

The New Peer Review

Isaac S. Kohane¹, Russ B. Altman² ¹Children's Hospital Informatics Program
Children's Hospital and Harvard Medical School, isaac_kohane@harvard.edu
²Stanford Medical Informatics, Stanford University, altman@smi.Stanford.edu

Public Access to Scientific Data

It is widely recognized that the Internet has fundamentally changed the dynamics of publication, and in particular, it is clear that there is no effective way to control the release of any web-based publication. The scientific and lay literature is now accessible to the public with unprecedented ease. Recent proposals to start a life sciences online repository of preprints highlights the trend towards "publish first, review later" that seems to be emerging. Does this mean that the peer review process is dead? It certainly suggests that there is a need for a change in how the process works. We discuss currently available technologies to enable the implementation of new, distributed peer review process benefiting multiple user communities.

In its most narrow application, the concept of peer review refers to the process of having qualified individuals read and evaluate manuscripts before they are released for general consumption by a larger community. The most consistent use of peer review has been in academic publishing, although other arenas (such as legal publishing) have found use for it as well. There are multiple benefits of this process. First, there are some assurances that the contents of the manuscript have been carefully assessed and are valid (with respect to some--often unstated--set of guidelines about novelty, credibility, reproducibility or other features) so that those who can not spend much time reading the manuscripts can expect that others have looked at them carefully and evaluated the contents. Second, the authors benefit from the knowledge that their colleagues have evaluated their work, labeled it as having a certain level of quality, and perhaps pointed out problems or errors that can be addressed much more easily before the release to a larger group. Third, publishers benefit because they can make claims about the quality of the contents of their publications in attempts to sell them to interested parties. The primary cost of a peer review system is that individuals must accept some "civic responsibility" to spend time carefully reading manuscripts on behalf of the interested community at large. Thus, the peer review system has worked for many decades, and seems to be a cost-effective mechanism for increasing the overall quality of the literature, and preventing (or making more difficult) the perpetuation of poor quality publications. It has also been an *ad hoc* method for controlling the release of publications.

The following scenario occurs with increasing frequency: A patient with type I insulin dependent diabetes mellitus (IDDM), for example, will search the web for information regarding novel therapies for IDDM. Today, she will find herself immediately awash in hundreds if not thousands of potentially relevant documents. If she has the patience, she might skim through these documents and focus on those which she considers the most authoritative, complete and up-to-date. Just what constitutes authoritativeness will depend a lot on her particular preferences. She might prefer documents that were written by her personal endocrinologist. She might also find acceptable, but to a lesser degree, documents written or endorsed by endocrinologists who had worked with her endocrinologist or who had published papers with her endocrinologist. Also acceptable, but perhaps of lesser-perceived importance would be the documents sanctioned by endocrinologists without any documented professional association with her endocrinologist. Alternatively, it could be, for a particular patient, that the most authoritative sanctioning is provided by individuals with type 1 diabetes mellitus rather than expert clinicians, perhaps because of the expertise borne of first hand experience. Similarly, a researcher in the molecular biology of type I IDDM may find an interesting reference on the tissue-specific activity of a glucose transporter in an obscure journal. If the findings are surprising or at odds with her understanding of transporter biology, the researcher will be inclined to trust the conclusions of the article much more if they were written by a respected member of her peer community, or reviewed by such a member. Or, if the authors and reviewers were not recognizable, some metric of their relationship with respect to this field would help determine how much effort the researcher would invest in verifying, or building upon the reported findings. Subsequently, she might wish to comment upon the reliability of the reported findings and have such comments made available broadly to aid other researchers.

