

## Open[Source]ing the Doors for Contributor-Run Digital Libraries

By Paul Jones  
University of North Carolina – Chapel Hill

What if you could wave a wand, in this very Harry Potter decade, and make libraries – at least digital libraries – more open, more easy 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? I believe that, thanks to the innovations from the Open Source community and perhaps more importantly the Free Software community, that we can have a contributor-run library at this very moment.

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 at [Metalab.unc.edu](http://metalab.unc.edu) which began in 1992 was that by removing nearly all barriers to submission and instituting instead 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 ( <http://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 that 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 ( <http://www.amazon.com/> ), and Barnes and Noble ( <http://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 <http://www.slashdot.org/> ), has instituted a rewards systems so that valued contributors and commenters accrue "karma" points which allow them to act as moderators of discussions and to rank comments and stories. Devices such as "karma" points serve as a hedge against trolls, group-take-overs, fakers, and the like.

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

By giving contributors and readers access to tools for evaluation, 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 ten 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 (see <http://sourceforge.net>), provide the tools that a project needs to get going on its own. Roadblocks to developers are removed by offering FTP and WWW 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. 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 so far of particular interest to digital library projects is that 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 been 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, which encourages community interaction and contributor involvement, digital libraries can open new horizons to new communities as well as greatly improve traditional services.<sup>5</sup>

---

<sup>1</sup> For a full description of the MetaLab Linux Software Archives and contributor demographics see Dempsey, Weiss, Jones and Greenberg. A Quantitative Profile of a Community of Open Source Linux Developers. [forthcoming in Communications of the ACM] or in draft at <http://metalab.unc.edu/osrt/develop.html>.

<sup>2</sup> Greenberg, J. Facilitating Author-Generated Metadata: Lessons Learned from an Analysis of Linux Software Maps (LSMs). [Draft report in progress.]

<sup>3</sup> For a good discussion on reputation capital in the Internet environment see Rishab Aiyer Ghosh's "Cooking pot markets: an economic model for the trade in free goods and services on the Internet" First Monday, Vol. 3 Issue 3 March 1998, **Error! Bookmark not defined.**

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

<sup>5</sup> For more on the Free Software and Open Source philosophies see Richard Stallman's groundbreaking work at the Free Software Project's philosophy pages <http://www.gnu.org/philosophy/> as well as Dibonna, Chris et al. Open Sources: Voices from the Open Source Revolution. O'Reilly and Associates, 1999.