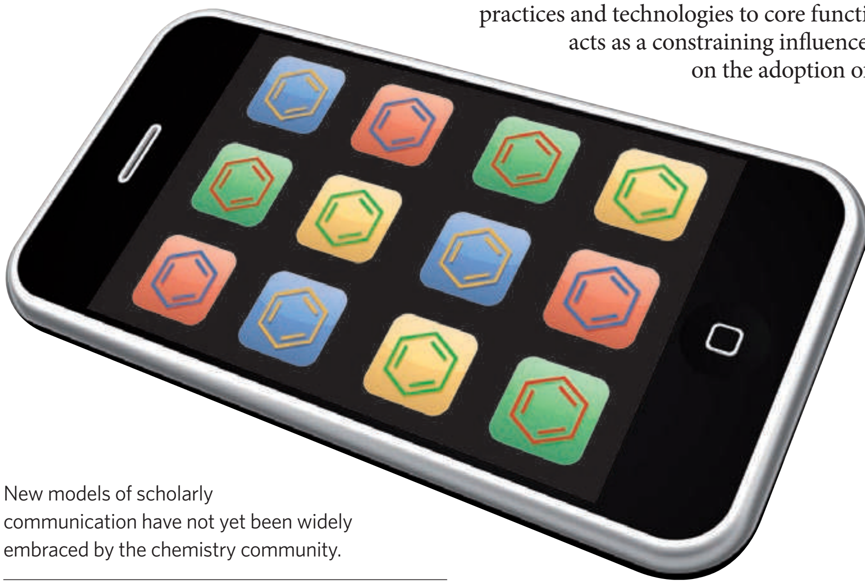
As a starting point for a complete examination of these questions, we need to understand the particular ways in which science communication supports knowledge production in chemistry, and how this can be meaningfully compared to other disciplines. The initial results of this analysis and comparison are fully articulated in the white paper, the contents of which are summarized below.

Initiatives for publishing and e-science

The broader context of innovation was analysed by reviewing publishing reforms and e-science initiatives — such as open access, data publishing, and preprint servers — in other disciplines. An understanding of the dynamics of transformation within these successful initiatives served as a useful benchmark for the analysis of the value of new science-communication models in chemistry. An underlying assumption of this review and the subsequent analysis was the recognition that science communication extends beyond standard formal channels, such as the publication of articles in peer-reviewed journals or conference proceedings, and includes informal communication between researchers in all kinds of settings. Furthermore, our analysis accounts for the fact that science-communication systems are 'socio-technical interaction networks'[12], which fulfil a set of functions deeply embedded in the nature of scholarship itself. These functions include the registration of new results, the certification or validation of these results, the distribution or exposure of them to peer communities, the granting of rewards such as tenure or promotion resulting from successful research results, and the preservation of the scientific record[13,14].

The issue of reward deserves special attention in our analysis because of its strong influence on scholarly behaviour and receptivity to communication innovations. In most disciplines, the publication of journal articles has become a key metric for career advancement. This coupling of the science-communication and reward systems in science has recently become even more closely linked with the reliance on journal impact factors and citation measures, such as the Hirsch index, to assess scientific quality and researchers' performance. This development, in combination with rigid academic stratification between and within institutions, presents a strong reason for scientists to be risk-averse when experimenting with new models of science communication[15]. Hence, the present reward system in science inhibits any research-driven change in science communication that focuses on its communicative function rather than its role as a proxy for the assessment of research performance.

Our analysis of the use of communication innovations and their effectiveness demonstrates other facets of the deep embedding of the communication system in scholarly practices. It also reveals the manner in which the linkage of existing practices and technologies to core functions acts as a constraining influence on the adoption of innovations across the scholarly landscape. We briefly review here some principal strands of innovation, and provide some observations on the way their uptake illustrates the dynamics of the transformation of science-communication systems.



New models of scholarly communication have not yet been widely embraced by the chemistry community.

