
ROLE OF OPEN ACCESS, OPEN STANDARDS AND OPEN SOURCES IN LIBRARIES : A STUDY

Bobby Goswami Baruah

Abstract

The key concepts open access, open source and open standards are playing a great role today when we talk about borderless libraries. Immense benefits and importance to libraries are being examined. Greater benefits include lowering of cost, greater accessibility and better prospects especially for long term preservation of scholarly works. The paper mainly discusses on the importance and utility of the three concepts i.e. open access, open source and open standards in libraries emerging, which has helped to widened the gap in resource sharing, making it accessible in a far better way, limiting the expenses and above all standardizing which helps different systems to share their resources.

Keywords: Open Access, Open Source, Open Standards, Borderless Libraries, Open Source Softwares

1 Introduction

Open Access, Open Standards and Open Source Softwares have been receiving great attention lately in the world of libraries. This is mainly due to high growth of information on one hand and shortage of time on the other. Besides, there are many lot factors contributing today to the importance of the three concepts making libraries far more easier to be accessible to the end users. Open access is seen today as a possible solution mainly to the increasing price of serials and also as a way for governmental funding agencies to receive a better return on investment. Open source software is helping libraries mainly in lowering initial and ongoing costs, eliminating vendor lock-in and allowing for greater flexibility. The main advantage of open source software is that it is generally available for free. Lastly, open standards allow for interoperability to exist between diverse library resources and eases data migration between different systems. All three of these are important to libraries individually and they can be even more beneficial when they are used simultaneously.

2 Open Access, Open Source and Open Standards

2.1 Open Access :

In the recent years much discussions and initiatives are taken in the area of open access. Open access, a philosophy facilitates availability and distribution of scholarly communication freely, as a

means to solve the problem of inaccessibility primarily due to financial constraint particularly in the context of developing countries. Many scholarly literatures are freely accessible now without any hindrance. Open Access endeavours to reduce barriers to scholarly communication. The open access literature available in various forms like open access archives, institutional repositories, open access journals and off late open courseware. The availability of open source software has accelerated this development. Various different, though similar, definitions of open access exist with the *Budapest Open Access Initiative* definition being the most widely used (Goodman 2004). Other definitions include the *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, the *Bethesda Statement on Open Access Publishing*, and the *Washington DC Principles for Free Access to Science*. While there are multiple definitions and flavors of open access, open access basically calls for scholarly publications to be made freely available to libraries and end users. Willinsky (2003) identified nine flavors of open access. The flavors are: 1) e-print archive (authors self-archive pre- or post-prints), 2) unqualified (immediate and full open access publication of a journal), 3) dual mode (both print subscription and open access versions of a journal are offered), 4) delayed open access (open access is available after a certain period of time), 5) author fee (authors pay a fee to support open access), 6) partial open access (some articles from a journal are available via open access), 7) per-capita (open access is made available to countries based on per-capita income), 8) abstract (open access available to table of contents/abstracts, and 9) co-op (institutional members support open access journals).

The growth of the Open Access movement is basically in response to the enormous costs of many scholarly journals. With traditional journal publication methods it is not uncommon for an institution to have to pay for an article twice. First they pay scholars to produce the work and then the institution's library pays to purchase the work back from the journal publisher. Anderson (2004) is correct in saying that there is no such thing as free information and that there are costs involved in producing scholarly information. However, with the advent of new technologies and software programs, it is becoming increasingly less expensive to compile and distribute scholarly information. By using different funding methods and electronic delivery of journals, the costs can be absorbed by alternative means to subscription fees. One of the great benefits to open access is that libraries in smaller institutions or in economically disadvantaged areas around the world can have greater access to these scholarly resources.

Open Access helps to ensure long-term access to scholarly articles. Unlike articles that are licensed in traditional article databases, libraries and others can create local copies and repositories of these resources. Libraries, by working together to make repositories of open access literature, can ensure continued access to these scholarly publications into the distant future.

