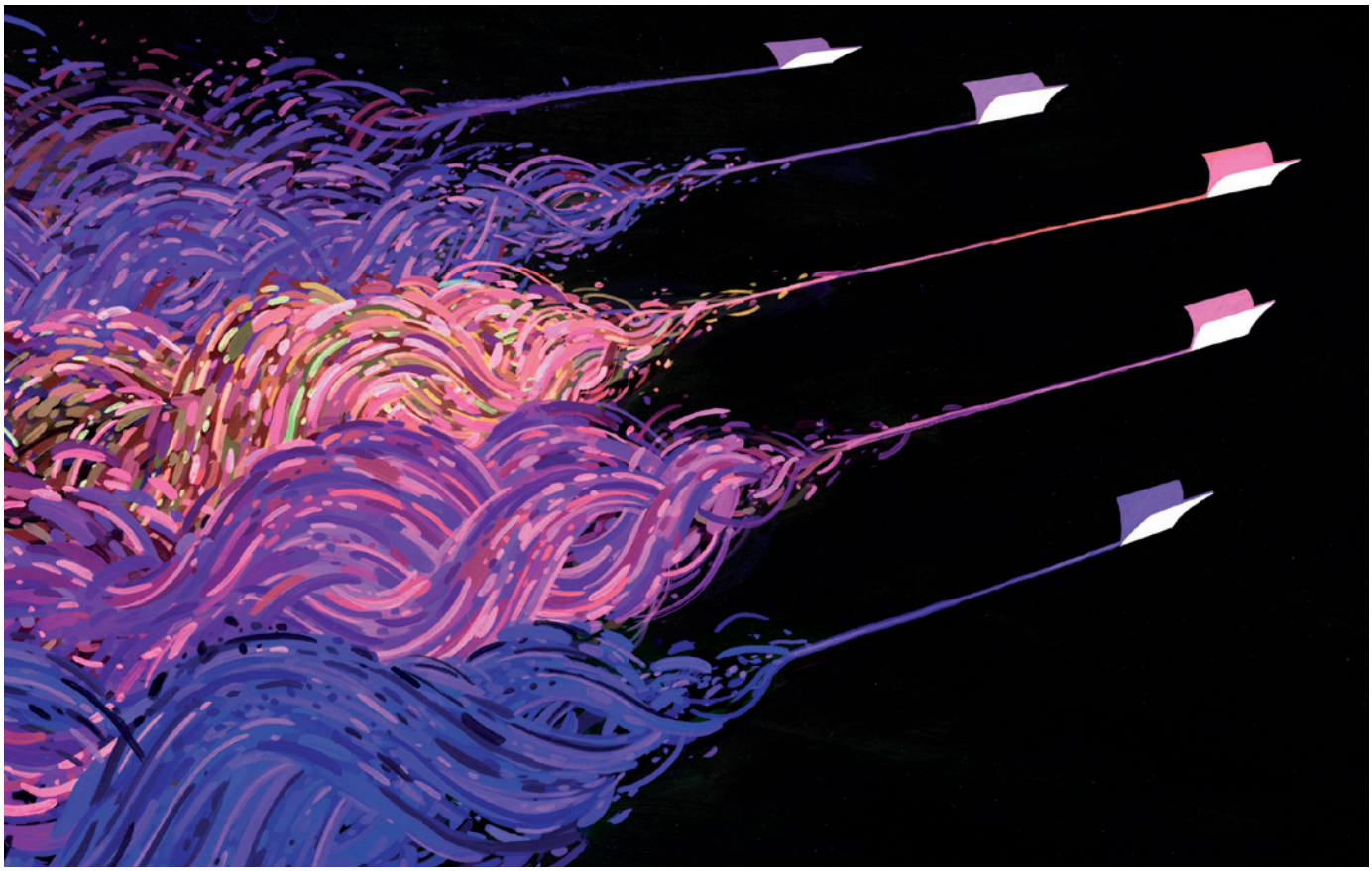
COMMENT

BRENDAN MONROE



Beyond the paper

The journal and article are being superseded by algorithms that filter, rate and disseminate scholarship as it happens, argues **Jason Priem**.