© ISTOCKPHOTO / ZIMA FILM

Open access. Open access is concerned with free-to-the-user web-based access to the results of publicly funded research to optimize their impact and ease of use and re-use. The underlying motivations for open-access movements and their various declarations[16–18] are the desire to develop an alternative to critical economic problems — the so-called serials crisis, whereby the correspondence of dramatically increasing subscription prices and decreasing budgets has forced libraries to cancel journal subscriptions[19,20] — and a belief that open access provides the foundation for new capabilities for linking and integrating the search results, for searching and mining information, and for reusing data. One approach to open access is to remove the limitations in access to the journal literature that arise from subscription costs by switching to a model in which readers can access published articles for free. An article processing fee from the author or the research funding institution covers production costs. A number of issues are associated with such a switch that play out differently in different disciplines. For example, in the humanities, costs of research equipment are usually low and an increase in publishing costs resulting from an article processing fee would be perceived as quite high and daunting.

Preprint servers. Several fields have adopted web-based mechanisms for the dissemination of preprints, working papers, or final peer-reviewed articles. In general, these can all be characterized as the dissemination of research reports that have undergone minimal quality control notably distinct from the peer-review process of the formal publication system.

Comprehensive access to the chemical literature and information about chemical substances is crucial in most areas of chemical research.

The most prominent of these is the arXiv[21] preprint server, originally introduced in 1991 by Paul Ginsparg for high-energy physics, and which now has been expanded to include a large number of quantitative sciences. Advocates of preprint servers claim a number of advantages, including immediacy of access to latest research results, their role as a comprehensive and self-contained archive of the literature of a field, and the fact that they provide a long tail of research results outside the limitations of peer review.

The success of preprint servers has varied widely across disciplines. For example, an attempt in the early 2000s by a subsidiary of the publisher Elsevier to introduce an arXiv-based preprint server for chemistry failed entirely (Chemistry Preprint Server[22]). In fact, the success of the model in physics and related disciplines most probably stems from the fact that the practice of exchanging preprints in those fields precedes the existence of the Internet. Notably, most of the content within preprint servers eventually appears also in the peer-reviewed journal literature, such that preprint servers serve a complementary function, and do not substitute for peer-reviewed publication venues.

Data publishing. The aim of data publishing is to make research data publicly available in a reusable format to support reuse and validation of existing research results. A critical issue for data publishing is the lack of uniformity of motivations of researchers to share data at some point. The factors involved include the stage of research, the type of data, the investment made in its generation, the effort needed to make it available in a useful form, and the value of the data for the community as a shared resource versus the competitive advantage of keeping it private for further exploitation. Because of these factors, different trade-offs exist for making one's data publicly available[23,24] that vary widely across fields.

Science blogs. An increasing number of scientists have adopted blogging as a means of informal communication. Typically, the writing style of blogs is conversational, and humorous content gets mixed with posts of a more serious tone. Some blogs are dedicated to educating lay audiences, others aim at an academic discussion, and many are like personal diaries. At this point in time, many science bloggers are assumed to be less than 30 years old, and are primarily journalists, teachers, graduate students or young researchers[25]. Hardly any established scientists maintain a blog — after all, blogging regularly is very time consuming[26]. The question remains open whether these will remain fringe phenomena or become part of the mainstream communication in science.

Our analysis shows that science communication is tightly interlinked with disciplinary practices and cultures. Although successful new science-communication models extend capabilities and support new practices in the contexts in which they have been implemented, they also represent continuity by building on pre-existing practices and values. As a result, when new models are transferred from one field to another (for example, from physics to chemistry), they may need to be significantly altered to match different pre-existing practices before they can become successful.

Chemistry communication system

Comprehensive access to the chemical literature and information about chemical substances is crucial in most areas of chemical research. The chemical societies such as the American Chemical Society (ACS), the Royal Society of Chemistry (RSC) and the Gesellschaft Deutscher Chemiker (the German Chemical Society) are chief providers of chemical information, in the form of journals and databases. These societies have a strong historical track record in shaping the science-communication system of chemistry. Scientific results may get communicated at conferences managed by these societies in talks, abstracts and posters, but the enduring scientific record is established through publication in peer-reviewed journals.

