An analysis of the present system of scientific publishing:

What's wrong and where to go from here

Dov Greenbaum[1], Joanna Lim[2] & Mark Gerstein[2,3]

[1]Department of Genetics,
[2]Department of Molecular Biophysics & Biochemistry
[3]Department of Computer Science
Yale University
P.O. Box 208114
New Haven, CT 06520-8114, USA.

Introduction

As recounted in Professor Guédon's work, "In Oldenburg's Long Shadow" scholarly journals where initially founded in order to preclude intellectual property disputes. The Philosophical Transactions of the Royal Society of London, first published in 1665, was to be a register of scientific ideas, and *the* arbiter of what was science; as a secondary goal, it would also disseminate scientific ideas[1]. Henry Oldenburg, inspired by Francis Bacon's *Novum Organum*, was the pioneer behind the journal, and the concept of peer review; Oldenburg would have articles sent to experts to review them prior to their inclusion in the *Phil Trans* [2]. The concept of peer review was later cemented as a requirement for publication almost 100 years later when the editorial process of the journal was taken over by the Royal Society [3]. These notions of wide dissemination and peer review have subsequently become hallmarks of scientific journal publishing. In addition to these, there are other objectives of scholarly journals including: the creation of archives for scientific data, a system to prevent plagiarism of other's works, and a sort of currency for scientists, demarcating their level of prestige as a function of the number and quality of the articles published[4]. Journals as we know them are becoming less important in the dissemination of scientific information (they are used more as a currency representing scientific ability rather than their initial purpose of information dissemination); better vehicles of communication, (e.g., more able to conform to the now diverse levels of collaborations that are the norm in present-day scientific research) are required [5]. Publishing scientific articles in general, in its present form, is slow, inefficient, costly and sometimes even a hindrance to research, and the flow of information[6]. In addition the paper, as opposed to digital medium used presently is

"difficult to produce, difficult to distribute, difficult to archive and difficult to duplicate."[7]

Problems with the Current System

`"Our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purposes" – Dr. Vannevar Bush, 1945[8].

Although there was no practical alternative in 1945 to the publication process, the internet presents an opportunity to reshape the scientific publication process. Still, the internet is only starting to make inroads into the methods of transmitting research, and much of the heretofore evolution of scientific information dissemination has resulted from a haphazard and undirected progression of research methodologies. For example, the web now allows researchers the ability to present much of their data in forums other than journals, such as private websites, pre-prints, databases, newsletters, reports, working papers, theses, conference proceedings. While not peer reviewed, this 'gray information/literature'[9] is gaining validity and importance in research as a source of scientific information. For example, the US departments of Energy and Defense, as well as other governmental agencies currently have well over 100,000 scientific and technical non-peer reviewed reports which they have integrated into a central repository: the GrayLit Network[10].

Nevertheless, to achieve a true paradigm shift in scientific publishing, we need a directed evolutionary event (Contrast with Ann Okerson's position[11]), a total and global unified

revamping of the system from the ground up. Although two-thirds of all journals already publish online[12], there are many issues with the present system of peer review academic journals, problems that cannot be solved by simply making PDF copies of the journal articles available online: "An electronic document is not {simply} the electronic version of a traditional paper document …{Rather it is} a document comprising a variety of different types of information presentations that are brought together by an author in order to present a comprehensive scientific argument …" [13].

This paper will examine some of the issues with the present system of scientific publication - such as rising costs, poor peer review and slow dissemination of information - and present a possible alternative to the present situation. The discussion is not novel, many groups have already attempted to tackle the issue and reform the world of scientific publishing and data dissemination (See for example: The Scholars' Forum[14], SPARC[15], or the 'Tempe Principles'[16]).

Issues with the Present Publication System

Formats

With the advent of high throughput experimental methodologies, molecular biology has become, like many other sciences, data intensive (See J. Rumble [17] for a list of examples). Consequently, experimental results more often than not will not fit within the rigid guidelines of journal formats, and very often, important data tables, if they are included, are regulated to on-line supplementary tables or associated websites.

Moreover, in their present state, journal articles are not easily parsed for data mining given the lack of any standardized formatting or ontology[18]. In addition the universal rigid format presently used in journals (e.g. abstract, introduction, methods, results, discussion and conclusion) may not be appropriate for the presentation of web tools or databases and future research methods and results.

Gray information

In addition, many laboratories choose to present their data on their own websites (irrespective of any particular publication), providing access to raw, unverified experimental data. This information is a rich source of cutting edge data, and its growing usage as a research tool blurs the boundaries between formal and informal publications[19]. These databases are slowly encroaching on the journals' position as disseminators of information. Still, as opposed to journal articles that are centrally indexed, it becomes very difficult to keep track of and locate new results that are published in these forums. While prior to this explosion of data, researchers could easily contact authors for additional individual data sets, with the advent of bioinformatics and the need to sift through and analyze multiple huge datasets, all of the data must be easily accessible in real time[20, 21].