This example illustrates the components of a growing problem that all consumers of medical information are faced with: 1) Documents on the web are rarely annotated and only for articles published in the traditional medical journals is there any significant review process. Yet the latter articles constitute only a small minority of the medical information offerings

currently available on the web. 2) Different communities and individuals might find different sources of authoritativeness. That is, the perception of what constitutes a reliable, authoritative and disinterested peer group will vary greatly by reader population. For some medical topics the boundaries between these populations of medical information consumers will be distinct and for others much less so. 3) Even for nominally peer-reviewed scientific literature, the issue of authoritativeness often is unresolved or unclear for even a well-informed reader. That is, the same problems can exist for the scientific community and scientific publishing as for health care consumers. Interdisciplinary work can be judged by representatives from numerous disciplines, and the "stamp of approval" that is desired by an individual scientist may be quite different from that desired by a colleague. For example, peer-generated reviews in biocomputing can have a very different flavor when the peers are biologists than when the peers are computer scientists.

Ideally, there should exist mechanisms by which any community of information consumers or providers should be able to create a shared set of trusted annotations or sanctions upon any collection of documents present on the web. Far more rapidly than had been anticipated, and for reasons having little to do with medical publication, these mechanisms are already available and in the public domain. We outline here the nature of these mechanisms and some of the considerations that will inform their use. In doing so, we suggest a recasting of our commonly held notions of peer review.

In 1997, Silberg et al. noted, "The problem is not too little information, but too much, vast chunks of it incomplete, misleading, or inaccurate, and not only in the medical arena. The Net—and especially the Web—has the potential to become the world's largest vanity press. It is a medium in which anyone with a computer can serve simultaneously as author, editor, and publisher and can fill any or all of these roles anonymously if he or she so chooses. In such an environment, novices and savvy Internet users alike can have trouble distinguishing the wheat from the chaff, the useful from the harmful." ¹

Consider the effect of web publishing on medicine. Patients regularly approach their clinicians with printouts of web pages some containing information highly relevant to their condition that the clinician might not have been aware of and some potentially misleading and dangerous.^{2, 3} Similarly, clinicians typically find it impossible to remain current with all but a small fraction of the relevant, available literature⁴ often leading to outmoded and ineffective patient management. At the same time, the process of traditional peer-review is perceived by some as a bottleneck for dissemination of useful information.

This bottleneck is increasingly problematic in view of the new sources of medical information made available by web technologies. These include alternative medicine sites, patient support news groups and mailing lists, disease-specific sites, and sites sponsored by employers, and insurers and all the other parties involved in the delivery of healthcare, each with their own agenda. *A priori*, all these sources are equal citizens of the community of medical information sources on the Internet. Therefore, even when consumers of medical information are able to retrieve useful information relevant to a medical condition, they may remain uncertain and/or unaware of the sanctioning of the information by sources they trust and respect. Also, the accountability of the authors/editors, or the authenticity of the provenance of these documents remains unknown. The volume of information available to patients and providers alike is so large that it immediately overwhelms, by several orders of magnitude, the resources available through traditional peer review. Additionally, there are large classes of medical information which would rarely be considered the province of the traditional peer review process such as patient-oriented "how-to" manuals.

What then can substitute for the inadequate resources of our current mechanisms of peer review and provide information consumers with information they can trust? Among the desiderata for any substitute mechanism: 1) Automating the means by which information consumers can obtain a quick and reliable assessment of the authoritativeness of any document they retrieve. 2) Providing a mechanism to distribute the editorial function to large classes of users/editors throughout the Internet without compromising the first goal. 3) Additionally, because of the international reach of the Web and the breadth of its readership, any editorial mechanism must satisfactorily address the concerns arising from multiple cultural, geographic or religious viewpoints.

Shape of a solution

There are already several informal mechanisms for reviewing content on the web that take advantage of its decentralized and distributed nature. There are large numbers of newsgroups and mailing lists devoted to specific topics. However, understanding which opinions or reviews articulated in these venues are authoritative requires a significant investment of time and effort. For instance, for each community which these newsgroups constitute, there are several roles which each participant takes. Although the majority may be passive readers, there will be individuals who become prominent or notorious within the community. Some will be seen as gadflies, some as careful and thoughtful authorities, and others as cranks or eccentrics. However, there is little formalization of these roles and there is no way other

than through time and experience for a participant to determine which voices/notes are authoritative and/or or consistent with the preferences of the participant. It is also difficult to interact with these newsgroups at a low frequency, since the shared assumptions and conclusions of the community are often not accessible to a novice who has an acute need for information.

