

September 2005

I Hear the Train A Comin' -- PubMed Central

Greg Tananbaum

The Berkley Electronic Press, greg@bepress.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tananbaum, Greg (2005) "I Hear the Train A Comin' -- PubMed Central," *Against the Grain*: Vol. 17: Iss. 4, Article 31.

DOI: <https://doi.org/10.7771/2380-176X.4534>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

GOOGLE GENESIS

by Sandra Beehler (Lewis & Clark College)

The styles of **Larry Page** and **Sergey Brin** clashed when they first met on the **Stanford** campus, but their later research collaboration had a better outcome. **Page** was interested in how Web links worked, and made the intellectual connection between Web linking and bibliometrics: the science of citations. That connection led to investigation of back links in a project **Page** called **BackRub**. As with academic citations, **Page** thought Web links could be ranked according to their importance. The challenge was creating the math to describe that ranking — and that's where **Brin**, a math prodigy, came in. Together **Brin** and **Page** created an algorithm (**PageRank**) that counts those back links, thus establishing the importance of any given Website. At that point they realized that **PageRank** could be adapted as a search tool. Comparing search results to those of existing Web search engines proved the superiority of the **PageRank** system. They set up a machine lab in **Page's** dorm room and a programming center in **Brin's**, begged and borrowed equipment, and used the **Stanford** broadband connection as their way into the Web. Named **Google**, after a mathematical term, their new search engine quickly caught on and now dominates Web searching.

See — “The Birth of Google,” by **John Batelle**. *Wired* 13.08 (Aug 2005), p. 103-110.



OPINING OPEN SOURCE

by Jane H. Tuten (USC Aiken Library)

Written by the founder of **Lotus Development Corp.** and the co-founder of the **Electronic Frontier Foundation**, this short article provides both a great explanation and offers some constructive criticism of the open source software movement. **Kapor** discusses the alternative nature of the open source software model including the decentralized and transparent nature of the software and any changes that might be made. Open source software encourages participation from folks around the world with the only requirement being a willingness to contribute. Anyone can participate and the role is dependent only on the individual skills of the participant. **Kapor** points out that the openness might also be a downside because there are certainly no guarantees that the open source product will provide the needed solution. Risk is involved. Open source projects need the proper technical infrastructure to succeed and a solid community infrastructure that will provide some oversight of the project. Additionally, **Kapor** notes that the open source community must have clear values and principles stipulated which provide oversight for the practices of the community. The academic world needs to become more involved in open source projects because both participants will benefit.

See — “How Is Open Source Special?” in Viewpoints section by **Mitchell Kapor**, *Educause Review March/April 2005*, pp. 72-73.

I Hear the Train A Comin' — PubMed Central

by **Greg Tananbaum** (President, The Berkeley Electronic Press, 2809 Telegraph Avenue, Suite 202, Berkeley, CA 94705; Phone: 510-665-1200 x.117; Fax: 510-665-1201) <greg@bepress.com> www.bepress.com

Welcome to the 1.0 edition of **I Hear the Train A Comin'**. My intention is to spend a few hundred words each month discussing, in an informal way, what's around the bend on the scholarly communications track. I will try to draw some general conclusions — and raise some general questions — by looking at a case study or two each column. Why me? As President of **The Berkeley Electronic Press**, I seem to spend a lot of time these days navigating the treacherous terrains of institutional repositories, alternative journal pricing models, open access publications, and new and emerging forms of scholarly communication. Why now? **Katina Strauch** asked me, and I know better than to refuse **Katina** — “stupendiferous” can become just plain “stupid” with a few strokes of her keyboard. And why “**I Hear the Train A Comin'?**” Beyond serving as an appropriate metaphor for a future intuited but as yet unseen, it serves as the opening line to **Johnny Cash's Folsom Prison Blues**, the coolest song ever recorded.

