

April 2010

# I Hear the Train A Comin' -- Cornell Hosting arXiv

Greg Tananbaum

*Anianet*, [greg@anianet.com](mailto:greg@anianet.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 (2010) "I Hear the Train A Comin' -- Cornell Hosting arXiv," *Against the Grain*: Vol. 22: Iss. 2, Article 42.

DOI: <https://doi.org/10.7771/2380-176X.5531>

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

a library to manage multiple PDA programs with multiple eBook vendors. As an approval and firm order vendor, YBP is in a unique position to know exactly what a library has already acquired in print and can de-duplicate the list of titles in the PDA program to eliminate the possibility of purchasing a title more than once.

### Benefits of PDA

Libraries can benefit from incorporating a variety of patron-driven acquisition tactics into their daily operations, expanding the ways in which patron needs are assessed and fulfilled. The UT Libraries have, in addition to implementing the PDA program through EBL as described above, purchased print materials on demand through its Interlibrary Services. And, as a next step, the UT Libraries are investigating the possibility of incorporating patron-driven acquisition into their print approval plan through BNA. (Macicak and Schell, S36)

Patron-driven acquisition, in all of its incarnations, offers a number of potential benefits to libraries. In her column in the January 2009 issue of *Booklist*, Sue Polanka states that “[a]mong the benefits of PDA are guaranteed usage of new titles and proven usage of purchased titles, automatic acquisition, and seamless access for patrons.” Polanka goes on to say, “In addition, PDA can save time and eliminate guesswork for selectors, and it can be a cost effective alternative to ILL.”

While Polanka’s column specifically addresses patron-driven acquisition of eBooks, many of her comments regarding the benefits of PDA apply to all methods of patron-driven acquisition. Whether a library is purchasing materials that have been requested through interlibrary loan or using patron-driven acquisition to select print books, the same benefits apply. Libraries acquire materials based upon demonstrated patron needs, and they can be assured that the materials that are purchased will be used at least once.

### References

Alder, Nancy Lichten. “Direct Purchase As a Function of Interlibrary Loan.” *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 18, no. 1 (2008): 9-15.

Macicak, Susan and Lindsey E. Schell. “Patron-driven, librarian approved: a pay-per-view model for e-books.” *Serials* 22, no. 3 (November 2009): S31-S38.

Morehouse, Sarah. “Two ILLiad Clients, One Desktop, Purchase on Demand; Sharing a University’s Collection, Staff, and Expertise.” *Journal of Access Services* 6, no. 4 (October 2009): 497-512.

Polanka, Sue. “Patron-Driven Acquisition.” in *Off the Shelf. Booklist* 105, no. 9/10 (January 1 & 15, 2009): 121.

Zopfi-Jordan, David. “Purchasing or Borrowing.” *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 18, no. 3 (2008): 387-394. 🌱

# I Hear the Train A Comin’ — Cornell Hosting arXiv

Column Editor: **Greg Tananbaum** (Founder and CEO, Anianet) <greg@anianet.com> [www.anianet.com](http://www.anianet.com)

**Column Editor’s Note: Teresa (Terry) Ehling is the Director of the Center for Innovative Publishing at Cornell University Library. I recently had the pleasure of speaking with her about arXiv, home to more than a half-million open-access e-prints in physics, mathematics, computer science, quantitative biology, quantitative finance, and statistics. Cornell, which has hosted arXiv since 2001, recently announced a significant change to its funding model. Terry was kind enough to explain the changes, as well as provide some insights into arXiv’s long-range planning. — GT**

### Why did arXiv need a new business model?

arXiv has been with us for nearly 20 years. It came to Cornell with Paul Ginsparg in 2001 and was supported initially by a mix of NSF funding and direct financial support from the Cornell University Library. In 2007 the Cornell University Library assumed full operational responsibility for arXiv, and the library has been arXiv’s primary source of financial support since that time. The recession and its impact on the library budget made it imperative that we identify and pursue alternative funding sources for arXiv. We developed a short-term business plan that relies on annual contributions from the top 200 users of arXiv, based on download data from the previous year. We believe that arXiv should be promoted as a shared investment in an academic resource that is vitally important for the researchers in the physics, mathematics, and computer science domains.

### arXiv’s costs translate into a cost per download of 1.3 cents, or a cost per submission of around \$7. How does this compare with other scholarly publishing structures?