Henry Oldenburg created the first scientific journal in 1665 with a simple goal: apply an emerging communication technology — the printing press — to improve the dissemination of scholarly knowledge. The journal was a vast improvement over the letter-writing system that it eventually replaced. But it had a cost: no longer could scientists read everything someone sent them; existing information filters became swamped.

To solve this, peer and editorial review emerged as a filter, becoming increasingly standardized in the science boom after the Second World War. This peer-review system applies community evaluation of scholarly products by proxy: editorial boards, editors and peer reviewers are nominated to enact representative judgements on behalf of their communities.

Now we are witnessing the transition to yet another scholarly communication system — one that will harness the technology of the Web to vastly improve dissemination. What the journal did for a single, formal product (the article), the Web is doing for the entire breadth of scholarly output. The article was an attempt to freeze and mount some part of the scholarly process for display. The Web opens the workshop windows to disseminate scholarship as it happens, erasing the artificial distinction between process and product.

Over the next ten years, the view through these open windows will inform powerful, online filters; these will distil communities' impact judgements ▶

algorithmically, replacing the peer-review and journal systems.

THE PRESENT

This move towards a more diverse set of outputs has already begun. Scholars now share their research data in repositories such as GenBank, Dryad and figshare (figshare is supported by Digital Science, which is owned by the same parent company as *Nature*). They use repositories such as GitHub to share code, analyses and even 'executable papers' that automatically knit together preloaded data, analysis and prose. They challenge the traditional article format by including blog posts, interactive graphics and video. And perhaps most significantly, academics are moving informal scholarly conversations from the faculty lounge to social media platforms such as Twitter[1]. In the next ten years, most scholars will join such networks, driven by both the value of improved networking and the fear of being left out of important conversations.

This shift from a paper-native to a Web-native system has three significant consequences. First, the flow of scholarly information is expanding by orders of magnitude, swamping our paper-based filtering system. Like journal articles, Web-native products are of inconsistent quality and require winnowing; unlike journal articles, the scale of these products overwhelms attempts at manual curation. In the late 1990s, commercial Internet services such as Yahoo! found that hiring experts to create vetted lists of web pages completely failed at Web scale; the same will be true of scholarship.

Second, the Web era is exposing the delicate tracework of ideas beneath the formal structures of the academy[2]. The ephemera of scholarship are reified in a constellation of data points: views on figshare, mentions in discussions between colleagues on Twitter, saves in a reference manager such as Zotero or Mendeley, citations in an open-access preprint, recommendations on Faculty of 1000, and many more. We can use these and other tracers of impact to generate new metrics of scholarly influence. Informed by these alternative metrics, or 'altmetrics', we will draw new maps of scholarly contribution, unprecedented in subtlety, texture and detail. Suddenly, the rocky plain of ideas — once navigated using cairns of citation — is covered in fresh snow. In the Web era, scholarship leaves footprints.

The third consequence proceeds from the first two. The editors and reviewers employed as proxy community assessors will be replaced by the aggregated, collective judgements of communities themselves. The information-overload problem supplies its own solution. This is the power behind the Web's great filter, Google. While Yahoo! was applying paper-native expert curation to the Web, Google was aggregating the collective authority judgements inherent in the structure of the network. Its PageRank algorithm weights hyperlinks from authoritative sources more heavily. To find which sources count as authoritative, the same algorithm is applied to each of the source's inbound links, and so on. This simple recursive algorithm has proved remarkably effective, and requires minimal manual tuning. It simply harnesses the quality judgements already being made by the community, implicit in their decisions to link to other pages. This core approach is also the future for scholarly communication.

THE FUTURE

Today's publication silos will be replaced by a set of decentralized, interoperable services that are built on a core infrastructure of open data and evolving standards — like the Web itself (see 'Reconstructing publishing'). This 'decoupled journal'[3,4] publishes promiscuously, then subjects products to rigorous review through the aggregated judgements of expert communities, supporting both rapid, fine-grained filtering and consistent, meaningful evaluation.