This month, let's focus on **PubMed Central's** recent efforts to capture publicly funded research in an openly accessible archive. As of early May, 2005, the **US National Institutes of Health** recommended, but do not require,

that all **NIH**-funded investigators submit an electronic version of their peer-reviewed final manuscripts, upon acceptance for publication, to **PubMed Central**. **NIH** asks that authors make these manuscripts available immediately after the final date of journal publication. Authors are given the option to delay the release of their manuscripts at a later time, up to 12 months after the official date of final publication. This policy set off loud debate within the academy, with most of the volume provided by one of two “true believer” camps. One camp argues that the **NIH** is stepping on private enterprise by seeking to make copyrighted materials freely available to the world. By offering a competing, free version of an article, this line follows, the government is on the path to state-run publishing, or even government-controlled science. The other camp believes that the couched language of the pronouncement, including recommendation rather than requirement and a 12 month delay, render it stillborn. Indeed, the original recommendation from **Congress** in the summer of 2004 was rather more stringent in its language. Before either side turns blue from shouting about the impending end of the world, let's step back and talk about some practical elements of the policy.

Researchers wishing to post their papers in **PubMed Central** must go through a number of straightforward steps. First, they log on to the **NIH** manuscript submission system at <http://nihms.nih.gov/>. Next, they follow simple Web instructions to enter basic metadata, as well as specific **NIH**-required information such as the grant number. Files may be uploaded in a number of formats, with images provided separately

in high resolution format. Finally, researchers must indicate the length of the embargo for withholding their papers from the general public. The submission statement through which researchers must click reads as follows:

I hereby submit an electronic version of my final manuscript that is the result of research supported, in whole or in part, with direct costs by the **National Institutes of Health**.

This manuscript has been accepted for publication in [JOURNAL NAME] and includes all modifications resulting from the peer review process. The manuscript contains confidential information and I request that it not be disclosed prior to the time indicated below.

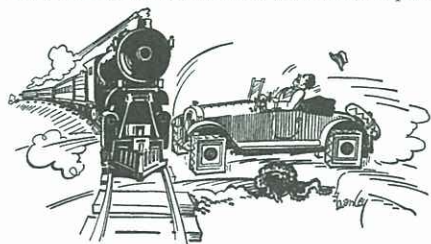
I request that this manuscript be publicly accessible through **PubMed Central** [DELAY PERIOD] after the publisher's official date of final publication and am notifying the publisher of this action.

I understand that this submission is voluntary and provides an alternative means for fulfilling the existing requirement to provide publications as part of **NIH** grant progress reports.

Upon submission, the **National Library of Medicine** converts the data into XML and sends notice back to the submitter to verify this new version. The paper is posted within **PubMed Central** upon confirmation from the submitter that the markup version is acceptable.

That is an executive summary of the submission process. Looking deeper, what clues

continued on page 79



I Hear the Train A Comin' from page 78

does this submission process provide into the policy's prognosis? Put another way, what practical factors may influence the policy's adoption for researchers who fall outside of either "true believer" faction? In these early days, perhaps the biggest selling point for the dispassionate author is the ease of the submission mechanism. The Web form is simple, quick, and nontechnical. The submitter need not have any special computer savvy to prepare the manuscript. The time and energy a submitter must exert to get a paper into **PubMed Central** is minimal. Indeed, according to early data released by the **NIH Public Access Working Group** of the **NLM Board of Regents**, 84% of submitters indicated a submission time of ten minutes or less.

There are several factors, however, which may limit widespread adoption among authors. For one, the submission process is confusingly silent on copyright matters. No copyright statement is included in the submission form. Separately, in the *NIH Public Access Policy Authors' Manual*, a section states, "Authors and/or their institutions should ensure that their final manuscript submissions to **PMC** are consistent with any other agreements, including copyright assignments that they may have made with publishers or other third parties." Authors unsure of their article's copyright status and the procedures for obtaining clearance may avoid submission altogether.

Another gating factor is the duplication of effort that **PubMed Central** submissions involve. Authors submit their papers to the journal. They submit them to their institutional repositories. Yet another upload of the same metadata, no matter how simple the **PMC** form is, may prove untenable to some. The **National Library of Medicine** no doubt hopes to work with publishers to create a common submission or data forwarding mechanism. Squaring away the motivations of each party would seem a delicate task.