Peer review

The peer review process, which is supposed to provide verification for the information found in scientific journals, and thus differentiate journal based information from the above mentioned gray information is under attack. Both Science and Nature have recently taken flack for publishing questionable material[22]. For the most part, research

scientists, and their students make up the cadre of peer reviewers, and with increasing pressure for these scientists to produce, there is less time and incentive to review articles thoroughly, and a greater chance of bad science slipping through the cracks.

Cost of acquiring journal articles

Journals are also becoming less available to the masses due to high costs. Journal prices are rising, significantly faster than inflation, and many are no longer within the price range of the average university library. The Association of Research libraries claims that the price for journals subscriptions skyrocketed 207% from 1986 to 1999[23, 24]. In conjunction with budgetary cutbacks, many libraries are forced to cancel several of their subscriptions[25]. As a result most refereed journals are not available to the average researcher[26]. The irony of the situation is that the universities are funding research, yet they can not afford to buy the results back from the journals [27]. Even the electronic versions of journals, which were supposed to be cheaper than print subscriptions, are just as unaffordable[27] (The high prices here have been attributed to the cost of customer support, as well as the continuing fixed costs of editing[28]). Yet even with all the cutbacks and cancellations science, technology and medical (STM) publishing has been the fasted growing media sub sector for the last 15 years[29].

Even with this incredible growth, journal-publishing houses that maintain high prices may be pricing themselves out of the market, and as such should also be interested in reform. Recent research has shown that researchers preferentially read and cite articles that are made freely, or at least, easily available. Many are not willing to pay for

expensive journals, nor are they willing to seek out printed copies of journals when they can access other journals effortlessly and freely online [30, 31].

Journals ought to be free to the scientific community. Still, given that the PubMed/Mediline database was only made freely available to the public in 1997[32], the concept of providing totally free access to all information may be somewhat premature. Even so, there are many groups presently working towards providing free access to scientific journals. These include: Pubmed Central [33 34], BioOne[35], the Public Library of Science[36], and the Budapest Open Access Initiative[37].

Too much information to be useful

The number of articles published annually has been doubling every decade or so for the last two hundred years[38]; there are, at present, approximately 20 thousand refereed journals producing in excess of two million articles each year[26]. Researchers cannot possibly, and surveys have shown that they do not, keep up with this deluge of data[39]- in fact, it has been found that they do not want to read the seemingly inexhaustible literature[40]. With this growing number of articles, it is becoming increasingly more difficult to effectively sift through the literature to find the desired information. Even with the growing desire, and the computing ability, to mine the literature for additional information[41-43], the incredible lack of uniformity within the literature in terms of ontologies and formats makes this method of research difficult to conduct.

Speed and biases in information transmission

The process of getting an article from submission to publication, especially in competitive fast-moving fields, is much too slow. With the fear of getting scooped by

their competitors, scientists are often publishing incomplete or partial research results so that they can stake their claim to potentially valuable research. Additionally, there is a general concern that too much power is held by the editors of journals and peer reviewers, such that their biases could potentially prevent the publication of important, novel, or avant-garde results.

Potential Alternative to the Present System

While only some of the concerns with the present system have been presented, it should be clear that Dr. Bush's statement [8], voiced over a half century ago, is all the more pertinent today. What is needed is a totally overhauled publishing structure. Below, we present an outline of what could be the next system of scientific dissemination. Following the presentation of a succinct framework, we flesh out some of the particulars and present some additional issues that need to be tackled.

Outline

We are not presenting a system similar to the present scheme where journals in print are also available online, rather a total and unmitigated shift from print to online; we envisage the following multi-tiered system: After completing a project, the researcher submits her paper to a web-based journal along with a standard reasonable submission fee to cover the initial costs of editing. The journal's editorial board decides whether the project and the paper fit their basic criteria for publication and, if so, the paper is uploaded to a limited access web site. Other researchers in the field who have registered for access to this site, and have expressed interest in the subject matter, are notified

automatically via email of the submission. Over the course of some flexible period of time, depending on the subject matter, other researchers can log in and evaluate the paper, posting their comments and suggestions; this online discussion is moderated by an editor assigned to the paper. Once this review period ends, the editor can decide, based on the comments, whether to accept the paper as is, request changes and send it back for another round of review, or reject it. Each draft of the article throughout the review process is saved and contains a unique identifier. Upon acceptance, the author is charged an additional fee to cover the costs of publication and archiving. The final paper, which should be immutable and authentictable[44], may be uploaded to the journal's website, but must be uploaded to a freely accessible archival web site, providing unlimited access to anyone.