The chemistry community and the societies are historically receptive to selected communication innovations. In fact, the ACS was one of the first chemistry

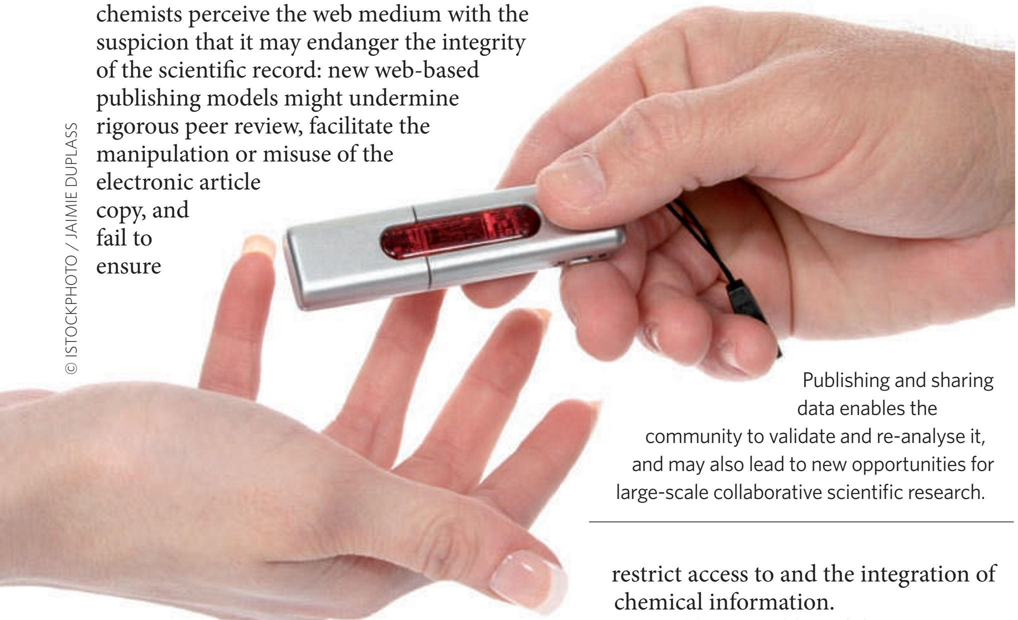
© ISTOCKPHOTO / DRA_SCHWARTZ

Web 2.0 services such as blogs and wikis offer new opportunities for disseminating and sharing research results and data.

publishers to experiment with electronic versions of research articles, as early as pre-web times in the 1980s[27]. Nowadays all principal chemistry journals are available online. They mostly reproduce the paper-based article in electronic form by offering printable PDF files and sometimes an HTML version of the article. The online versions of the journals play to the strong visual orientation of chemists for processing information[28] by offering graphical abstracts in the table of content listings.

Experience has shown that online-only journals have not found wide acceptance among chemists. The RSC launched the online-only journal *PhysChemComm* in 1998 to take advantage of the web for rapid publication and to include colour images, animated graphics and movies in their article. At the end of 2003 the journal was discontinued. The RSC explained that the submitted articles contained fewer features for electronic enhancement than they had expected, and that print journal production had sped up to such an extent that "it has become clear that there is now no real advantage to an e-only journal in the physical chemistry field"[29].

Certification of articles through peer review organized by journals is clearly highly valued in the chemistry community. Peer review is seen as preventing factual mistakes (if not always), and enhancing the editorial quality of articles. It has been suggested that chemists perceive the web medium with the suspicion that it may endanger the integrity of the scientific record: new web-based publishing models might undermine rigorous peer review, facilitate the manipulation or misuse of the electronic article copy, and fail to ensure perpetual access[30]. As a result, no 'preprint culture' exists in chemistry that assigns value to chemists who publicly disseminate their not-yet-peer-reviewed manuscripts — on the contrary, to disclose information before priority has been established through a formal journal publication seems to be perceived as too risky.