2.2 Open Source

Open Source Software is software that includes source code and is usually available at no charge. There are additional requirements besides the availability of source code that a program must meet before it is considered open source including: the software must be free to redistribute; derivative works must be allowed; the license can not discriminate against any persons; and the license cannot discriminate against any fields of endeavor. Software that is licensed under an open source license allows for a community of developers from around the world to improve the software by providing enhancements and bug fixes.

Libraries can realize many advantages by using open source software. One of the most obvious advantages is the initial cost. Open source software is generally available for free (or at a minimal cost) and it is not necessary to purchase additional licenses for every computer that the program is to be installed on or for every person who is going to use the software. Open source software not only has a lower acquisition cost than proprietary software, it often has lower implementation and support costs as well.

It is easier to evaluate open source software than proprietary software. Since open source software is typically freely available to download, librarians and systems administrators can install complete production-ready versions of software and evaluate competing packages. This can be done not only without any license fees, but also without having to stick to a vendor's trial period, evaluate a limited version of the software, or deal with the vendor's sales personnel. If the library likes an overall open source package but would like a few added features they can add these features themselves. This is possible because the source code is available. Even if a library does not have in-house expertise they can benefit from source code availability because another library may be able to provide them the fix or they can hire a consultant to make the changes that they desire. Fuchs (2004) points out that if a proprietary program "is deficient in some way [the user] must wait until the vendor decides it is financially viable to develop the enhancement – an event that may never occur." With open source software the user can develop the enhancement themselves.

Open Source Software allows for more support options. Proprietary software vendors often package service with the product. This is particularly true of proprietary library-specific software. When support from a vendor is inadequate it is an additional expense to purchase another tier of support, assuming that it is even available. Open source software allows for different vendors to compete for support contracts based on quality of service and on price. Access to the source code also allows for self-support when practical and desired.

The amount of vendor lock-in is dramatically reduced with open source software. The large initial costs often associated with proprietary software makes it difficult to reevaluate the choice of software when it does not live up to expectations. Proprietary software can lead to a single point of

failure. If a vendor goes out of business or decides not to support a program anymore there is often nothing an user can do. Organizations using the software could provide self support or other vendors can come in and fill the void left by the previous vendor if the program were available as open source software.

As discussed by John Lee, August 2007 in the article "Open standards and software for bibliographies and cataloging" in sourceforge.net the list below represents different open source softwares available. They are-

Open Source Software

Building Blocks

BibTeX and LaTeX

LaTeX is the traditional scientific document markup and typesetting system, still widely used today. BibTeX is LaTeX's bibliographic partner, which allows you to keep your bibliographic database(s) separate from your documents. Almost everything LaTeX- and BibTeX-related can be found on

CTAN

BibTeX can be entirely replaced with newer formatting engines, but the large number of .bst files (which specify formatting styles) available for BibTeX means that many people prefer to keep BibTeX, and replace only the front-end (with a graphical reference management application) and/or the data format (with an XML format).

m-bib and ConTeXt

ConTeXt is a TeX macro package similar to LaTeX, and has a built-in XML parser. The bibliographic module (m-bib) is a TeX-based replacement (for the most part) of BibTeX. Version 2 is intended to be rewritten to use XML data and formatting files.

ibibproc

From the site:

ibibproc is a bibliography processor similar to BibTEX. Its primary distinguishing features are:

- Internationalised. References can be multilingual and multiscript.
- Customisable and extendible. Styles, database formats, front-ends, and back-ends can be tailored and extended.
- Multiple front- and back-ends. Citations and references can be formatted for use with several document processing systems.

RefDB

Currently supporting PostgreSQL, MySQL, and SQLite:

RefDB is a reference database and bibliography tool for SGML, XML, and LaTeX/BibTeX documents. It allows users to share databases over a network. It is lightweight and portable to basically all platforms with a decent C compiler. And it's released under the GNU General Public License.