The Journal

Historically journals have played many important and essential roles in the dissemination of information. In their simplest form they are archives of information; one can dig up ancient copies of journals in any well-equipped library to find data. In the pre-internet era they were the easiest way to distribute new information to the broadest possible audience; anyone who was interested in learning the most recent accomplishments in their field could flip through a copy of the appropriate journal and read a description of the research. Usually, the research was (and for the most part still is) presented in a common format which included an abstract, introduction, methods, results, discussion,

conclusion and references; readers knew where to look in the article for the information they needed.

Journals act as gatekeepers to the scientific archive, keeping out undeserving or plagiarized research. The fact that an article appears in a journal indicates that it has gone through some sort of peer review that had provided some sort of validation to the purpose, necessity and results of the research. The fixed costs of publishing a journal are thought to be a barrier to entry for journals that have not reached a level of public acceptance or academic stature. Journals also provide some sort of qualitative comparative measure to the research. The more prestigious the journal, the more important and conclusive the research is thought to be.

With the prospect of creating a long-term digital archive of all scientific data (as opposed to the present paper archive) it doesn't make economic sense for individual journals to maintain their own archives (See later for a discussion of the issues of maintaining a digital archive). Instead we envisage a much smaller yet important role for journals in our potential solution; As described, journals presently perform both a repository and an information service function[45]. In our proposal they would retain a portion of the service function, and spin off their repository functions. That is, they would retain only their most important and irreplaceable role as editors and facilitators of peer review. (Although some have claimed that the editorial process actually diminishes the value of an article[46].) Rather than having each journal maintain copies of their articles, a system

has to be developed to maintain an easily accessible archive that would promote interoperability that would allow for large scale and mining of scientific literature.

Journals should, though, maintain their banner on the top of their specific articles in the archive as the journal's name is somewhat indicative of the quality of the article.

We assume that many journals may decide to continue publishing online, still there should be a universally accepted framework that would demand that the articles be deposited in an archive shortly, if not immediately, after publication. Some journals might also choose to continue to publish paper versions of online articles, possibly for the small but persistent Luddite population. Journals might also publish smaller, single page, abstract-like versions of their online content in print journals; for example, the FASEB journal publishes short summary versions in print but longer articles online[47].

Nevertheless, research articles ought to be provided to the scientific public for free. Journals claim that providing free and unlimited access through a provider other than the journals to online articles will deplete an economically important source of revenue for the journals, could lead to loss of quality control, abuse of content, and will put too much control within a centralized organization, rather than what they claim is a more stable system where hundreds of journals provide independent access[48]. Additionally, the transfer and duplication of information from the journal to the archive could potentially corrupt the data[49]. Journals claim that they can maintain profits by instead of providing their information right away freely to the public, that they instead wait 6 months where

they can charge for access, after which they will provide the article for free on their website, where they can control and monitor access

We propose a more research friendly profit making approach: To prevent lost of profits, journals will retool their revenue mechanisms. One possible solution is to charge authors for the cost of editing. Given the general inelastic demand for publishing articles, journals should be able to charge enough to be profitable. Anyway, the authors will just pass the cost to their funding agencies and the costs should not limit the ability of a researcher to publish. Moreover, given that the economic system of publishing tends to favor those who pay, a system wherein the author is paying is a system that will reflect the goals of the author, i.e. broad dissemination[50] . Additionally, by not maintaining any archival functions, the journals do not have to fear that the copy that they submit to the archive will be corrupted through reproduction, instead, the journal should submit their copy immediately to the archive.

Peer Review

The peer review process, existing in its present form really only since World War II[51], has been coming under fire for many of its failings[52] for quite some time. Some of the issues with the peer review process include: (i) falsified data has gotten past reviewers[53];(ii) reviewers have been suspected of holding up the review process either out of spite or while they themselves published similar results[54]; (iii) plagiarism[55] ;(iv) sharing confidential data with others[56]; (v) researchers are overwhelmed by their reviewing responsibilities and either do not do a thorough job or do so very slowly; (vi)

the anonymity of the review process does not give the reviewer the feeling of accountability [51]; (Although contrast this with Steven Harnad's comments in [57]); (vii) the lack of credit given to the unpaid labor force of reviewers; (viii) reviewers are given too much power in (and their biases may be affect) the dissemination of scientific information; and (ix) the review process is a large portion of the cost of publishing costing anywhere between 500 and 1000 dollars per article[58].

However, with all of its faults, the peer review process is integral for scientific research. It provides assurance to the authors, general public and the publisher that the submitted work is of a minimum quality. At the very least, it provides a process wherein works are improved by the incorporation of outside ideas.