© ISTOCKPHOTO / JAIMIE DUPLASS

Publishing and sharing data enables the community to validate and re-analyse it, and may also lead to new opportunities for large-scale collaborative scientific research.

Access to databases is vital in all areas of chemical research. To fulfil this need, since the 1960s, comprehensive databases have been serving the chemistry community both in academia and in industry. These databases can be broadly classed into three categories of information: literature and patent databases, structure and reaction databases, and factual databases (chemophysical properties including spectra); a number of databases provide access to several of these information types. Prominent among these are the Chemical Abstracts Service (CAS), Beilstein and Gmelin, which are all widely used by most practising chemists, and in particular synthetic chemists. They have a century-long history, and have been built into valuable resources deeply rooted in chemists' research practices.

Recently, a number of web-based innovations and experiments have emerged to enhance the communication and management of chemical information. So far they have not found a wide uptake such that they would change or extend science-communication practices in chemistry in a fundamental way. One of the main hurdles, according to many of the pioneers, is the existence of the proprietary regimes in which most chemical data reside that restrict access to and the integration of chemical information.

An abbreviated list of the innovations designed to encourage the communication and management of chemical information is as follows:

Semantic chemistry web. The vision for a semantic web in chemistry is driven by interests in large-scale data mining in cheminformatics to support drug discovery[31,32], as well as the perceived benefit of integrating work of small scale labs by better managing their diverse data. Semantics aid the discovery and reliable re-use of data, and reduce ambiguity for later automatic processing of data[7]. Although there are a number of challenging technical problems to be overcome for the realization of a semantic chemistry web, the chief obstacle is the access problem. Unfortunately, the vast majority of accumulated chemistry knowledge is locked up in proprietary silos in a manner that prevents semantic linking — the field "has ceded the dissemination of data and knowledge almost entirely to commercial entities in the form of publishing businesses"[33].

Electronic- and open-notebook science. Electronic lab books help to capture data at the source, when they are created in the laboratory. One example is the semantic electronic lab notebook, which pursues a semantic-web approach in the capture of data and facilitates internal management and eventual publication of the data created[34]. The emphasis in this approach is on capturing the data during the research process in digital format with as little effort for the researcher as possible (for example, by having instruments providing the data in standard formats that can be read into the electronic lab notebook). This way a complete-as-possible provenance trail of data could be established that later would facilitate re-use of the data as well as public dissemination if so intended. Another experiment in this area is the open-notebook-science vision[35], which makes the entire primary record of a research project available almost in real time, as it is created. Its supporters expect an increase in efficiency of the scientific process through greater transparency.

Data publishing, discovery and access. The idea of data publishing is to make experimental data publicly available in a way that they (i) can be referred to and cited by a stable identifier, and (ii) are provided in a reusable format. This would increase the value of experimental data by allowing the scientific community to do a number of things, such as double-check a result that has been reported in the literature, conduct alternative analyses of the data possibly not anticipated by the creators of the data, recalculate quantities from the data when new calculation methods become available, or aggregate data for data mining. One route to increasing the availability of data is the post-hoc extraction of data from published literature. Crystal Eye[36] is a web-based service that demonstrates the benefit of this extraction, and uses a standard format for data publishing.

An alternative to the post-hoc approach of extracting data from the published literature is to make data sets publicly available by their creators through direct deposit in a web-based repository. In this manner, data sets are published by themselves, independent of or in parallel with a journal publication, thereby increasing significantly the amount of available data. There are a number of initial efforts in this area, including eCrystals[37] in the UK and Reciprocal Net[38] in the USA. As the amount of open web-based chemical data grows, search engines and web crawlers that aggregate chemical information and provide services on top of these different resources become more relevant. An example of a chemistry search engine under development is the National Science Foundation (NSF) funded project ChemXSeer[39], which is modelled after the CiteSeer search engine for computer-science literature, but with added intelligence to deal with chemical information.