Pybliographer

Though Pybliographer is most frequently used as a graphical reference manager application (see below), it was designed as a general-purpose library: ...a simple framework that provides easy to use python classes and functions, and therefore can be extended to many uses (generating HTML pages according to bibliographic searches, etc).

bp

An elderly Perl 4 bibliographic formatting library. Input and output to many formats, character set conversion, etc.

MIT

There is an impressive collection of efforts at MIT that could have profound implications for bibliographic applications and data.

DSpace

DSpace is a newly developed digital repository created to capture, distribute and preserve the intellectual output of MIT.

DSpace currently uses qualified Dublin Core to store metadata, and PostgreSQL as its storage engine, but support of additional metadata standards (including MODS) is likely to grow as a result of the SIMILE research project.

SIMILE

Simile will leverage and extend DSpace, enhancing its support for arbitrary schemas and metadata, primarily through the application of RDF and semantic web techniques. The project also aims to implement a digital asset dissemination architecture based upon web standards. The dissemination architecture will provide a mechanism to add useful "views" to a particular digital artifact (i.e. asset, schema, or metadata instance), and bind those views to consuming services.

Cheshire II project

Note that parts of the web site are out-of-date. The current version of the system is available from the Cheshire ftp site.

Work is underway on the development of Cheshire III.

The Cheshire II project is developing a next-generation online catalog and full-text information retrieval system using advanced IR techniques. . . . The Cheshire II system was designed to overcome twin problems of topical searching in online catalogs, search failure and information overload as well as to provide a bridge between the purely bibliographic realm of previous generations of online catalogs and the rapidly expanding realm of full-text and multimedia information resources. The system incorporates a client/server architecture with implementations of current information retrieval standards including Z39.50 and SGML and XML.

Haystack

Another MIT offering.

Our research seeks to bring modern information management and retrieval technologies to the average computer user in order to make computers a more compelling place for users to interact with their information. Haystack looks into the use of artificial intelligence techniques for analyzing unstructured information and providing more accurate retrieval. We also deal with the modeling, management, and display of user data in more natural and useful ways.

BibDesk

BibDesk is a GUI BibTeX bibliography manager, making extensive use of Mac OS X-native technologies – including integrated PDF viewing, a citation-completion Service, etc. Capable of publishing RSS feeds of reading lists.

Pybliographer

Pybliographer is a tool for managing bibliographic databases.

The current stable version is Pybliographer 1. Work is in progress on Pybliographer 2.

Pybliographer 1 allows retrieving, editing, searching and citing bibliographic records using a GNOME graphical interface. Input and output to BibTeX, ISI, Medline, Ovid and Refer formats is provided, plus citation to LaTeX, LyX, HTML and plain text formats. Particularly good support for BibTeX / LaTeX and LyX. Medline queries direct from the application. Customizable BibTeX templates, Language / charset support. Pybliographer 1 is limited to databases of a few thousand records (this will be fixed in Pybliographer 2).

Open Office Bibliographic Project

Scheduled for inclusion in the next major release of the open source office project Open Office, this will be an integrated bibliographic storage and formatting module. The **OpenOffice technical committee** on XML file format specification for office applications is also relevant.

RefDB web client

RefDB focuses its attention on a relational database backend and support for markup languages, but includes a simple web client. Supports a variety of structured markup documents (including TEI, DocBook and LaTeX).

Kaspaliste

Bibliography manager for KDE. The OpenOffice people have this to say about it: Kaspaliste is a literature and knowledge database. It handles all kinds of books, articles, journals, web pages etc. But the database goes beyond simply storing bibliographical information. There is the possibility to create annotated links between pieces of information (like the content of a book chapter) and to group the links in categories. It is based on KDE and uses the Postgres relational database. It is a promising project but currently does not have import or export functions other than a BibTeX export. There are no links to other programs such as OpenOffice.

JabRef

JabRef is a GUI for managing BibTeX databases. ... JabRef works on all platforms and requires java 1.4.2

JBibtexManager

Import from ISI Web of Science, Medline/PubMed XML, Scifinder, OVID and INSPEC formats. Export to BibTeX, HTML and plain text. Searching, editing, sorting, duplicate detection, automatic key generation, customizable BibTeX templates, language support.