I would use the word “contrast” rather than “compare.” arXiv is a unique scholarly communications initiative. Unlike conventional publishing projects, arXiv’s MO is not to add value, in the traditional sense, to the content it delivers. For example, submissions to arXiv are not peer-reviewed. According to Paul Ginsparg, it was designed from the outset to operate several orders of magnitude less expensively than the journal system. From publicly available data we can also say with confidence that arXiv is less costly to operate, on a per-article basis, than most (if not all) institutional repositories. arXiv was and remains a rapid distribution system for e-print material in the

sciences and related fields. It was designed and optimized for the community it serves by members of that community.

### Can you succinctly explain how the new model works?

Cornell University Library has launched a three-year effort to solicit voluntary, annual support from libraries at academic institutions, government research labs, international research centers, and other organizations that are the heaviest users of arXiv, based on the prior year’s download statistics. We have devised a tiered pricing model for the top 200 users; the suggested contribution for the top 100 institutions is \$4,000, \$3,200 for institutions ranked 101-200, and \$2,300 for those ranked below 100.

### Why did you select a use-based model? What were some of the other possibilities you considered?

After extensive feedback from arXiv stakeholders at libraries around the world, we decided to base our pricing structure on use, specifically downloads. Given that we are committed to sustaining arXiv as a fully open-access academic resource, we felt this would be the most equitable model we could implement. We may consider a more nuanced user-based model in the future; for example, delimiting institutional use based on both submission and download data. We might also incorporate support from scholarly societies, an endowment, or funding agencies such as the NSF.

### What has the response been from the institutions with heaviest use?

The response to date (early March) has been overwhelmingly positive. We’ve received funding commitments from 45 institutions, domestic and international, since we launched our campaign in mid-January.

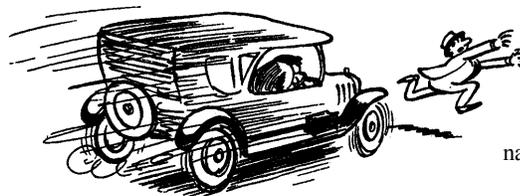
### How concerned are you about the free rider problem, with some institutions shouldering less than a fair share of the costs?

We have and always will have free riders. We don’t expect every one of the nearly 4,000 institutions we can identify by domain to contribute to arXiv’s support. arXiv’s top 200 users represent approximately 75% of arXiv’s download activity during the calendar year, so it made sense to target this group during the 2010 launch year.

### The new model is meant to be a three-year, interim plan. Why not make it permanent?

We need time to explore our income options. Our goal is to realize a diversified revenue portfolio for arXiv. Ultimately we hope our funding docket will include a blend of ongoing underwriting from the Cornell University Library and financial support from the academic library

*continued on page 85*



## I Hear the Train A Comin' from page 84

community and research centers. An endowment, as well as support from scholarly societies or funding agencies, may also be part of our mature portfolio.

*Is there any role for commercial entities in the long-term sustainability model? If so, what might that look like?*

Possibly. We've had some preliminary discussions with a few of the discipline-relevant scholarly societies about how they might be involved. We hope to continue the conversation with them over the course of the next year or two. Participation by the societies or commercial publishing entities must, of course, be informed by terms and conditions that are acceptable to both the arXiv stakeholders and the outside agencies.

*For other disciplines in which the culture of mass distribution of open-access content by authors is not as strong, are arXiv's recent financial developments encouraging or a cautionary tale?*

Both. The mass distribution of content is costly. There are a variety of ways to recover those costs, and it's both prudent and shrewd to devise a model that will ensure that the resource will realize revenue from a variety of differ-



**ScholarNext**

Consulting Services at the Intersection of  
Technology, Content, & Academia

- ✓ Web 2.0 Functionality
- ✓ Strategic Planning
- ✓ Business Development
- ✓ Policy Analysis
- ✓ Publishers
- ✓ Information Providers
- ✓ Technology Companies
- ✓ Libraries

(510) 295-7504      Greg Tananbaum      greg@scholarnext.com

[www.scholarnext.com](http://www.scholarnext.com)

ent sources. It's critical that the stewards of scholarly resources understand scale and scope economies, think strategically instead of tactically, and be willing to shift directions when their winds of fortune change. 🌸