Chemistry distinguished

We posit that the interplay of several factors distinguishes chemistry from other disciplines and constitutes a unique context of science communication in chemistry that may influence the adoption of new communication models.

> We posit that the interplay of several factors distinguishes chemistry from other disciplines and constitutes a unique context of science communication.

The distinguishing characteristics of chemistry can be grouped into two classes: research practices and socio-economic organization. Some examples of research practices include:

Focus on creation. Chemistry has been characterized as distinct from other sciences by being primarily the "art, craft, and business of substances and their transformation" and only subsequently a science[40]. Producing new chemical substances is a central research activity in chemistry[41]. The emphasis in papers where the synthesis of these new substances are reported is on the synthesis itself, and the application of the new substance in generating other new substances[42] — not on the analysis or development of a theory[40,43]. There is limited emphasis on reproducibility or reusability and thus little incentive for using web technologies to publish data more widely in a reusable manner.

Long-tail science. Chemistry can be characterized as a long-tail science (http://tinyurl.com/33t82z). It is a field of research dominated by large numbers of small research-producing units (in contrast to the large-scale collaborations of high-energy physics). The predominantly non-collaborative mode of research in chemistry reduces the incentive to make use of new technologies to facilitate data sharing and research collaboration.

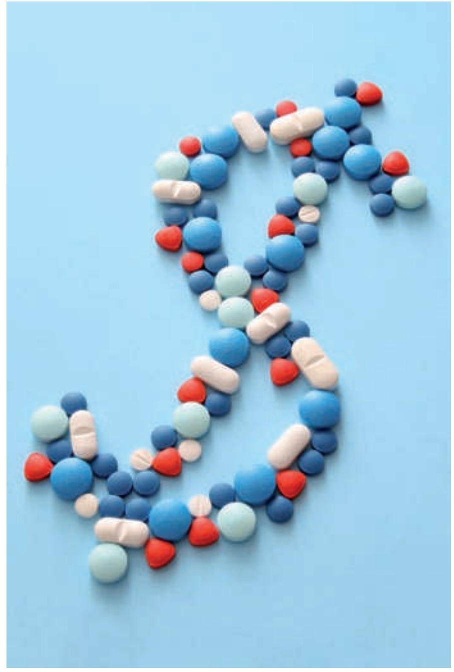
Longevity of scientific literature and data. Another hallmark of chemistry as a science is the enormous knowledge base of scientific data that has been accumulated over more than a century. The millions of data sets reported in the literature and captured in databases that record chemical structures, and the physical and chemical properties of chemical substances retain immediate value for most of the present research in chemistry. Thus, for any transformation of the science-communication system, incorporating the legacy data is not just a feature that is nice to have, but essential for the use of a new system to chemists.

Non-digital practices. Preparative chemistry as 'art and craft' emphasizes manual practices. Generally, the computer is not the central tool for generating chemical knowledge. Thus, it is a significant challenge to integrate digital data capture into the workflow of chemists in the laboratory.

Diversity of research cultures in chemistry. Research cultures in chemistry are all but unified. There is a plurality of historical research traditions, methods and goals to which research fields in chemistry adhere, as well as a variety of interdisciplinary projects. A principal division is between those people who make molecules and determine their structure, in contrast to those who study their properties. In the first group are synthetic organic and inorganic chemists, and polymer chemists; in the second are physical, theoretical, and analytical chemists. Synthetic organic and inorganic chemists probably publish a little more, although the unit of publishable research is smaller. An alternative division is synthesis ('I made it!'), analysis ('what do I have?'), mechanism ('how did it happen?') and theory ('why, oh theorist, why?'). The synthesis and analysis people share one subculture, the mechanism and theoretical people another. A consequence of these divisions is that the perceived need for improvement, and readiness to innovate will differ across subfields in chemistry, and correspondingly any technical innovation must support a diversity of communication practices.