The transformation of scientific data from paper to the internet can help democratize the review process, make it more efficient, and more discriminating. The present peer review process requires the editors of a journal to select reviewers based on their perceived fields of expertise, contact these reviewers and request them to review a paper. Often reviewers are slow to respond and may not have the time or desire to review. We propose a system wherein reviewers would be notified automatically via email if a new paper was submitted in their field. Moreover, in addition to the present incentives to review, (e.g. the desire to keep bad science out of the field, or a feeling of responsibility) journals could provide monetary incentives to review in the form of some sort of credit towards the publication of the reviewer's next piece. In addition to providing an incentive, this method will also result in a situation wherein the more prestigious journals (where more

people would like to publish and would be more appreciative of the credit) will have more people reviewing the submissions, in essence, providing more substantiation for the work in better journals.

Addressing the issue of anonymity, reviewers will have to register to access these presubmission pieces, and their access to the papers will be logged, thus allowing for a paper trail in a case where a reviewer is suspected of stealing information. Moreover, authors of papers will no longer be held up by the procrastination of individual reviewers. The review process will be for a finite period of time, after which the editor for the piece will review the comments.

Of course there will be cases where the editor may feel that the paper is not garnering enough attention for a comprehensive review. At this point she may step in and actually assign reviewers for the piece or reject the piece outright. Still, as the success of sties such as eopinions.com shows, people are more than willing to give their opinion on anything.

This system also allows for the authors to collect a wide range of comments on their piece from a significantly larger audience; reviewers will not be limited to a small cadre of researchers that are selected by the journal, rather anyone can register and include their opinion.

Reviewers will also be able to increase their 'street cred', and the credit towards future publishing in the journal. Akin to the system already in place on amazon.com, readers of reviewers' comments will be able to evaluate the comments and note whether or not they were helpful, helping to highlight the important comments and weed out the inane comments often seen when the reviewer does not truly understand the paper. A reviewer who consistently presents strong comments will receive more credit for their review (bad reviewers could be barred from the forum), in essence also providing an incentive for people to put in well thought out comments.

The review process can also be simplified by requiring reviewers to stick to a specific syntax and format, answering a list of directed questions. Given the automation of the system there can be significant cost savings in this step of publishing.

Finally, to prevent frivolous submissions from overwhelming the reviewers, there can be some sort of automated check to determine an author's authors previous publication record, institutional affiliation , research grant status and other background information that can act as an automatic first level of discrimination to at least determine that the paper is of 'refereeable quality'. New authors could resort to alternate paths of entry, i.e. referrals from other credentialed authors[59].

Although it might be argued that such a peer reviewing system is faulty in that it relies on fellow authors volunteering to review articles instead of journals requesting experts in that field, this system rewards reviewers by giving them the opportunity to become

known to the journal, whether they are or are not already well-known for their research accomplishments. This system of peer review allows for a greater breadth of response to each article, allowing all kinds of perspectives, from many related to provide feedback and possibly even create future collaborations.

The Format:

One of the main strengths of our framework is the possibility of creating a homogenous body of scientific literature that will allow for thorough searching and data mining[48]. To this end it is imperative that a set of universal standards for the formatting of scientific articles be established. In addition it is also important to create a standardized language to describe the information contained within the articles[18, 60].

With all of the text of each article available online large scale literature searchers, similar to database searches, will allow users to integrate and incorporate disparate information for analyses. Large scale global searches will allow users to pick out key words or gene names from the entire body of scientific literature. To facilitate more powerful searches, we envision a standardization of formats and key words – similar to MESH terms in the NCBI's Entrez/Pubmed system[61].

Within the potentially unlimited extant of cyberspace, articles will expand and provide not only more information, but more information in a more efficient manner. One

potential way of setting a internet journal format is to have the data presented in multiple different layers; articles are accessed by a wide variety of readers (e.g. experts, non-experts and casual readers), all of which have different information requirements which could be satisfied by different layers of the article. (The concept of different layers within an article has been suggested by Dr. Paul Ginsparg, founder of the arXiv physics pre-print archive [7].) For example, the first layer might include the primary data, the information on which the article is based with little or no textual information, thus allowing experts to quickly scan and retrieve data. A second layer would provide more information regarding the material and methodology. The third layer would resemble a short article providing, succinctly the data, methods, and some discussion and conclusion. Finally a fourth layer might include information that might be necessary for the uninitiated reader, including a longer introduction, methods, discussion, conclusion and supplementary materials. While presently space limitations force authors to either leave out information or publish it as supplementary material, a wholly online format would allow researchers to incorporate all their data and textual information into the article.

In addition to the extra space an online format would allow authors and editors to integrate hyperlinks into the papers providing readers with access to further information on the subject at hand, both within the article itself, to other sites, gray information, articles, and, importantly, erratum[62]. Furthermore, a list of citations as well as links to derivative works can be continuously and dynamically updated[63]. Moreover, readers should have the opportunity to post comments on individual articles, organically growing what on paper would have been an inert document.