This suggests the need for a distributed and decentralized mechanism of formalized sanctioning of documents or collections of documents on the web. That is, groups or populations with common interests or shared points of view must be able to generate a consistent set of reviews across a set of documents across the entire world wide web. Additionally, this distributed mechanism must allow rapid, concise and trusted communication of this information. This notion is consistent with the overall growth and governance of the Internet, which has depended on local autonomy bound by a common set of protocols. The new idea here is that the egalitarian principles emerging from the web be embraced: peer review becomes not just the activity of a small group of scientists, but becomes a mechanism for any group of "peers" who can identify and specify criteria for membership can add value to the web by labeling its contents.

The tools are here

What would it take implement a widely distributed peer review mechanism? In addition to the social structures and compacts required (touched upon below), there is a required technological infrastructure. Fortunately, at the technological level the requirements are relatively modest and, furthermore, many of the solutions have already been developed. As is often the case, the imperatives of Internet commerce have resulted in a collection of technologies that were not designed to meet the needs of peer review but are particularly useful. These technologies include: a mechanism for irrefutable annotations, a formal syntax for these annotations, and search mechanisms that exploit preferences and search strategies of communities to customize the information returned to the individual. These are briefly described here.

A web document labeling mechanism, the Platform for Internet Content Selection (PICS)⁵ standard promoted by the worldwide web consortium was originally designed to meet the need of families that were concerned that their children might be exposed to inappropriate content while "surfing" the web. PICS has been widely adopted for that use. PICS allows the annotations of web pages with a variety of rating scales. It designed to be able to directly annotate the page it is describing. Or, as is often the case, if the annotator is not given access or permission to directly modify a particular web document, PICS allows the

annotation to be entered into a "label sever." The label server can subsequently return the rating corresponding to the address (URL) for any annotated web document. PICS has already been adopted by a wide range of content providers, Internet Service Providers and manufacturers of web servers. All these capabilities are applicable to the annotations that could be added to reviewed documents. The structured annotations can describe both the content of the documents as well as specify the value of those documents along several dimensions such as relevance, completeness, and authoritativeness.

How then can such annotations provide consumers of information with documents that they will find relevant according to their own context or beliefs? There are now a variety of search engines that allow some degree of personalization. Some of these employ manual methods in which the user explicitly defines preferences over a particular set of content subjects, authors, or time of web "publication". More successful have been efforts based on choices the user has made in prior search episodes, or by comparing the set of choices the user has made to the similar choices or preferences of other users (e.g. collaborative filtering⁶). Most recently, topological analysis programs seem to be particularly promising, especially in their most recent implementation at IBM as the Clever system⁷. The last programs exploit the insight that popular and authoritative web pages will include references to authoritative and useful directory sites (compilations of links to web sites) and vice versa. These techniques can be used in conjunction with the content annotations and sanctioning enabled by PICS to generate collections of documents that are rated as highly relevant, accurate and authoritative to a particular community. Diffusion of distributed annotations will encourage the sharing of search profiles that are particularly well suited to the interests of specific communities. It is possible that a mixed strategy in which certain features are mandatory (e.g. the number and credentials of reviewers) are mixed with features that are desirable (e.g. the occurrence of particular phrases).

There are many commercial and other motivations for organizations or individuals to try to "game" such an annotation system to direct consumers of medical information away or to particular documents. For this reason, the annotations made must be done in a manner that irrefutably authenticates the authors. Again, for motivations unrelated to medicine, the interests of Internet-borne commerce have promoted the use of "strong" cryptographic schemes such as Public Key cryptography (PKC) that permit the irrefutable authentication of electronic documents such as documents on the web. Also, PICS supports the use of PKC to authenticate its annotations thereby providing a mechanism to meet the

aforementioned requirement of irrepudiability of medical peer review annotations.