JReferences

Reference database with file and MySQL backends, oriented towards Docbook. Seems to be dormant.

gBib

gBib is a user-friendly editor and browser for BibTeX databases. You can use it also to insert citations inside a LyX document.

zNote

From the site:

...a web-based bibliography-management tool built with Zope and XML.

zNote is intended to ultimately be a replacement for tools like EndNote, ProCite, and to a certain extent, BibTeX. It uses a hierarchical XML data format which is more flexible than flat data, and it works using a set of pretty simple DOM calls to format, edit, etc.

CMFBibliography

A basic BibTeX-centric bibliographic plone module. Designed for scientific users (this is true of many of these applications, in fact).

biblioz

Plone-based personal book list and reference manager with support for journals and automatic retrieval of data from Amazon.

CitationManager**Tyrannio**

Z39.50-capable book cataloging and information retrieval application with command-line, GUI and web front ends. Support for barcode scanning with the :CueCat scanner. PostgreSQL backend.

Bookcase

Book collection manager for KDE 3.x. File-based XML system: there's no database backend. There are plans for extension to other types of collections (CDs, for example), and for Z39.50 client capability. Has it's own simple DID.

Bookcase is a KDE application for keeping track of your book collection. Ultimately, I'd like it to be similar in capability to AVCataloger or Readerware, although it's still got a ways to go.

Koha

Made in New Zealand by Katipo Communications Ltd. and maintained by a team of volunteers from around the globe, the Koha system is a full catalog, OPAC, circulation and acquisitions system.

Greenstone

Greenstone is a suite of software for building and distributing digital library collections.

OpenBiblio

OpenBiblio is an easy to use, open source, automated library system written in PHP containing OPAC, circulation, cataloging, and staff administration functionality. The purpose of this project is to provide a cost effective library automation solution for private collections, clubs, churches, schools, or public libraries.

PhpMyBibli (in French)

Web-based library cataloging and circulation database, including administration and OPAC functionality. Apache/MySQL/PHP-based.

iVia

Web-based library cataloging database, with an emphasis on electronically-stored collections ('virtual library system'). In addition to human-created catalogs, it crawls the web to do automated catalog creation.

2.3 Open Standards

Pountain (2003) defines an open standard as "a standard that is independent of any single institution or manufacturer, and to which users may propose amendments." This definition is a good starting point, but in reality the term "open standard" means different things to different people. Three key characteristics of open standards identified by Coyle (2002) are 1) that anyone can use the standards to develop software, 2) anyone can acquire the standards for free or without a significant cost, and 3) the standard has been developed in a way in which anyone can participate. When a standard has the first two of these characteristics (the ability to use the standard and to obtain it with out a significant cost) it can be said to be an open standard in an utility sense. That is to say that an open standard is a standard that is not encumbered by a patent, does not require proprietary software, and can be utilized by anyone without cost. Proprietary standards can sometimes be expensive and it may be cost prohibited to purchase access to a proprietary standard if it is ever needed. Many people consider a standard to be sufficiently open as long as it is open in a utility sense. Others take this a step further and consider a standard to be open only if the process meets the criteria of being created and modified in an open process as well. An example of a standard that fits the definition of a standard that is open in utility but not in process is XHIML. In order to help

develop the XHTML specification one has to be a member of W3C. In order to become a member of W3C businesses pay between \$5,000 and \$50,000 per year (Coyle 2002). Conversely, Dublin Core is a completely open standard that is open both in utility and in process. All one has to do is show up and participate in order to contribute to the development of Dublin Core.

It is important for libraries and other cultural institutions to ensure long-term access to digital information. The rapid growth in digital technologies has led to new and improved applications for digital preservation. However at the same time it has also led to some problems as well. Two of these problems are obsolescence and dependency issues. The obsolescence problem is caused by the advances in hardware and software making many computers obsolete within three to five years (Vilbrandt et al. 2004). Dependency problems can arise if tools that are needed to communicate between systems or read file formats become unavailable. In order to account for obsolescence and dependency problems organizations must be able to migrate data into new systems. Data migration, however, cannot occur without access to data file formats.