Present paper-based articles have static tables and figures. An online literature will allow for interactive vibrant and informative figures where users will be able to zoom in on parts that they may be interested in or rotate 3D protein structures. Additionally, the internet allows for dynamically updatable tables that will be available for bulk downloads[64]

As all new ideas take time to be accepted, some scientists may balk at the idea of "layering" their articles, but in the end such formats would benefit themselves when they need to access other people's work. Such formatting also requires an integrity of work, laying bare all research and results for scrutiny, allowing for no ambiguity.

Moreover, some authors may be averse to having to carefully structure their articles to conform to some seemingly arbitrary standards. These authors must understand that computers are much more capable of parsing and handling structured and well designed information, and their minimal efforts will go a long way in providing significantly more functionality. In the long run, it is in the interests of the author when her works can be communicated more widely[65].

Archives

With the journals providing only the editing and peer review portions of their original functions, the issue of presenting and archiving the data needs to be addressed. Will there be one central archive, i.e. a 'megacenter' for the whole body of scientific knowledge akin to the Pubmed abstract archive, or will there be a system of federated

archival libraries, e.g. the Biomed Archives Consortium[66], Project Muse[67], Highwire Press[47, 68] or CrossRef[69]? Will it be privately (as is the case now with journals) or publicly controlled? Should the archive include only peer reviewed information, or gray literature as well?

One commonly used example of a central archive that has done exceptionally well is the physics preprint archive. In 1991 Paul Ginsparg launched this groundbreaking archive of physics preprints, http://arXiv.org (Formally operating out the Department of Energy's Los Alamos National Laboratory now working out of Cornell University). The archive, which receives tens of thousands of papers annually functions to rapidly and efficiently distribute articles as soon as they come out, even before they are published[70].

While the international nature of scientific research would seem to make the concept of a centralized database politically unlikely [71]. Still central archives have their proponents. Matt Cockerill of Biomed Central claims that it is imperative that data be stored within a central location for there to be efficient searches of the data. Additionally, a central repository can provide for a simple and interoperability friendly interface; fears of lost data can be limited if there are multiple mirror sites[72]. The costs of maintaining any long term digital archive favor a centralized archive over some balkanized system of small independent and non-interoperatable systems.

CrossRef, which aims to not only include journals but gray information as well such as, books, reference works, and databases, claims that they can achieve the same degree of

interoperability, through the use of consensus standards, that a centralized archive can achieve, yet at the same time avoid many of the limitations inherent in a central system[69]. SPARC (Scholarly Publishing and Academic Resources Coalition), is another example of a decentralized group. It is composed of universities that publish and archive an aggregate of leading research journals at prices that are 'sensitive to the interests' of publishers and subscribers accessible journals[73, 74].

A digital archive in whatever final form it takes will have many advantages over the present day paper archives in libraries around the globe. For example, in contrast to present day libraries that cannot curate their physical stacks to remove wrong, misleading or outdated information, the dynamic nature of an online archive allows for the sequestering and possible removal of bad data. Moreover, similar to present online databases, the archive will be organic, growing and evolving based on the present and future needs of the research community.

The role of present day libraries will change from being physical repositories of information to being a 'gateway of information' providing advanced search systems and an 'expertise center' in terms of knowing how to access the different levels of the chain of information in the archives[75].

FUTURE ISSUES

In addition to the question as to who should archive is the potentially more impotent question of how to archive data. Given the rate of technological change, it is highly unlikely that any system implemented today will be similar to whatever system is used to archive the data in a couple of decades; media decays, standards change, software and the machines that can run them become obsolete and lost. The US Census information from 1960, originally stored on digital tapes, in addition to hundreds of other reels of tapes from multiple departments in the government have already become obsolete [76]. Any long term archive will need significant recurring investments to keep it operational.

Long term archiving requires that the data be maintained, easily accessible, displayed and recreated. Moreover, one cannot just print out hard copies of the archive as this defeats the purpose of a digital archive and, it in many cases, much of the information cannot be meaningfully displayed on paper (i.e. hyperlinks)[77]. The issue of data archiving is complex and mostly beyond the scope of this paper, but we will present, succinctly, some of the options.

It is imperative that whatever system is used, that it allow for easy migration of the data from one system to another, bearing in mind the exponential growth of the archived data. The ability to transfer the data from one system to another, dynamically recreating the entire archive on the new technology is very important in light of the fact that much of the media used to preserve digital data is unstable and does degrade, without active preservation, as opposed to paper archives. Even within the lifetime of the present technologies being used, the storage media on which the digital information is stored

have finite lives; data will degrade or be corrupted[78] [79] . Additionally, as the archive grows and technology changes, newer, cheaper and better media will become available for use in storage.

What is needed is a long term solution, one that does not call for heroic efforts or continual interventions to maintain it over the longterm[77]. One idea is to use some sort of semi structured representation of the data, which would include basic information with each digital object, such as the attributes of the data – its structure and physical context, information regarding the organization of the information, and information regarding the display of the information, (e.g. a user interface)[80]. The use of platform independent technologies such as XML[81] can be used to both describe and provide a simple and flexible format, and as a subsequence, longer lifetimes for the data[82].