Challenges

In the above discussion, we have conveniently avoided some thorny problems in implementing such distributed peer review mechanisms. To start with, there is no apparent consensus as to the dimensions along which one might annotate/review a document, nor the vocabulary or terminology with which one would assign values across these dimensions. For example, one person might describe a document as “irrelevant” along the Relevance dimension and another person would describe it as “meaningless” along the “Applicability” dimension. A priori, there is no way for a search engine to “understand” just how close these two annotations are. Certainly, homogeneity of annotations can be enforced through the design of appropriately constraining annotation tools and the adoption of standardized vocabularies for peer review. The vocabularies that seem most necessary are those for describing *relevance* to the peer group of interest, *importance* of findings, as well as those describing the credentials of the reviewers. In scientific publication, reviewers are currently credentialed in an ad hoc way by an editor who knows enough about the reviewer to feel comfortable requesting a review. Often editors rely on the fact that the manuscript in question refers to the work of the potential reviewer or is the subject of other manuscripts by the potential reviewer. It is not hard to imagine automated systems for performing this type of credentialing, or at least for providing some support for a human decision maker. We suspect that, like many other phenomena of the Internet, consensus on these terminologies may emerge *de facto* as one particular set of conventions becomes particularly popular.

Another significant challenge is that the public key cryptography infrastructure required to support the widespread authentication of all annotations is nascent and has only begun routine use for financial transactions. The infrastructure required for any individual to reliably and safely access and use their private keys is still incomplete but, again, rapidly developing to meet the exigencies of Internet commerce.

Social Structures

The Internet has generated multiple communities and social structures that were not anticipated by the architects of the Internet and web. We do not claim clairvoyance on how the opportunities for peer review will become realized. However, a few questions might help us frame how we view developments in this domain. Where will the annotators come from? Currently, reviewers (or annotators) serve their community from a sense of civic responsibility, and

for reasons of self-promotion. Mechanisms will be required to track the participation of “peers” in this process, and to allocate credit for participation in the system at an appropriate level, but also with safeguards to avoid over-participation or under-participation.

What combination of self-appointed, nominated, or hired annotators will provide the critical mass of reviewers? The current system of peer reviews employs an “editor” who credentials the reviewers manually. Initially, this model can be used, but the more powerful model would involve some sort of automatic credentialing process. How do editors determine who is appropriately positioned to perform a review? Probably from a combination of valid criteria and some that result from personal biases. For some forums, specific disciplinary training will no doubt be required, for others the criteria are much less exclusive.

What mix of business-driven annotations or purely volunteer-initiated annotations will prevail? This will again be a function of peer group and its goals. In science, volunteerism will most likely be the most valued. There will likely be a bootstrapping process required for creating these peer groups. For example, the IOM could be the first for medical science to provide an operational definition for journals and conferences that exist, and then bootstrapping from that, with occasional mid-course corrections by consensus-building within peer communities. Beyond volunteerism, there are several conceivable incentives that could be used to drive wide-scale review process. Information systems with particularly good reputations might pay practitioners to annotate healthcare websites as a means of marketing their “brand.” Companies interested in promoting cheaper and more effective products will annotate sites accordingly. In the field of medicine, pharmaceutical and biotechnology companies may promote labeling initiatives just as they currently support scientific symposia and publications, and with similar issues about the independence and impartiality of such efforts. Each of these incentives, however, will either be explicitly made part of the annotation or regardless, the authoritativeness of different schemes will become eventually known and formalized among various peer communities.

To what degree will different peer-review communities use the annotations of other communities in their own searches? It is likely that the public will demand technologies that allow the opinions of more than one peer group to be mixed in order to create personalized ratings of documents. A publication on the biology of cancer that wins approval from both the basic science community and the medical oncology community is likely to be of

great importance to someone following developments in cancer.