Dissemination. The concept of 'pre-publication' will be forgotten as frictionless, Web-based dissemination increasingly pervades the research process. Conversations, data collection, analysis and description will be born published. Many researchers are already practising this 'open-notebook science'. For instance, population biologist Carl Boettiger describes his day's research and provides in-progress code, analysis and writing, and chemist Jean-Claude Bradley's lab at Drexel University in Philadelphia, Pennsylvania, publishes its entire output in near-real time. New tools are emerging to facilitate this 'share early, share often' approach. The journal *Push*, for example, lets scholars build journal articles incrementally, with each version tracked and open online, available for collaboration and comment throughout (see http://push.cwcon.org).

"Tools are emerging to facilitate this 'share early, share often' approach."

As the former constituents of the article — data, tables, figures, reference lists and so on — fracture and dissolve into the fluid Web, they will leave behind the core of the article: the story. Authors will create these stories like blog posts: lean, fast and heavily reliant on the published tools, data and analyses of the entire community. These stories will carry forward the synthesis and narrative functions of the traditional article. Like all decoupled journal products, they will be openly available, redundantly archived, easily attributable and enveloped in an open network of links, comments, citations, annotations, discussions, saves and other interactions.

Certification. Authors will select from among diverse paths of certification for their research products. For specialized products such as daily lab results, no certification is needed — use of these products will be limited to narrow subfields in which researchers are qualified and motivated to consume unfiltered results[5]. For products aimed at broader audiences, authors will rely on algorithmic filters to flag their work to others. *The Journal of Digital Humanities*, for example, does not take submissions; rather, it highlights the best content already published online, often pulling from relatively obscure blogs and web pages. Importantly, this selection process relies heavily on altmetrics (such as number of page views, tweets and trackbacks) as a first-pass filter before manual curation.

Similarly, forums such as MathOverflow use community-bestowed votes to promote or demote content. Many MathOverflow posts have been cited in the peer-reviewed mathematics literature, and some mathematicians have begun to report high MathOverflow reputation scores as evidence of scholarly standing.

Qualitative peer review will move into the open and become yet another public scholarly product — complete with its own altmetrics, citations and even other peer reviews. The reviewer will metamorphose from gatekeeper to interlocutor and collaborator.

Start-up services such as Rubriq already offer 'standalone peer review', sold to authors as a tool to improve their work and help it to find its best audience. As peer review is unhitched from the creaky machinery of paper-native publishing, many more start-ups will offer other innovative variations. Some may specialize in statistical review or copy editing, or emphasize a quick turnaround. Many will offer automated review derived from altmetrics, while others will leverage existing journal brands.

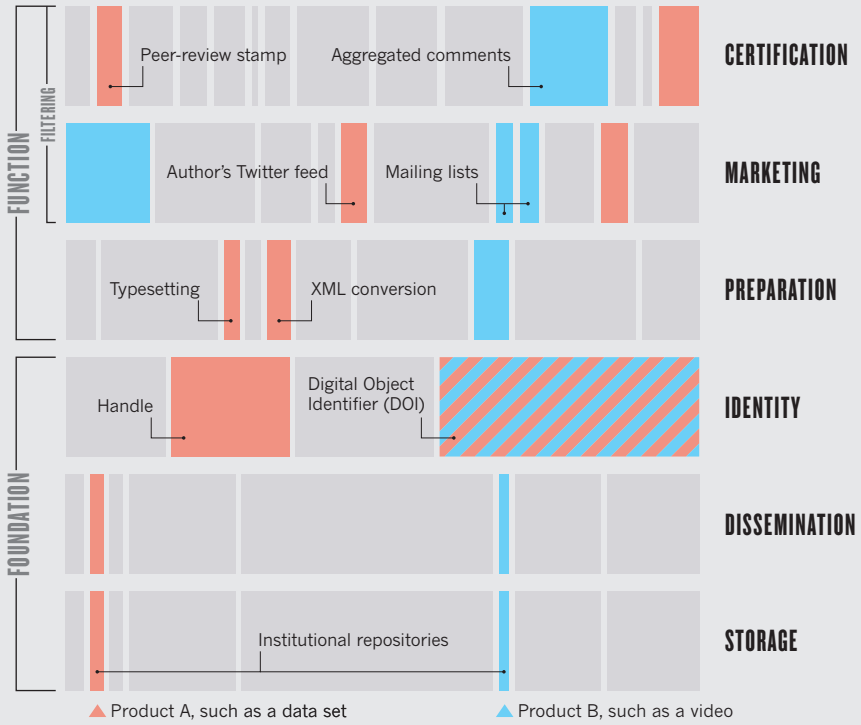
Authors will select the best services for each product and purpose, pinning certifications to their products' web pages like county-fair ribbons. For instance, an article might receive a 'B+ overall significance' badge from *Nature*, a 'statistically rigorous' badge from a specialist statistical review service and a 'highly read' or 'highly influential' badge from an altmetrics aggregator.

