



— Paul Jones

# OPEN(SOURCE)ING THE DOORS FOR CONTRIBUTOR-RUN DIGITAL LIBRARIES

What if you could wave a wand and make libraries—at least digital libraries—more open, easier to manage, cheaper, and even more eclectic and democratic? What if content contributors could submit, catalog, index, manage, rate, and rank materials in large collections themselves? Thanks to innovations from the open-source community and perhaps more importantly the free-software community, I believe we can have contributor-run libraries today. —

In fact, there are several very successful examples from which we can draw not only best practices, but also—that grail of the programmer—working code. But better still, these projects are also examples of vibrant, lively, noisy, democratic communities.

The first step in contributor-run libraries is to allow people to contribute. This may sound obvious, but many collections try to control or gatekeep from the onset. Our experience with the Linux Software Archive (MetaLab.unc.edu), which began in 1992, was that by removing nearly all barriers to submission and instituting some simple verification procedures, we were able to accept (and later distribute) some very high-quality software with a very low rejection rate.<sup>1</sup> Submissions are accepted by a simple FTP upload to a secure area. Along with the software, we require some basic metadata called the “Linux Software Map” (metalab.unc.edu/pub/Linux/LSM-TEMPLATE) to identify the author, title, and describe the software. There are only 12 fields in all and only four are required. Our rejection rate due to missing or improper metadata is at

a low 4.5% although we have contributors from every corner of the globe.<sup>2</sup>

What this experience tells us is that opening the doors to contributors may not be as scary as we may have been led to believe. Of course, digital libraries don’t have the same shelf space problems as physical ones. But the fact the metadata and the attendant organizational assistance taken directly from contributors are reliable and immediately useful is encouraging.

But others have found that encouraging contributors to rank and comment on the contributions of others adds great value and creates a favorable environment for a noisy, active, democratic community to develop and grow. Large book wholesalers, including Amazon (www.amazon.com), and Barnes and Noble (www.bn.com), add value to their offerings by collecting and ranking both user comments and comments on those comments.

Other sites, most notably Slashdot.org (see www.slashdot.org), have instituted a reward systems so valued contributors and commenters accrue karma points that allow them to act as moderators of discussions and to rank comments and

<sup>1</sup>For a full description of the MetaLab Linux Software Archives and contributor demographics, see [metalab.unc.edu/orst/developro.html](http://metalab.unc.edu/orst/developro.html).

<sup>2</sup>See draft report by J. Greenberg at [ils.unc.edu/~janeg/lsmstudy](http://ils.unc.edu/~janeg/lsmstudy).

*WHAT THIS EXPERIENCE TELLS US IS THAT  
OPENING THE DOORS TO CONTRIBUTORS MAY NOT BE AS SCARY AS  
WE MAY HAVE BEEN LED TO BELIEVE.*

stories. Devices such as karma points serve as a hedge against trolls, group takeovers, fakers, and the like.

More sophisticated structures such as Advogato's trust metric (see [www.advogato.org/trust-metric.html](http://www.advogato.org/trust-metric.html)) and other schemes to evaluate reputation capital offer an even stronger and more reliable community structure for ensuring rich and useful ranking and evaluation.

By giving contributors and readers access to tools for evaluating, ranking, and managing the collections, we are not just off-loading work; we are building communities of intellectual discourse. Strong community members are recognized by reputation capital and trust metrics and are rewarded.<sup>3</sup>

Digital libraries can give back to contributors as well. By sharing collected information, contributors can see which items (manuscripts, songs, and software) are most in demand in the form of top-10 lists or most recommended. This enhances not only the referral services, but helps new contributors understand what is considered a good item.

More sophisticated sites for contributors, such as SourceForge for open-source software developers ([sourceforge.net](http://sourceforge.net)), provide the tools a project needs to get going on its own. Roadblocks to developers are removed by offering FTP and Web hosting, list services, project status pages, version control software, backups, and discussion forums. By supplying these simple tools, SourceForge became one of the largest collections of open-source projects in the world within a matter of months. Last fall, SourceForge added Advogato's trust matrix for contributors. While SourceForge directs its energy toward software developers, their needs are similar to those of contributor communities in any medium or genre.

What makes the tools described here of particular interest to digital library projects is they are open source and free (issued under the Free Software Foundation's General Public License)<sup>4</sup> for the most part. In the great tradition of public libraries, the tools and sites can be shared, built upon, and adjusted to local or particular circumstances. The tools and the concepts they use have proven useful and effective in live

and vocal communities. They have produced real and effective collections and more importantly real and effective communities in the best democratic sense.

By adopting not only the open-source tools, but also the open-source philosophy encouraging community interaction and contributor involvement, digital libraries can open new horizons to new communities as well as greatly improve traditional services.<sup>5</sup> ■

---

PAUL JONES ([pjones@ibiblio.org](mailto:pjones@ibiblio.org)) is Director of [ibiblio.org](http://ibiblio.org)—a collaborative project of the University of North Carolina—Chapel Hill and the Center for the Public Domain.

---

© 2001 ACM 0002-0782/01/0500 \$5.00

---

<sup>5</sup>See Richard Stallman's groundbreaking work at the Free Software Project [www.gnu.org/philosophy](http://www.gnu.org/philosophy).



## Looking for a great job?

We need great people to help take our ferocious, web-text-extracting technology to the world masses! High-growth opportunities for **Machine Learning Researchers and Senior Software Engineers** in the mountains of Utah and in Pittsburgh. Check our web site and you'll want to apply.



[whizbanglabs.com](http://whizbanglabs.com)

<sup>3</sup>For a good discussion on reputation capital in the Internet environment, see [www.firstmonday.org/issues/issue3\\_3/ghosh](http://www.firstmonday.org/issues/issue3_3/ghosh).

<sup>4</sup>For the Free Software Foundation's version of how various licenses work, see [www.gnu.org/philosophy/license-list.html](http://www.gnu.org/philosophy/license-list.html).