Finally, it is not clear how the **National Library of Medicine** might scale to meet the potential success of the policy. At present, all XML conversions are performed by the **National Library of Medicine**. With an adoption rate among funded researchers at less than 5%, **NLM** resources have not yet been taxed. Whether such labor-intensive efforts are sustainable at higher adoption rates is a very open question. Any restriction placed on file types, or conversion requirements pushed back to the submitters will undoubtedly dent uptake among the authoring community.

In concluding my look at the practical application of the **NIH** policy, I again pull away from the extremes. The system works well, but not perfectly. It is a good "1.0," but is not so developed as to render participation costs nil. It is neither perfect nor flawed enough, in sum, to provide succor to either tribe of true believers.

Where will **PubMed Central** go from here? I'll speculate that adoption rates among funded researchers will remain low unless one of two shifts occurs. The first shift would be the re-


placement of the recommendation for deposit with a requirement as a condition of funding. It is generally true that a rule yields greater adoption than a request. The second shift would be the recruitment of major publishers as strategic partners in the initiative. Publisher delivery of papers to **PMC** in automated fashion could reduce the author's role to checking an opt-in box. Piggybacking papers' markup on existing journal production systems would also alleviate scalability concerns.

The first shift relies on legislative intervention. Given the history of the original policy's wording, in which key language was progressively watered down over the course of a year (see <http://www.earlham.edu/~peters/fos/nihfaq.htm> for history) this may be a non-starter. On the other hand, perhaps the modest early results will embolden the cause's political champions to strengthen the provisions. As someone who voted for **Gray Davis**, **Bob Dole**, and **Paul Tsongas**, I fear my intuitions on matters political are insufficient to offer a reasoned prediction here.

The second shift would require the true believers within the **NIH** and the publishing industry to decide that the risks of collaboration are preferable to the alternative. Why might this be? Fear, mostly. Fear among the publishing community that prolonged antagonism will push the government to proceed with mandated participation. Fear among the **NIH** that publisher resistance will subtly suppress participation for long enough that the enterprise will be labeled a failure. Fear among all parties that the spilling of this debate into public consciousness will damage their standing among the researchers that provide their lifeblood.

What might the give and take of collaboration resemble? One can surmise that the publishers would seek to limit the free duplication of their proprietary content. This might mean extending the moving wall to 18 or 24 months. Alternatively, the publishers might push for public access to only the most current articles, with any content older than, say, two months pushed back to restricted access. Publishers would also likely seek to reinforce the value of their proprietary copies as the versions of record. For its part, the **NIH** might push for all relevant manuscripts from participating publishers to be forwarded unless the author explicitly declines (i.e., changing the aforementioned opt-in check

box to opt-out). They might also require a standard delivery mechanism from all participating publishers to streamline production.

It is not immediately clear whether such collaboration will occur. One factor that bears watching is the mindset of the funded researcher. It may be that author indifference to **PubMed Central** continues over time, or that participation grows dramatically as the concept is demonstrated. Such an attitudinal shift would no doubt influence the strategies of each vested party. 

My hope and expectation is that this column will spark discussion and debate. To that end, I welcome your thoughts — my email address is <greg@bepress.com>.

**The APS
Journal
Legacy
Content**

**SAVES you 200
Linear feet of
SHELF SPACE**



**14 Journals of the American
Physiological Society**
www.the-aps.org

International Dateline — Aux Amateurs de Livres: How to Get Hold of French Books Painlessly

by **Claire-Lise Benaud** (University of New Mexico, Albuquerque, NM 87131)
<clbenaud@unm.edu>

Introduction

As a librarian and selector for French and francophone literature at the **University of New Mexico**, I had the opportunity of doing a four-month internship at the bookstore **Aux Amateurs**

de Livres International in France during a sabbatical. To work for the other side is always an interesting experience. I went from working for an institution whose job is to spend money wisely

continued on page 80