Open source software for libraries in thirty minutes

This presentation describes what open source is and what role it can play in libraries, all in thirty minutes. On my mark. Get set. Go!

Community

Open source software is about community and it works because of the 'Net.

Open source software is about sharing one's expertise with others. It is about solving computing problems in an environment where others have the same problems. By working together the community solves common problems and grows through the process. Open source software works because the Internet facilitates communication across a large number of people. It flattens institutional hierarchies and enables diverse interests to coalesce into larger communities.

The open source software community works similar to the peer-review process in academia. Works are put forward, people examine the works and make suggestions for improvement, the works are edited, and the process begins anew. In the open source software world, it is said that, "Given enough eye-balls, all bugs are shallow."

Some people are leery about open source software because of the apparent lack of technical support. "Who are you going to call?" How satisfied are you now with the technical support you get from your commercial software vendors? How much do you annually pay for this support? Do you feel like you are getting your money's worth? Open source software is supported through mailing lists and some commercial agreements. If you demonstrate that you have tried to resolve your issues after reading the documentation and if you are still having problems, then mailing lists do work. Well-worded questions get responses

Free kittens

Open source software is as free as a free kitten.

You see a kitten and a sign next to it that says free. The kitten purrs. It plays with a ball of yarn. It is cute and fuzzy. You take it home. You then buy cat food. You take it to the veterinarian for shots. It claws your furniture. It then escapes into the night and returns the next day. While you didn't pay money for the kitten you did incur costs both financial as well as emotional. Open source software works the same way. While you don't pay money for it up front you do pay in terms of supporting hardware, emotional time and energy, and personnel. This is true for commercial software too, but with commercial software you have additional costs, the initial cost and the ongoing costs of licensing.

Investment in personnel

Investment in open source software is an investment in personnel because the learned skills are transferable.

More often than not open source software is standards-compliant; there are few "special features" in open source software that try to lock you into a particular product. There are few proprietary file formats in open source software. Files are usually saved in plain text (human-readable) formats, and binary file formats are well-documented. All of this means there are no "black boxes" in open source software and users of open source software need to learn a basic set of computing techniques: reading and writing plain text files, maintaining content in databases and writing reports against them, making content searchable with indexers, and transforming content into human-readable forms.

By combining standard file formats (such as MARC, XML), established computing technologies (such as relational databases and indexers), with open protocols (such as OAI, Z39.50 or SRW/U) it is more than possible to create modular digital library collections and services. It is then a library's responsibility to learn these computing tasks and mix-and-match them to meet their particular needs.

Put another way, because everything is "open" and there are no "black boxes", open source software enables you to control your computing environment, not the other way around.

Greater responsibility

Open source software requires a greater degree of computing responsibility than commercial software.

The use of open source software requires a greater degree of computer sophistication. Libraries, as a

whole, will need to know the fundamentals of relational database design. They need to understand the subtle differences between search with a database and search with an indexer. Libraries need to know how to read, write, and transform raw MARC records and XML files. Good reference librarians should know a bit about MARC records. Good catalogers are aware of the breadth and depth of the local collection. Good bibliographers understand the needs of the library patrons. None of these library specialties are islands unto themselves; each knows a bit about the work of the others. Similarly, libraries need to have a greater degree of computing knowledge, but no one is expecting every librarian to become a computer programmer.

Innovation

Open source software makes it easier for libraries to innovate.

People's expectations regarding data and information have dramatically changed with the advent of globally networked computers. Libraries never were the center of the information universe, and they are less so even now. People don't come to libraries to do their learning and scholarship as much as they did twenty years ago. People expect most of their information needs to be accessible through email messages and Web browsers.

At the same time there is still a need for the fundamental services of libraries, namely the collection, organization, preservation, and dissemination of data and information. Through the use of open source software libraries can mix and match computer technologies to provide these services. Libraries, as a community, can do this in lieu of a vendor who is only going to create products and services after they have become well-established and well-articulated needs. The vendor community is not geared nor designed to keep up with the changes in user expectations, only librarian expectations.

Open source software also allows libraries to go beyond collection, organization, preservation, and access. With the advent of the Internet individuals can do much of this without libraries. This is a good thing because it enables librarians to evolve the definition of librarianship to include the use data and information, not just its access.

Links

The open source software community is probably larger than you think.

Here are few general-purpose open source software tools:

- Apache - a Web server
- MySQL - a relational database
- Postgres - another relational database application
- Linux - an operating system
- Perl - a scripting/programming language
- Python - another scripting/programming language
- Swish-e - an indexer/search engine
- Lucene and Plucene - another indexer/search engine

Here are a few links pointing open source software communities:

- OSS4Lib - a mailing list and collection of library-related open source software
- SourceForge - probably the largest collection of open source software located in one place
- Electronic Information For Libraries - a foundation advocating the use of electronic resources by libraries and for library users in developing countries

These are library-related open source software projects of note:

- MARC::Record - Perl modules for reading and writing MARC records
- MyLibrary - Perl modules for managing links to just about any type of information resource
- DSpace - an institutional repository system
- Koha - an integrated library system
- Greenstone - a digital library system good for managing content
- YAZ - A Z39.50 client and server toolbox

Eric Lease Morgan
University Libraries of Notre Dame

February 26, 2006