Filtering. To channel this torrent of Web-native scholarship, readers will rely on personalized recommendation engines to aggregate both altmetrics and qualitative badges, producing a bespoke, curated stream. Imagine every morning getting an e-mail from your personal filtering service with the five most important things for you to read that day. Content from your narrowest subfield — say, the evolution of toxins in Scorpaeniformes (an order of ray-finned fish) — is relatively unfiltered, resembling a simple RSS feed. At this level of detail,

SOURCE: REF. 3

RECONSTRUCTING PUBLISHING

In future, authors and funders will select services that are best for a given research product (for example, A and B). The product will be stored in and disseminated from an institutional repository, and given a unique identifier. Most traditional journal functions will be performed independently by an ecosystem of service providers. Some examples of services are labelled below.



information scarcity is as much a problem as information overload, and you want to stay aware of nearly everything your closest collaborators and competitors do. As you move out in scope (to venomous fish, then fish, then marine biology, for example), the content becomes increasingly filtered: you see only products that have been read, discussed, saved, cited and recommended by the other narrow subdisciplinary communities that are (like you) reading everything in their specific areas. These recommendations constantly improve as the system learns from your interactions (as Google search does already), but also allow for serendipity by delivering a customizable percentage of unexpected content from outside your narrow circles. Given such a system, why would you ever bother reading another journal?

Reward. The reward structure of scholarship will change too. Built on tools such as ImpactStory (of which I am a co-founder), scholars' CVs will transform from static lists of formal products to real-time displays of transparent, normalized metrics that reveal diverse impacts. Funders, ever eager to maximize value for their money, will demand more Web-native scholarly production. The US National Science Foundation, for example, has already begun requesting "products" rather than "articles" in the biographical-sketch section of grant applications[6]. The widespread availability of open altmetrics will fuel this trend, as progressive scholars begin to tell — and to be rewarded for — compelling, data-driven stories about the impact of their diverse products.

Tenure and hiring committees will adapt, too, with growing urgency. Ultimately, science evaluation will become something that is done scientifically, exchanging arbitrary, biased, personal opinions for meaningful distillations of entire communities' assessments. We can start to imagine the academic department as a sports team, full of complementary positions (theorists, methodologists, educators, public communicators, grant writers and so on). Coming years will see evaluators playing an academic version of *Moneyball* (the statistical approach to US baseball): instead of trying to field teams of identical superstars, we will leverage nuanced impact data to build teams of specialists who add up to more than the sum of their parts. We will wonder at how meaningful evaluation could have been accomplished in the slow, coarse, paper-native world of the journal.

The peer-reviewed journal is the product of some 350 years of selective breeding, a thoroughbred exquisitely tuned to wring the best possible performance from the paper-based world of its creation. But the Web has irrevocably changed our information environment — it is no longer the habitat the journal evolved in.

