

OR2010 Presentation Proposal:

Repository sustainability: arXiv business model experience and implications

Simeon Warner and Oya Yildirim Rieger
DLIT, Cornell University Library
Ithaca, NY 14850, USA
simeon.warner@cornell.edu, rieger@cornell.edu

In January 2010 Cornell University Library moved to expand the funding base for arXiv by requesting support from user institutions. We hope that this voluntary support model will engage the institutions that benefit most from arXiv while maintaining arXiv's open access mission as a service free to readers and submitters alike. The development of a business model has made us look closely at arXiv's sustainability from both operational and technical standpoints. The engagement of supporting institutions creates new requirements to demonstrate value to these institutions as separate from arXiv's understood value to the community in general. In this presentation we will briefly describe options considered in development of the business model, the model chosen, uptake and feedback. We will then focus on the implications for arXiv's operation, for the long term development of our platform, and new reporting facilities.

1 Context

Many repositories and other scholarly information services are started as side projects of research projects with grant funding in the early phase. Over time they may become larger, increasingly expensive to operate, and more infrastructure than research. They may also become indispensable to the community they serve. This is the story of arXiv, which was created in 1991 and has grown to be the key scholarly communication venue for some areas of physics, and an important service for a much broader community of physicists, mathematicians and computer scientists. arXiv moved to Cornell University in 2001 and Cornell University Library has provided the bulk of the funding to support arXiv's operation.

2 arXiv business model

During 2009 we considered many possible business models for arXiv (see [2, 3] for background). We surveyed similar projects and consulted with a number of librarians especially about library support models. In January 2010 we announced a voluntary support model in which we are requesting support from the heaviest user institutions (described in detail in [1]). The response so far has been extremely positive and we will describe the process, status and feedback obtained. One motivation for choosing the institutional support model was the desire to keep fundraising costs under control. Initial effort for communication with potential supporting institutions and administration has been quite high but we hope this will stabilize. We will summarize costs associated with implementing

the model and compare with expectations.

One concern expressed about our model is that it might open a floodgate of requests for support. We will argue that there are a number of criteria and standards that a repository should be expected to meet in order to justify community support, and thus that relatively few repositories would qualify.

3 Platform

The arXiv software has evolved throughout its 18 year history and predated most other repository systems. It includes some elements that are specialized to our community and service: the TeX processing system and our optimized administrative workflow for example. Other elements are more generic and are good candidates for replacement with standard components which we hope will reduce costs or free developer effort for the development of new features and services. There are many ways to split this problem and no single system offers all of the facilities required. For example, would arXiv be best adopting a very standard component such as Fedora for underlying repository functionality, or the Invenio system common within the physics community (Developed at CERN, being adopted by ADS) which has more appropriate features built in an offers opportunities for shared development with partner projects?

4 Focus on the institution

He who pays the piper calls the tune or so the saying goes. arXiv has traditionally focused on the individual and relied little upon institutional affiliation except as a means to help verify membership within the academic community (by providing an email address for example). Most submitters to arXiv do not include affiliation information within the author metadata they enter. Now that arXiv is requesting institutional support there is a need to focus somewhat more on the institution and to demonstrate value at that level.

4.1 Institutional reporting of submissions

The most common request since announcing the new business model has been for statistics relating to institutional use of arXiv. In preparation for the announcement we calculated download statistics by institutional domain for 2009 (see: http://arxiv.org/help/support/2009_usage) and will refine that with IP ranges supplied by supporters for 2010. However, arXiv does not have data from which to calculate accurate submission statistics by institution. Should arXiv demand that submitters identify the affiliation of each author? Should administrative effort be spent normalizing this information given the lack of institutional name authority services? Could institutional effort be co-opted to help curate this data?

4.2 Account management

Although arXiv will remain free to readers and submitters, the institutional funding model results in a need to manage institutional accounts in a manner very similar to that of a toll-access service. We need to obtain IP address ranges from supporting institutions in order to accurately track and report usage. With this information we may also relatively easily provide acknowledgement of local support with a customized banner. The same account management system can be used to track and register local OpenURL endpoints, something arXiv has not done up to now because of the additional administrative overhead it would introduce.

4.3 Interoperability with institutional repositories

Institutions with institutional repositories are usually keen to have them used, and would like to avoid the need to for authors to make multiple deposits. arXiv implemented the SWORD [4] protocol for automated deposit over a year ago. This protocol enables both multiple deposit from a single tool and deposit from another repository but has yet been used to address the multiple deposit problem while it has been successfully used by journals and conference systems depositing in arXiv. What are the impediments? What else is needed to make this work for multiple deposit?

References

- [1] arXiv. arXiv Business Model White Paper, Jan 2010. <http://arxiv.org/help/support/whitepaper>.
- [2] Raym Crow. Income Models for Open Access: An Overview of Current Practice, 2009. <http://www.arl.org/sparc/publisher/incomemodels/>.
- [3] Kevin Guthrie, Rebecca Griffins, and Nancy Maron. Sustainability and Revenue Models for Online Academic Resources - Final Report, May 2008. http://www.ithaka.org/ithaka-s-r/strategy/sca_ithaka_sustainability_report-final.pdf.
- [4] SWORD (Simple Web-service Offering Repository Deposit) profile of the Atom Publishing Protocol. <http://www.swordapp.org/>.