## Standards Column from page 86

provides the linkage of different manifestations of the same content. For example, the first ISTC was assigned to **Charles Darwin's** *On the Origin of Species*. This text has been reprinted by numerous publishers in various forms, both print and electronic, since it was first published in 1859. A search of the **WorldCat** catalog found 281 different editions with **Charles Darwin** as the listed author, with the latest edition appearing in 2009. While few books have this many editions, in today's digital environment, if each version is unambiguously identified (presumably with unique ISBNs), it's very likely that every single title published today could have at least a dozen different versions and an equivalent number of ISBNs. The ISTC will be the tool that links these different versions together, both for discovery and collection management. On the publisher side, the ISTC also assists with rights management. If the ISBN is used consistently for disambiguation, the ISTC can provide the umbrella framework for all of the multiple product references and could help to

streamline metadata creation and improve its quality. The ISTC standard was published in early 2009 and is currently being used by a few publishers but needs to gain wider adoption to be broadly useful. Additionally, best practices need to be adopted regarding the sharing and consistent application of metadata between the ISBN and the ISTC.

The **International ISBN Agency** recently released a white paper (<http://isbn-international.org/news/view/29>) describing the problems of ISBN assignment to eBooks, emphasizing the need for consistent practices for assignment of ISBNs, and recommending some solutions. From the perspective of a centralized and globally adopted system, ensuring the consistency of the system's application is paramount. This applies not only to different formats, but also to sub-items, such as chapters or other fragments of works, when these items become part of the supply chain. The paper also suggests that there might be a need to identify the generic publication file produced by the publisher that is then used for further re-formatting and distribution. The music industry has such a system in place for the original track, which can then be placed on a variety of albums or even sold separately.

This original track — defined as “an abstract entity representing a bundle of one or more Digital Resources compiled by an Issuer for the purpose of electronic distribution to individual consumers, directly or through intermediaries” — is identified using a Global Release Identifier, or GRID. The GRID standard ([http://www.ifpi.org/content/section\\_resources/grid-standard.html](http://www.ifpi.org/content/section_resources/grid-standard.html)) clearly states that “the Release is not itself a Product.” The **International ISBN Agency** is now discussing with various stakeholders whether a similar system of using generic eBook release identifiers is both feasible and of value.

Although the digital distribution of content has been gaining momentum for more than a decade and a half, the eBook is still in its relative infancy. But eBook production is growing faster than its predecessors, such as the e-journal. As a result, the problems in the supply chain are getting out of control. We need to act quickly to develop a consensus solution. If you're a stakeholder in this arena, the **International ISBN Agency** (<http://isbn-international.org/>) and the **International ISTC Agency** (<http://www.istc-international.org/>) are good places to start. 🌸

## Rumors from page 63

back to the Palmer House and ALAP (**ALA Players**) was born. Every ALA winter and summer conference there is a dinner dance put together by **Fred Goodman** at LLSI usually on Monday night to commemorate the event. And, it's a hoot to go to the ALAP homepage about the history of this whole event. It's complete with a photo gallery, and memories of the 20th anniversary gala in New Orleans in 1998 [homepage.mac.com/gailmcgovern/PhotoAlbum23.html](http://homepage.mac.com/gailmcgovern/PhotoAlbum23.html).

Speaking of Iceland — The **Steering Com-**

**mittee of Iceland Consortia** for electronic subscriptions has selected **ProQuest Central** to provide electronic resources nationally for the next three years. On behalf of the Ministry of Education, Science and Culture, the **Steering Committee** is responsible for the evaluation and procurement of resources which can then be accessed by all Icelandic citizens from any computer within the country. The project is funded by the Ministry along with more than 200 libraries, educational institutions and private firms. [www.proquest.com/](http://www.proquest.com/)

Yours truly was just interviewed by **Jack McHugh** and the interview is posted on his Website. [www.johnbmchugh.com/](http://www.johnbmchugh.com/). Click on

expert interviews and mine is the fourth one down. Would love comments.

Was interested to see a report on eBook devices from the floor of the **Consumer Electronics Show (CES)** in Las Vegas. Try this link for a look at some of the devices. Ho-ha! [aptaracorp.com/context/Aptara\\_Context\\_Feb\\_2010.pdf#page=3?elq=369ac855817c4b66b93d7b10007dfcc](http://aptaracorp.com/context/Aptara_Context_Feb_2010.pdf#page=3?elq=369ac855817c4b66b93d7b10007dfcc)

**Last but not least! Baby Georgia Helen Walser is here! Born March 7, 2010 at 6:49pm. 7 lbs 10 oz; 19.5 inches long. This is Katina's and Bruce's second grandbaby! And Georgia Helen skyped her father in Iraq in the first 15 mins of her life! She is obviously a genius! ;-)** 🌸