The next decade will see a dramatic dieback in journals and a Cambrian explosion of communication species evolved for the Web- and network-centric world. Most of these will be failures. We see this already in the dozens of 'online journal clubs' and 'Facebooks for science'; many are virtual ghost towns. But critics of these early failed systems miss the point. In the early twentieth century, there were more than 100 car makers; by mid-century, there were barely a dozen. But they were the dozen who had got it right. Similarly, we can expect the intense selection pressure generated by scholars' hunger for relevant information to drive rapid evolution in systems for certification, recommendation and assessment.

CONCERNS

The new era of scholarly communication will pose challenges. One significant concern is the ease of 'gaming' the algorithms we will rely on for filtering. What is to keep me from recruiting a few friends or students to tweet, read and save my new article, for example? Several companies have already emerged to sell social media follows and likes. Will I need to spend grant money on this sort of payola to compete?

Of course, any important metric will be gamed, just as citations are[7]. But altmetrics are not the easy target they may seem, thanks to their diversity and volume. Gaming a single metric may be easy; tampering with dozens is not. Products with high numbers of tweets but no downloads, for instance, will stand out. And not just to humans: given enough data, machines turn out to be excellent at finding discrepancies in standard usage patterns. Credit-card companies, for instance, use 'algorithmic forensics' to nab identity thieves. The powerful pattern detection of constantly evolving algorithms also maintains the robustness of Google search results and online advertising click counts — both of which are heavily gamed by highly motivated hackers[8]. These defences work in scholarly contexts, too: the Social Sciences Research Network and the Public Library of Science have already effectively used algorithmic shields around download counts[9] (see go.nature.com/binjfk).

A second criticism is that the very idea of quantifying scientific impact is misguided. This really will not do. We scientists routinely search out numerical data to explain everything from subatomic physics to the appreciation of Mozart; we cannot then insist that our cogitations are uniquely exempt. The ultimate judge of scientific quality is the scientific community; its judgements are expressed in actions and these actions may be measured. The only thing to do is to find good measures to replace the slow, clumsy and misleading ones we rely on today. The great migration of scholarship to the Web promises to help us to do this.

A final concern is that in seeking the wisdom of crowds, we will end up governed

by the madness of mobs. Will a crowdsourced scholarship be dominated by provocative pap that fills without nourishing?

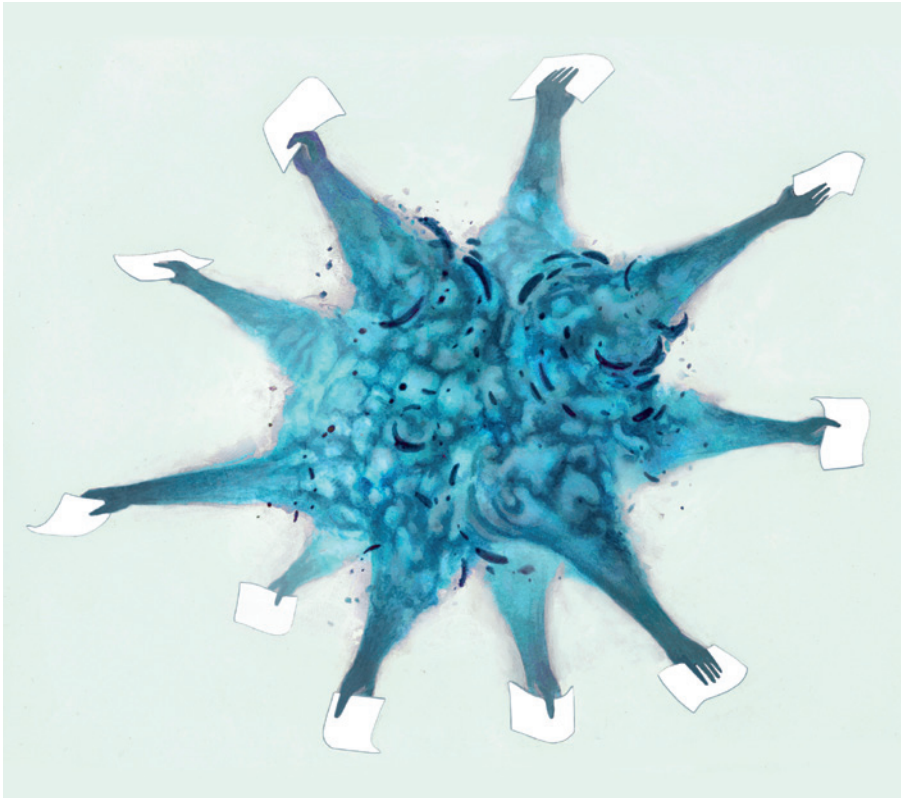
Here we must recall first that scholarship has always been a community enterprise, driven by building consensus among experts[10]. So the question is not 'Should we crowdsource?' but 'How should we crowdsource?'. Second, we must dispose of the straw-man argument that hundreds of uninformed readers' opinions will count for more than one Fields medallist's recommendation. Authority and expertise are central in the Web era as they were in the journal era. The difference is that whereas the paper-based system used subjective criteria to identify authoritative voices, the Web-based one assesses authority recursively from the entire community.

