

All things open

Things "open" abound. Open source software. Open access publishing. The open archives initiative. OpenURL. Some of these things are fundamental to the inner workings of the Internet. Others are a natural consequence of it. Some groups of people "believe" in things open with an almost religious fervor. At the other end of the spectrum are some people who see the same things as a drain on intellectual property. The key to progress lies in a middle ground. This presentation describes "all things open" in greater detail, elaborates on how they affect librarianship, and finally demonstrates some of their applicability in librarianship.

Open source software

Open source software is about: 1) community, 2) things "free as a free kitten", 3) an investment in personnel, 4) taking responsibility for your computing environment, and 5) greater opportunities for innovation.

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.

Open source software is as free as a free kitten. 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.

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.

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 needs to know 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.

Open source software makes it easier for libraries to innovate. Open source software 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.

Open Access Publishing

Open access publishing is about freely accessible content, usually scholarly in nature. Open access publishing is another opportunity for libraries to direct their own future. Open access publishing represents an opportunity to more actively participate in the scholarly communication process. There are opportunities for collecting, organizing, preserving, and disseminating content. Successful projects are project where libraries are proactive in the process and provide value-added services to the content. Examples include the creation of author and departmental citation lists, the syndication of content (via email messages, RSS feeds, the campus-wide portal, etc.), the creation of What's New? services, the demonstration of impact by listing Google PageRank integers or links from remote sites to local content, or the creation of reports such as "These people looked at my article this past month".

To some degree, a library's participation in open access publishing is akin to the collection of the "gray literature". Remember the gray literature? What is old is new again.

Open Archives Initiative

The Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) is a standardized and well-accepted method for sharing and harvesting metadata. With this goals in mind a protocol -- an agreed upon communication method -- was designed. The protocol would work over the Web (HTTP) and therefore exploit the client-server computing model. One computer (the metadata harvester) would send another computer (the metadata repository) commands, usually in the form of a URL. The second computer would then respond with an agreed upon XML stream. In OAI there are only six commands, called "verbs", a harvester can send:

1. Identify - Who are you?
2. ListMetadataFormats - What controlled vocabularies do you support?
3. ListSets - How have you organized your data?

4. ListIdentifiers - Return the unique keys of your records
5. ListRecords - Return a set of records
6. GetRecord - Return a single record

Using these commands is possible to collect the metadata from one or more repositories, copy it to one or more central caches, and provide enhanced services against them.

OpenURL

OpenURL is an ANSI/NISO standard (Z39.88) used to describe information resources in an unambiguous way in order to facilitate information services against those resources. In the form of link resolvers, OpenURL was and still is primarily used to address the "appropriate copy problem" created by our globally networked computer environment.

At the expense of increased complexity, OpenURL can provide the same functionality as the others "appropriate copy problem", but it does so in a more robust/non-proprietary fashion, and it provides for greater functionality. OpenURL is usually implemented in the form of HTTP GET requests (URL's), but they can be manifested as HTTP POST requests or XML streams sent as SOAP messages. The URL's are made up of three parts: 1) the scheme (almost always http or https), 2) the resolver (the host plus the full HTTP path to the resolver application), and 3) the ContextObject (the information resource in question). The ContextObject is the most complicated part of the URL and it is a set of key/encoded-value (KEV) pairs. These KEV's must denote at least one "Referent" (the information resource in question), but they can also denote things such as: the "Requester" (the name of the user requesting services), the "ReferringEntity" (like an article with a footnote pointing to the Referent), the "ServiceType" (get full text, find in a catalog, initiate an ILL request, etc.), the "Resolver" (just like the second part of the URL), and the "Referrer" (the resource that generated the ContextObject).

An URL would look like this when it is a part of an OpenURL system:

```
http://www.example.com/resolver?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.atitle=Librarianship:%20Profession%20of%20Opportunity&rft.date=1999&rft.volume=3&rft.issue=1&rft.spage=8&rft.epage=13
```

Opportunities ("open-unities")

All things open (open source software, open access publishing, open archives initiative, and OpenURL) and librarianship boils down to another word beginning with the letter "o" and that word is opportunities or maybe "open-unities".

Our society is moving (or has moved) from production-based society to service-based society. Increasingly the economic drivers are not the creation of steel, cars, and refrigerators. Instead it is based on services such as banking, entertainment, and computing. The output of these services are data, information, and knowledge. Couple this with almost ubiquitous networked communication and you can see how the "knowledge worker" has come into her own.

In the 1980's it was cool to be a librarian/knowledge worker. You had access to a computer and computer systems like DIALOG. You were the conduit for indexed bibliographic information. People paid a lot of money for your services. Our profession was almost required because few others had access to the necessary information so conveniently. In the 1990's this became less true because that same bibliographic information was now being distributed on CD's. For a while those CD's were only located in the library, but eventually they became networked and accessible from faculty member's offices. For the most part, people still had to come to the library to actually get the cited articles. Then came the World Wide Web and now a proverbial flood of data and information suddenly became available. With this increased accessibility also came a changing set of user expectations. People expect to get the information they need to do their work, their learning, their teaching, and their scholarship through their email programs and Web browsers. They expect to find it now with queries one or two words long, not through controlled vocabulary and fielded searches complete with Boolean operators. They expect access to full text. They expect it to be free and immediately accessible. These expectations are a far cry from the ways libraries have traditionally operated.

This does not spell the demise of librarianship. Instead it is the fuel for a metamorphosis. Through all things open librarians can transform themselves and the things they do to go beyond traditional library collections and services. Through all things open it is possible to enhance the meaning of librarianship, to empower users to a greater degree, and to more proactively support life long learning and the pursuit of truth through scholarship. Change is not easy. Think of the time and energy a caterpillar goes through to become a butterfly. Think of your own personal development from childhood through adolescence to adulthood. Librarianship may be going through the same thing, and it will continue to grow as long as the majority of us learn to adapt to the environment, build on what we have already created, and take advantage of our "open-unities".

Eric Lease Morgan
University Libraries of Notre Dame
March 24, 2006

(See also: <http://infomotions.com/musings/all-things-open/>)