Will fear of retribution or litigation stifle the use of irreputably signed annotations? Academic peer review has, for the most part, used anonymity to guarantee the freedom to criticize. However, the anonymity is only partial: the author of a publication may not know the reviewer, but the moderating editor does know the identity. This acts as an important safeguard for abuse of anonymity, and protects authors from random, unsubstantiated criticism. It is therefore unlikely that moderators can be removed entirely from the process, at least for academic review. The selection of moderators can be by vote of the peer community, either directly or by rating the quality of reviews. It is also possible that some of the functions of moderators can be assumed by software programs. For example, the volume of reviews and the overall positive or negative tone of reviews can be monitored automatically (with the availability of appropriate controlled terminologies for review), and reviewers can be audited (securely) to be sure that they do not abuse their positions.

How will the traditional peer-review process of journals be affected by the new peer review? Current academic publication review is a "special case" of the new peer review, and can be supported in exactly the present form with the new technology. Editors selected by the community (or a publisher) are given the right to select reviewers, reviewers provide reviews that are anonymous to the author (but not the editor), and the editor makes a decision about "branding" the publication after a suitable round of revisions in response to the reviewers. The branding is guaranteed with authentication technologies that guarantee that others can not imitate this brand technically, and thus the status of the publication on the web is guaranteed. The new peer review model, however, does admit of new ways for peer review. Papers can be published without the traditional process, and can be branded by ad hoc groups who define their own criteria. In addition, papers that are rejected by standard peer review procedures can nonetheless be published--even with the negative reviews--in order to focus the attention of a field on new and controversial ideas.

In the end, will consumers of information benefit or will they just be confused? With generalization of the new peer review methodologies, there is likely to be a proliferation of methods to compute authoritativeness and other metrics based upon the wide varieties of review annotations. We expect that there will be a variety of search engines or "agents" each more or less "tunable" or customizable for a particular information consumer. This is analogous to the current range of information retrieval engines currently available for searching the Web, each with

different properties, sensitivities and specificities, and each preferred by somewhat different user communities.

Conclusion

The new peer review is already occurring. In newsgroups, on mailings lists, ad hoc communities of information consumers, information providers. Within medicine, for example, patients, care providers, and other participants in the modern ecology of medical information are describing, summarizing, rating, applauding, recommending and condemning websites, newsgroup threads and printed paper articles. At this time, the process is *ad hoc* and the efforts of one community are not easily communicated succinctly and reliably. We have described here some of the kinds of tools, available now, that will soon enable the new peer review process to become scaleable to the full dimensions of the Internet. This process is happening, and the scientific community need only to decide whether to participate in it (and live with whatever is developed) or ignore it.

References

1. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of information on the Internet: Caveant lector et viewer--Let the reader and viewer beware [editorial] [see comments]. *Jama* 1997;277(15):1244-5.
2. Kassirer JP. The next transformation in the delivery of health care [editorial] [see comments]. *N Engl J Med* 1995;332(1):52-4.
3. McClung HJ, Murray RD, Heitlinger LA. The Internet as a Source for Current Patient Information. *Pediatrics* 1998;101(6):e2.
4. Hafner K. Can the internet cure the common cold. *The New York Times* 1998;Sect. E1.
5. Krauskopf T, Miller J, Resnick P, Treese W. PICS Label Distribution Label Syntax and Communications Protocols. W3C Recommendation: World Wide Web Consortium; 1996 October 31st, 1996. Report No.: REC-PICS-labels-961031.
6. Moulinier I. A Framework for Comparing Text Categorization Approaches. In: Hearst M, Hirsh H, editors. *AAAI Spring Symposium on Machine Learning in Information Access*; Palo Alto, CA; 1996.
7. Kleinberg J. Authoritative sources in a hyperlinked environment. In: 9th ACM-SIAM Symposium on Discrete Algorithms; San Francisco, California; 1998. p. 668-77.