We now have a unique opportunity as scholars to guide the evolution of our tools in directions that honour our values and benefit our communities. Here's what to do. First, try new things: publish new kinds of products, share them in new places and brag about them using new metrics. Intellectual playfulness is a core scholarly virtue. Second, take risks (another scholarly virtue): publishing more papers may be safe, but scholars who establish early leadership in Web-native production will be ahead of the curve as these genres become dominant. Finally, resist the urge to cling to the trappings of scientific excellence rather than excellence itself. 'Publication' is just one mode of making public and one way of validating scholarly excellence. It is time to embrace the Web's power to disseminate and filter scholarship more broadly and meaningfully. Welcome to the next era of scholarly communication. ■

Jason Priem *is an information scientist at the University of North Carolina at Chapel Hill, North Carolina, USA, and co-founder of ImpactStory.*
e-mail: jason@impactstory.org

1. Priem, J., Costello, K. & Dzuba, T. *figshare* http://doi.org/kvx (2012).
2. Cronin, B. *J. Inform. Sci.* **27,** 1–7 (2001).
3. Priem, J. & Hemminger, B. H. *Front. Comput. Neurosci.* **6,** 19 (2012).
4. Smith, J. W. T. *Learn. Publ.* **12,** 79–91 (1999).
5. Esposito, J. J. *J. Electron. Publ.* http://dx.doi.org/10.3998/3336451.0011.203 (2008).
6. Piwowar, H. *Nature* **493,** 159 (2013).
7. Wilhite, A. W. & Fong, E. A. *Science* **335,** 542–543 (2012).
8. Priem, J. & Hemminger, B. M. *First Monday* (5 July 2010).
9. Edelman, B. G. & Larkin, I. Harvard Business School NOM Unit Working Paper No. 09-096 (2009).
10. Kuhn, T. S. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1962).

The author declares competing financial interests: see go.nature.com/szgqx9 for details.



BRENDAN MONROE

A fool's errand

Objections to the Creative Commons attribution licence are straw men raised by parties who want open access to be as closed as possible, warns **John Wilbanks**.

Copyright licensing is a topic usually left to law review articles, or obscure terms of service on websites, or agreements between publishers and libraries. But it is an essential element of the move towards open access — the free, immediate online availability of scholarly articles coupled with the right to use them fully in the digital environment.

An article that is free to read is not necessarily open for all uses — often, it cannot be reused for text mining or in derivative works, for example. The permitted uses depend on the copyright licence used by the author.

In my view, for an article to be considered truely open access, it has to meet the widely accepted definition in the Budapest Open Access Initiative — a set of recommendations laid out by leaders of the open-access movement in 2001. That is, users must be able to "read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

Traditional publishing licences tend to place restrictions on at least one of these uses, and it isn't easy for a reader to figure out what those are. If the reader is a computer, as is more and more prevalent, the restrictions are a spanner in the works.

The use of the Creative Commons attribution licence (CC-BY) fulfils the community definition of open access and avoids a future morass of articles with murky legal provenance and concomitant unclear reuse possibilities. CC-BY was launched in 2002, 2 years before I started a 7-year stint as head of science initiatives at Creative Commons in