A similar idea is, as digital archives are inherently software dependent, that the original software should be kept and, as technology changes, it should be run under emulation on the future systems; present systems also have a short physical life and as such cannot be maintained to run the software.[77] Alternatively, instead of creating emulators of outdated software, software could be designed to run on some 'universal virtual computer' that would be standardized and maintained[83].

In addition to the issues concerning storing the data, there is a more basic issue of what deserves to be stored. As stated above there are already archives that are focused on informal publications, the so called gray literature. What of the gray literature deserves

to be archived? Is all scientific data pertinent to the future and worth the cost of storage; for example, will they play an important role in terms of deciding who is deserving of scientific accolades and/or intellectual property rights for results. Additionally, even within the so called formal literature, the peer reviewed articles, how many versions of an article deserve to be preserved, (e.g. pre reviewed or drafts in progress) and should they, like the final copy of an article be preserved indefinitely.

Finally, another issue that has to be dealt with prior to the establishment of an archive is that of ownership of the articles, and the underlying research results. Although we assume that scientific results and especially those funded by the governmental grants are intended for the public domain, this is often not the case. As a result of the Bayh-Dole Act[84], universities have been encouraged to protect and profit from their research by exercising intellectual property rights. One present area where the idea of ownership for scientific fact is hotly debated is in regard to databases [85]. With regard to the archive in particular the issue of who should own should own the copyright of the article continues to be debated.

The copyrighting of scientific articles, like the patenting of scientific results funded by government funds has been termed a "public taxation for private privilege"[86]. It goes against the spirit of the law "to promote the progress of Science and the Useful Arts" by limiting the dissemination of research results. The United States Supreme Court has already ruled some time ago in Universal v Miller that research results cannot be copyrighted. Still, a trend has developed over time for journals publishers to require that

the authors sign over all their copyrights to the journal. Authors acquiesced to this Faustian bargain wherein they would hand over copyrights and in return receive affirmation that their work would be disseminated and protected in perpetuity [87]. In 1996, Congress, in the National Information Infrastructure Copyright Protection Act (H.R. 2441, and S. 1284), considered expanding the rights of owners of copyrighted articles at the expense of the academic community[88].

Recently it has been proposed that authors maintain their copyright, either through new legislation requiring the author of government funded research to do so[89, 90], or through a grass roots campaign where the authors were encouraged to not sign over copyrights[91], and in cases where they were forced to, to boycott the journal[92]. Alternatively, it has been suggested that the journals maintain copyrights only for a very limited time, after which the copyrights are transferred over to a central journal repository[23]. With the growing trend of more collaborative works of scientific research, practically, it has become significantly harder to even determine who has copyrights to what[93].

References

1. J. Guédon: 'In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing', 2001,
2. E. R. Wertman: 'Electronic Preprint Distribution: A Case Study Of Physicists And Chemists At The University Of Maryland' Virginia Polytechnic Institute and State University, 1999.
3. R. Spier, *Trends Biotechnol*, 2002, **20,** 357-8.
4. C. Tenopir and D. W. King, *Nature*, 2001, **413,** 672-4.
5. A. M. Odlyzko: in 'Access to Publicly Financed Research: The Global Research Village III' (ed. S. Wouters P, P.), 273-278; 2000, Amsterdam, NIWI.
6. A. De Kemp: in 'The Impact of Electornic Publishing on the Academic Comunity' (ed. I. Butterworth), 4-9; 1998, London, Portland Press.
7. P. Ginsparg: Creating a Global Knowledge Network 'Symposium on Electronic, Scientific, Technical, and Medical Journal Publishing and its Implications', Washington, D.C., 2003, The National Academies Committee on Scientific Enigineering and Public Policy.
8. V. Bush, *The Atlantic Monthly*, 1945, **176,** 101-108.
9. **Grey Literature: an annotated bibliography**: http://personal.ecu.edu/cooninb/Greyliterature.htm
10. W. Warnick, *Nature webdebates*, 2001, www.nature.com/nature/debates/e-access/Articles/warnick.html.
11. A. Okerson **What Price Free?**: http://www.nature.com/nature/debates/e-access/Articles/okerson.html
12. Editorial, *Nature Neuroscience*, 2001, **4,** 1151.
13. J. Kircz: New Practices for Electronic Publishing: How to Maintain Quality and Guarantee Integrity 'Electronic Publishing in Science', Paris, France, 2001, ICSU-UNESCO International.
14. A. M. F. Buck, R.C. Coles, B **Scholars' Forum: A New Model For Scholarly Communication**: http://library.caltech.edu/publications/ScholarsForum/default.htm
15. **SPARC**: http://www.arl.org/scomm/tempe.html
16. **Principles for Emerging Systems of Scholarly Publishing**: http://www.arl.org/scomm/tempe.html
17. J. Rumble: Publication and Use of Large Data Sets 'Electronic Publishing in Science', Paris, France, 2001, ICSU-Unesco International Conference.
18. M. J. Gerstein, J. **Blurring the Boundires Between Scienctific 'papers' and biological databses**: http://bioinfo.mbb.yale.edu/e-print/epub-debate-nature/text.html
19. A. N. Correia, M.: The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA : a case study by the National Library of Portugal 'Proceedings of the Workshop on Electronic Media in Mathematics', Coimbra:, 2001, Departamento de Matemática da Universidade de Coimbra.
20. N. M. Luscombe, D. Greenbaum and M. Gerstein, *Methods Inf Med*, 2001, **40,** 346-58.
21. D. Greenbaum, N. M. Luscombe, R. Jansen, J. Qian and M. Gerstein, *Genome Res*, 2001, **11,** 1463-8.