Properly created open standards for file formats are less likely to become obsolete (Vilbrandt et al. 2004) and are more reliable and stable than proprietary formats (Breeding 2002). In the event that an open standard file format does become obsolete, having access to the file format would allow anyone to easily, and legally, create a data conversion utility. File formats that use open standards can assist in long-term archiving because they allow for software and hardware independence. Open standards help alleviate issues caused by obsolescence or dependency problems since files created in formats that adhere to open standards are "more likely than proprietary formats to be readable twenty or fifty years from now" (Baker 1999). This allows for greater flexibility and easy migration to different systems in the future.

The use of open standards can help assure interoperability of diverse systems. There are various software packages that are being used to create digital libraries, online library catalogs, and other resources that libraries relay on. These various systems need to be able to interact in order to provide the best possible service to patrons. The way to make certain that these diverse systems, and any future systems, can communicate with each other is by using open standards to help achieve the "free flow of information through interoperability" (The Open Group 2005).

Many different organizations are advocating open standards. One of the most prominent organizations is The Open Group which created the *Developer Declaration of Independence*.

The hope is that the *Developer Declaration of Independence* will help pull together the information technology industry in support of open standards. Some library-centric initiatives, including the Open Archives Institute (OAI), also support open standards. OAI's mission is to develop and promote "interoperability standards that aim to facilitate the efficient dissemination of content" (Open Archives Institute 2005). OAI has created a Protocol for Metadata Harvesting (OAI-PMH) that provides an application-independent interoperability framework based on metadata harvesting. Other common open standards for information retrieval relevant to libraries include Digital Object Identifier System (DOI), Dublin Core Metadata Initiative (DCMI), and OpenURL.

While open standards have garnered increased attention in libraries recently, the use of open standards in librarianship is not new. The use of open standards in librarianship can be traced all the way back to the first American Librarian Association meeting in 1877 when the dimensions of the card catalog were standardized to 7.5 x 12.5 centimeters (Coyle 2002). A more modern example of an open standard used by libraries is the Machine-Readable Cataloguing (MARC) record. Other common open standards for bibliographic data include Metadata Object Description Schema (MODS), Metadata Encoding & Transmission Schema (METS), and the XML Organic Bibliographic Information Schema (XOBIS).

Putting Open Access, Open Source, and Open Standards Together

Open access, open source software, and open standards each individually offer a number of significant benefits to libraries. When they are combined the results can be even greater. Open source and open standards can help libraries provide patrons with easier access to open access materials and other resources. There are literally thousands of open access titles available and without open standards it would be very difficult to find what one is looking for or to view various articles. Imagine the difficulty, and costs involved, in maintaining a library's information technology infrastructure if each electronic journal required a separate, proprietary piece of software to read or search the journal. Open standards make it possible to create interoperable systems to access the literature in various open access journals seamlessly.

Open standards and open source can help preserve long-term access to open access and other types of electronic journals. Libraries working together can use open source software such as LOCKSS to ensure continued access to these scholarly publications long into the future. LOCKSS (short for "Lots Of Copies Keeps Stuff Safe") is a system that caches copies of digital collections around the world. As current computers, software, storage media, file formats, and other types of information technology become obsolete, it will be necessary to

migrate open access articles and other data to new systems. Without the assistance of the software manufacturer (who may or may not even still be in business, let alone willing to help) proprietary software and file formats may make migration practically impossible. By utilizing open source software and open standards from the beginning, libraries can assure that this type of systems migration will be possible years down the road.

Not only has the growing cost of serials caused libraries to drop journal subscriptions, it has also factored into a 26% decrease of monograph acquisitions by the typical research library between 1986 and 1999 (Create Change 2002) . Library budgets can be reallocated to monographs and other areas because of the lower costs typically involved with open access, open source, and open standards.