The other class of distinguishing factors relate to the socio-economic organization, and includes the following examples:

Proprietary chemical information. Chemistry is distinguished from most other disciplines in that the chemical information that is produced in everyday academic research is of considerable relevance for a huge, profitable chemical industry, which is the prime user of that information. Thus, it has been suggested that chemists are more secretive about details of their research in formal and informal communication than scientists in many other disciplines.



Chemical information is at the heart of a number of very profitable industries and this shapes the publishing and data-sharing practices of many chemists.

Industry–academia balance in chemistry. Some chemists point out that academia produces two vital inputs for the chemical industry, trained PhD-level scientists and published scientific results, without proper compensation. Industrial researchers read the scientific literature but they publish only sparsely themselves, because their careers do not depend on it, and to keep their research strategies and goals secret from competitors. This may explain why some academic chemists are particularly sceptical of present proposals for open-access business models, as they feel that industry would profit inappropriately.

ACS's global responsibility. The world's largest scientific society is the ACS, which has about 155,000 members, including 19,000 international members. It is a non-profit organization, but nevertheless behaves very much like a commercial entity with regard to the information services it develops and offers, such as its CAS and proprietary identifier system. This behaviour may derive from the dominant role of members and customers from the commercial sector in the society, who tend to perceive of chemical information mainly as an economic asset, rather than as a common good. It could be argued that CAS's proprietary policy towards large-scale use of the identifier system undermines widespread experimentation and innovation by third parties that rely on the accurate integration of chemical information. Such integration is necessary for a web of shared chemical resources.

Concluding remarks

We are disseminating the white paper and publishing this Commentary to provide a preliminary starting point for a more extensive analysis that addresses the full breadth of research in chemistry. This is a complex and controversial topic that would benefit from review and discussion by a broad range of chemists from academia and industry, and from other professionals with an interest in the matter (publishers, information service providers, representatives of scientific societies, information and social scientists, and funding agencies). We are working with others to find a way to bring these various stakeholders of science communication in chemistry together at a second, international workshop.

We believe that a broad involvement of practising chemists is needed to reflect on the opportunities provided by new technological capabilities to improve science communication in chemistry, to examine and critically appraise innovative proposals and prototypes, and to develop models that may go beyond what is being discussed by stakeholders invested in the present system. Attention to this subject is essential to ensure that (i) opportunities are not missed, (ii) failures can be avoided and resources not wasted, and (iii) future initiatives in chemistry and elsewhere to develop new communication models recognize the importance of particular, discipline-specific contexts and non-technical challenges. We believe that the results of this in-depth analysis will benefit the future of chemistry research and scholarship. ❐

Theresa Velden and Carl Lagoze are in the Information Science Program, Cornell University, Ithaca, New York 14850, USA.
e-mail: tav6@cornell.edu; clagoze@gmail.com