22. D. Adam and J. Knight, *Nature*, 2002, **419,** 772-6.
23. D. Shulenburger: Principles for a new System of Publishing for Science 'Electronic Publsihing in Science', Paris, France, 2001, ICSU-Unesco International Conference.
24. R. Smith, *Bmj*, 2001, **322,** 627-9.
25. Symposiym on Electronic Scientific, Technical and Medical Journal Publishing and its Implications Washington D.C., 2003, The National Academies Committee on Science, Engineering and Public Policy.
26. S. Harnad, *Nature*, 2001, **410,** 1024-5.
27. A. Cetto: The Role of Peer Review, An Alternative View 'Electronic Publishing in Science', Paris, France, 2001, ICSU-Unesco International Conference.
28. Costs of Publication 'Symposium on Electronic, Scientific, Technical, and Medical Journal Publishing and its Implications', Washington, D.C., 2003, The National Academies Committee on Scientific Enigineering and Public Policy.
29. P. Gooden, M. Owen, S. Simon and L. Singlehurst: 'Scientific Publishing: Knowledge is Power', Morgan Stanley, London, UK, 2002.
30. B. Bjork and Z. Turk, *Journal of Electronic Publishing*, 2000, **6,** http://www.press.umich.edu/jep/06-02/bjork.html.
31. B. Bjork and Z. Turk, *Electronic Journal of Information Technology in Construction*, 2000, **5,** 73-88.
32. **Press Release: Free MEDLINE**: http://www.nlm.nih.gov/news/press_releases/free_medline.html
33. **PubMed Central (PMC)**: http://www.pubmedcentral.nih.gov/
34. R. J. Roberts, *Proc Natl Acad Sci U S A*, 2001, **98,** 381-2.
35. **BioOne**: www.bioone.org/
36. **Public Library of Science**: http://www.publiclibraryofscience.org/
37. J. E. Till, *J Med Internet Res*, 2003, **5,** e1.
38. A. M. Odlyzko, *International Journal of Human-Computer Studies*, 1995, **42,** 71-122.
39. C. Tenopir and D. W. King, *The Journal of Electronic Publishing*, 1998, **4,** http://www.press.umich.edu/jep/04-02/tenopir.html.
40. H. Roosendaal, P. Geurts and P. ven der Vet, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/roosendal.html.
41. H. Yu, V. Hatzivassiloglou, C. Friedman, A. Rzhetsky and W. J. Wilbur, *Proc AMIA Symp*, 2002, 919-23.
42. M. Krauthammer, P. Kra, I. Iossifov, S. M. Gomez, G. Hripcsak, V. Hatzivassiloglou, C. Friedman and A. Rzhetsky, *Bioinformatics*, 2002, **18 Suppl 1,** S249-S257.
43. V. Hatzivassiloglou, P. A. Duboue and A. Rzhetsky, *Bioinformatics*, 2001, **17 Suppl 1,** S97-106.
44. M. Frankel, Elliott, R. Blume, M. Bourgois, J. Hugenholtz, B. Lindquist, M. Morris, S. Sandewall, E **Defining and Certifying Electronic Publication in Science: A Proposal to the International Association of STM Publishers**: www.aaas.org/spp/sfrl/projects/epub/define.shtml
45. R. Johnson **Whither competition?**: http://www.nature.com/nature/debates/e-access/Articles/johnson.html