Open Source Solutions are the emerging trend today and started dominating the information industry. In particular, the libraries are the great beneficiaries of these open source technology. The following are some the benefits of open source systems are

- Enables librarians and libraries to have more control over their information systems and services.
- Provides the means to explore and implement ways of doing librarianship without the reliance on a software vendor.
- Illustrates how librarianship can facilitate library services and collections that go beyond book lending.
- Enables libraries to meet the quickly changing information needs, desires, and expectations of users.

Fortunately for Digital Library Applications, there are lot of open source systems available namely DSpace, Green Stone Digital Library (GSDL), e-prints, fedora, Linux, XML etc. They are not only cost effective but also capable of handling DL challenges like content/metadata management, information dissemination etc. Extensibility is achieved from plug-ins that is available under GNU public license. The types of open source systems may include

- Operating Systems
- HTTP Servers and clients
- SMTP Servers and clients
- Database Systems
- Programming Languages
- XML Processors and other application

Conclusion

These benefits of open access, open source, and open standards are numerous. The benefits include lower costs, great accessibility, and better prospects for long-term preservation of scholarly works. Libraries should embrace all three of these concepts now and in the future. By supporting open access, open source, and open standards libraries not only can help ensure that their current and future patrons will have easier and more comprehensive access to scholarly research, they will also be helping other libraries around the world, including those in disadvantaged areas, to have access to important scholarly research. Open Source movement has tremendously influenced Information industry, in particular the Digital Library Environment where the Information Systems and services are inevitable. This has helped the Library and Information. Centers in developing User Interface to render web based services to the patrons. Wherever the network infrastructure and the target users are in place, the design and development of IR user interface using Open Source system for a specific application is very much possible.

References

- 1 Anderson, R. 2004. Open access in the real world: confronting economic and legal reality. *College and Research Library News* 64 (4) . [Online] . Available: <http://dlist.sir.arizona.edu/archive/00000351/> [Accessed March 25, 2005] .
- 2 Baker, T. 1999. *TIAC White Paper on Appropriate Technology for Digital Libraries*. Bangkok: Technical Information Access Center.
- 3 Breeding, M. 2002. Preserving digital information. *Information Today* 19 (5) : pp. 48-49.
- 4 Coyle, K. 2002. Open source, open standards. *Information Technology and Libraries* 21 (1) : 33-36.
- 5 Create Change. 2002. Coping Strategies. [Online] . Available: <http://www.createchange.org/librarians/issues/coping.html> [Accessed May 11, 2005] .
- 6 Fuchs, I. 2004. Learning management systems: are we there yet? Syllabus. [Online] . Available: <http://www.campus-technology.com/article.asp?id=9675> [Accessed: April 1, 2005] .
- 7 Goodman, D. 2004. The criteria for open access. *Serials Review* 30 (4) . [Online] . Available: <http://dlist.sir.arizona.edu/archive/00000798/> [Accessed: March 11, 2005] .
- 8 Open Archives Initiative. 2005. Organization. [Online] . Available: <http://www.openarchives.org/organization/index.html> [Accessed: March 17, 2005] .
- 9 Pountain, D. 2003. *The Penguin Dictionary of Computing*. New York: Penguin Putnam.
- 10 The Open Group. 2005. Developer Declaration of Independence. [Online] . Available: <http://>

www.opengroup.org/declaration/declaration.htm [Accessed: March 29, 2005].

- 11 Vilbrandt, T., et al. 2004. Cultural heritage preservation using constructive shape modeling. *Computer Graphics Forum* 23 (1) : 25-41.
- 12 Willinsky, J. 2003. The nine flavors of open access scholarly publishing. *Journal of Postgraduate Medicine* 49: 263-267.

ABOUT AUTHOR

Ms. Bobby Goswami Baruah presently working as a librarian in Guwahati College of Architecture, Guwahati has completed her masters in Library and Information Science from Gauhati University in 2003 and PGDLAN from University of Hyderabad in 2006. She has worked as a guest lecturer in the department of library and information science, GU from 2003 till 2006.