References

1. Arms, W. & Larsen, R. *The Future of Scholarly Communication: Building the Infrastructure for Cyberscholarship* (NSF/JISC, 2007); available at <http://www.sis.pitt.edu/~repwkshop/SIS-NSFReport2.pdf>.
2. Gunnarsdottir, K. *Soc. Stud. Sci.* **35,** 549–579 (2005).
3. Benson, D., Karsch-Mizrachi, I., Lipman, D., Ostell, J. & Wheeler, D. *Nucleic Acids Res.* **35,** D21–D25 (2007).
4. Pöschl, U. & Koop, T. *Information Services & Use* **28,** 105–107 (2008).
5. http://hdl.handle.net/1813/14150
6. Guha, R. *et al. J. Chem. Inf. Model.* **46,** 991–998 (2006).
7. Frey, J. *Drug Discov. Today* **14,** 552–561 (2009).
8. www.galaxyzoo.org/
9. Todd, M. *Chem. Cent. J.* **1,** 3 (2007).
10. Van House, N. A. *Annu. Rev. Inform. Sci.* **38,** 3–36 (2004).
11. Hilgartner, S. *Sci. Commun.* **17,** 240–263 (1995).
12. Kling, R., McKim, G. & King, A. *J. Am. Soc. Inf. Sci. Tec.* **54,** 47–67 (2003).
13. Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C. & Warner, S. *D-Lib Mag.* **10,** (2004); available at <http://tinyurl.com/nlz3s>.
14. Roosendaal, H. E. & Guerts, P. A. T. M. in *CRISP 97: Cooperative Research Information Systems in Physics* (eds Karttunen, M., Holmlund, K. & Hilf, E. R.) (Oldenburg, 1997).
15. Armbruster, C. *Learn. Publ.* **20,** 259–266 (2007).
16. www.soros.org/openaccess/
17. www.earlham.edu/~peters/fos/bethesda.htm
18. http://oa.mpg.de/openaccess-berlin/berlindeclaration.html
19. Cope, W. W. & Kalantzis, M. *First Monday* **14,** Number 4–6 (2009); available at <http://tinyurl.com/ctta9o>.
20. Edlin, A. S. & Rubinfeld, D. L. *Exclusion or Efficient Pricing? The "Big Deal" Bundling of Academic Journals* (Univ. California, 2004); available at <http://repositories.cdlib.org/blewp/art167/>.
21. www.arxiv.org
22. www.sciencedirect.com/preprintarchive
23. Birnholtz, J. P. & Bietz, M. J. in *International ACM SIGGROUP Conference on Supporting Group Work* (eds Tremaine, M. & Simone, C.) 339–348 (ACM Press, 2003).
24. Hilgartner, S. in *Intellectual Property Rights and Research Tools in Molecular Biology* (ed. National Research Council) Ch. 4, 28–39 (Natl Acad. Sci. USA, 1998).
25. Bonetta, L. *Cell* **129,** 443–445 (2007).
26. Wilkins, J. *Trends Ecol. Evol.* **23,** 411–413 (2008).
27. Garson, L. *Acc. Chem. Res.* **37,** 141–148 (2004).
28. Hoffmann, R. & Laszlo, P. *Diogenes* **37,** 23–51 (1989).
29. *Phys. Chem. Chem. Phys.* **5,** v–xviii (2003).
30. Searing, S. & Estabrook, L. *Portal-Libr. Acad.* **1,** 77–96 (2001).
31. Neumann, E. *Drug Discov. World* **6,** 25–33 (2005).
32. Tetko, I. V. *Drug Discov. Today* **10,** 1497–1500 (2005).
33. www.semanticuniverse.com/articles-semantic-chemistry.html
34. Hughes, G. *Org. Biomol. Chem.* **2,** 3284–3293 (2004).
35. Bradley, J-C., Lancashire, R. J., Lang, A. S. I. D. & Williams, A. J. *J. Cheminf.* **1,** 9 (2009).
36. http://wwmm.ch.cam.ac.uk/crystaleye/
37. http://ecrystals.chem.soton.ac.uk/
38. www.reciprocalnet.org
39. http://chemxseer.ist.psu.edu/
40. Hoffmann, R. *Synthèse* **155,** 321–336 (2007).
41. Schummer, J. *Hyle* **4,** 129–162 (1998).
42. Lipkus, A. *et al. J. Org. Chem.* **73,** 4443–4451 (2008).
43. Schummer, J. *Scientometrics* **39,** 125–140 (1997).

Acknowledgements

Funding for the New Models for Scholarly Communication in Chemistry workshop held in Washington DC on 23–24 October 2008 was provided by the NSF through grant IIS-738543 SGER: Advancing the State of eChemistry. Further follow-on support for research and report writing was provided by Microsoft External Research through the oreChem Project. The authors gratefully acknowledge contributions to the white paper from workshop participants Roald Hoffmann, Leah Solla, Hilary Spencer, Jeremy Frey, John Wilbanks, Clifford Lynch and Jean-Claude Bradley. The authors also acknowledge the additional participants in the workshop: Colin Batchelor, Christine Borgman, Jane Hunter, Peter Jerram, Susan King, Peter Murray-Rust and Heinz Saller. This Commentary and the related white paper do not represent the views or opinions of all contributors, or workshop participants or of the organizations with which they are affiliated.