46. P. Brown: What Must Scientists Do to Exploit the New Environment 'Symposium on Electronic, Scientific, Technical, and Medical Journal Publishing and its Implications', Washington, D.C., 2003, The National Academies Committee on Science, Engineering and Public Policy.
47. M. Keller: The Changing Role and Form of Scientific Journals 'Electronic Publishing in Science', Paris, France, 2001, ICSU-UNESCO International Conference.
48. Editorial, 2001, **291,** 2318b.
49. I. Mellman, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/mellman.html.
50. P. Bolman: The Effects of Open Access on Commercial Publishers 'Symposium on Electronic, Scientific, Technical, and Medical Journal Publishing and its Implications', Washington, D.C., 2003, The National Academies Committee on Scientific Engineering and Public Policy.
51. F. Godlee, *JAMA*, 2002, **287,** 2762-5.
52. F. J. Godlee, T. eds.: 'Peer Review in Health Sciences', 1999, London, BMJ Publishing Group.
53. E. J. Lerner, *The Industrial Physicist*, 2003, **8,** 12-17.
54. Editorial, *Nature*, 2001, **413,** 93.
55. T. Gura, *Nature*, 2002, **416,** 258-60.
56. R. Dalton, *Nature*, 2001, **413,** 102-4.
57. S. Harnad, *Science*, 1980, **208,** 974, 976.
58. W. Arms, *Journal of Electronic Publishing*, 2001, **8,** http://www.press.umich.edu/jep/08-01/arms.html.
59. P. Ginsparg **Can Peer Review Be Better Focused**: http://arxiv.org/blurb/pg02pr.html
60. **Task Group on Access to Biological Collection Data (ABCD)**: http://www.bgbm.org/TDWG/CODATA/default.htm
61. J. McEntyre and D. Lipman, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/lipman.html.
62. S. Hitchcock, L. Carr, W. Hall, W. Harris, S. Probets, D. Evans and D. Brailsford, *D-Lib*, 1998, http://eprints.ecs.soton.ac.uk/archive/00000746/.
63. D. M. Eagleman and A. O. Holcombe, *Nature*, 2003, **423,** 15.
64. M. Frankel: 'Seizing the Moment Scientists' Authorship Rights in the Digital Age', American Association For the Advancement of Science, 2002.
65. T. Berners-Lee and J. Hendler, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/bernerslee.html.
66. **Biomed Archives Consortium**: http://140.234.1.105/
67. **Project Muse**: http://muse.jhu.edu/
68. **Highwire Press**: http://highwire.stanford.edu/
69. E. Pentz, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/pentz.html.
70. M. Sincell, *Science*, 2001, **293,** 419-21.
71. R. Luce, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/luce.html.
72. M. Cockerill, *Nature webdebates*, 2001, http://www.nature.com/nature/debates/e-access/Articles/cockerill.html.

73. Editor, *Information Today*, 1999, **16,** 32.
74. S. Michalak, *Serials Review*, 2000, **26,**
75. A. Klugkist: The Changing Role of the Librarian - A Virtual Library and a Real Archive? 'Electronic Publishing in Science', Paris, France, 2001, ICSU-UNESCO International Conference.
76. J. Rothenberg, *Scientific American*, 1995, **272,** 42-47.
77. J. Rothenberg: 'Avoiding Technological Quicksand', The Council on library and Information Resources, 1999.
78. M. Dementi, *The Journal of Electronic Publishing*, 1998, **3,** http://www.press.umich.edu/jep/03-03/dementi.html.
79. Guthrie, *Educase*, 2001, www.educause.edu/ir/library/pdf/erm0164.pdf.
80. R. Moore, C. Baru, A. Rajasekar, B. Ludascher, R. Marciano, M. Wan, W. Schroeder and A. Gupta, *D-Lib*, 2000, **6,**
81. B. Ragon, ***Computers in Libraries***, 2003, **23,** 10-14.
82. F. Achard, G. Vaysseix and E. Barillot, *Bioinformatics*, 2001, **17,** 115-25.
83. R. Lorie: 'A Project on Preservation of Digital Data', IBM Almaden Research Center, 2001.
84. **P.L. 96-517; The Bayh-Dole Act** (35 USC 200-212) 1980
85. D. Greenbaum, *Albany Law Jounral of Science and Technology*, 2003, **13,** 431-515.
86. D. Kreeger, *Law and Contemporary Problems*, 1947, **12,** 714-7445.
87. S. Harnad and M. Hemus: in 'The Impact of Electronic Publishing on the Academic Comunity' (ed. I. Butterworth), 18-27; 1998, London, Portland Press.
88. S. Colbert and O. Griffin, *Albany Law Review*, 1998, **62,** 437-465.
89. S. Bachrach, R. S. Berry, M. Blume, T. von Foerster, A. Fowler, P. Ginsparg, S. Heller, N. Kestner, A. Odlyzko, A. Okerson, R. Wigington and A. Moffat, *Science*, 1998, **281,** 1459-60.
90. C. McSherry: 'Who Owns Academic Work? Battling for Control of Intellectual Property', 2001, Cambridge, Harvard University Press.
91. L. Guernsey, *The Chronicle of Higher Education*, 1998, **5,** A29.
92. M. Wadman, *Nature*, 2001, **410,** 502.
93. R. Dreyfuss, *Vanderbilt Law Review*, 2000, **53,** 1162-1232.