

Prospects of Open Source Software in LIS Area of Assam

Sanjay Kumar Singh

Prasanta Kumar Deka

Abstract

The choice of library management programme is constituted by social, economic and or political factors that result in selection of a system that is not suitable for the library requirements, characteristics and functions. Open Source Software or OSS has revolutionised the modus operandi of library software for library automation. OSS is quickly becoming a solution, owing to the freedom to copy, modify and distribute it and freedom from contracts as well as for greater opportunities for interoperability with other application. The paper mainly discusses about different OSS. It is a relatively new idea for library automation. People often advocate OSS because it is free. Some of the selected OSS is highlighted. OSS is an umbrella term used to refer to a wide range of software products, services and communities.

Keywords : OSS, Koha, obiblio, DSpace, Greenstone, Linux, MySQL,

1. Introduction

Computers are capable of introducing a great degree of automation in operations, functions since they are electronic, programmable and are capable to control over the processes being performed. Open Source Software (OSS) has revolutionized the modus operandi of library software for library automation. There is a paradigm shift in the role of libraries from mere storehouse of information/knowledge to knowledge disseminations to knowledge providers and further transformed as digital library. Information / knowledge itself is of no value. It is the use of information that makes it valuable. The role of computers and their associated peripheral media are being increasingly used in library and information services for acquisition, storage, manipulating, processing and repackaging, dissemination, transmission, and improving the quality of products and services of library and information centres. There is a global change in the libraries, which indicate a shift from owning technology to using technology. OSS is currently one of the options preferred by some of the libraries, which are not financially sound because of the facilities available in OSS. This new development towards the use of OSS in libraries is also incorporated in LIS curriculum in the universities of India. Keeping in view of this increasing use of OSS in libraries and easily accessibility, it is freely available in the net, which can be downloaded for use.

The modern libraries and information centres facilitate free communication because access to information has become a fundamental right of the clientele. The automation is economically feasible and technologically required in libraries to cope up with the requirements of new knowledge, the

enormous increase in the collection of materials, problems of their acquisition, storage, processing, dissemination and transmission of information. The capabilities of computer associated peripheral media and its application in library activities and services led to a highly significant quantitative and qualitative improvement in on-line technology.

2. Library Automation

Computers and automation have become an established fact for the library operations, and this glamour machine of the information age, is a unique solution for storing and distributing large amount of information. Libraries need computers, application software and trained manpower for computerizing their activities. To run any academic library or information centre systematically, computer based automated facilities are now the most effective system and necessity.

As number of members are increased in library, cost of information materials, services and growth of information or information explosion, the budget of the libraries is also raised. That also allows us to automate the library activities and make maximum utilization of the library funds.

3. Areas of Automation

With the application of ICT in the area of libraries there has been tremendous improvement in the library services offered to the users by libraries. Following are the main areas of library automation:

1. Library administration and management.
2. Library house keeping operation, and
3. Information Retrieval System.

4. Library Software Packages

To offer the complete satisfaction of users and perform the library activities, we must select competent and suitable software which can meet out our requirements or can be developed on contracted basis by any software company or can be developed by professional of the institution keeping in view the requirements of the library. An increasing number of library software companies and their attractive advertisements / propaganda's confuse the librarians/ libraries which software will very much meet their needs. Libraries and information centres must keep in mind some basic aspects before selection of library software.

1. A library automation committee is to be constituted which includes library and computer software/ hardware experts.
2. Prepare a list of the library's areas, activities, services and functions, which are to be automated.
3. A profile of library software's is to be made.

4. Consult with some organization/ libraries about software and their functions, which are already, automated their services. The experiences of the other librarian who have used software are more valuable than the assurance of the manufacturers.
5. Let the vendors demonstrate their product. Library and computer software experts should prepare an observation report of library and software facts, which meet or how much meet or do not meet our requirements and submit it to the chairman library automation committee.
6. Consider the services after installation, but do not believe on unjustifiable assurance, made by companies' representatives and make agreement of all fact.
7. Consider the market reputation of manufacturing company, software or vendor.
8. Software selection is a very complicated issue, on the observation of experts, the discussion should be made by the selection committee and most suitable in regard of flexibility, capacity, expandability, security, economically, users friendly, module based and updated with the latest technology is to be procured. There are some leading names of the software packages which are available in the market may be competent with our needs.

Some of the leading library software are LIBSYS, SOUL, VTLS, e-Granthalaya, SANJAY, MAITRAYEE, OASIS, SLIM ++, MINISIS, WINISIS, CDS/ISIS, Open Source Software, etc. Because of cost factor, to overcome the problems of library automation software one can think of using open source software, which is freely available on the net and can be easily downloaded.

5. OSS and the ILS

Libraries are a lot about the collection, storage, organization, dissemination, and sometimes evaluation of information and knowledge. With the advent of computer technology in libraries many of these processes have been implemented through a library's integrated library system or ILS. The primary purpose of the ILS seems to be the management of lists of MARC records and the facilitation of services against these lists. The online public access catalogue (OPAC) provides searching functions against the list. Cataloguing provides functions to add and edit items on the list. Acquisitions provide some accounting functions. Reserve room modules, circulation modules, and interlibrary loan modules allow the locations of items to be temporarily moved from one place to another. Serials modules provide functions for inventory control.

6. Some Selected OSS

The principles and practices of open source software are very similar to the principles and practices of modern librarianship. Both are value free and having equal access to data, information, and knowledge; value the peer review process; and advocate open standards. The use of open source

software in libraries enables libraries to have greater control over the automation. No one is saying that librarians should know how to compile relational database programs. If librarians want to be privileged in the fields of information and knowledge, then libraries need to know how to take advantage of the current technology that makes this happen. Open source principles, practices, and results can assist librarians in their fulfilment of day-to-day activities as well as the goals of the profession. OSS shows a method for librarians to maintain control over automation. There are numerous OSS directly relevant to the libraries. Some of the important OSS are given below:

6.1 Apache (<http://httpd.apache.org/>)

It is the most popular Web (http) server on the Internet and a standard open source piece of software. It is made up of many modular parts to create a consistent whole. The latest version of Apache is 2.0.

6.2 Linux

It has come to be the conventional piece of open source software. It is an operating system just as Windows is an operating system. It is a software that handles the most basic operations of a computer like keyboard input, display via monitors, reading and writing to the hard drive, and getting input/output from network devices. Linux came about when a college student, Linus Torvolds, desired Unix (another operating system) functionality on this inexpensive Intel-based hardware. Today Linux is one of the most powerful operating systems ever produced and definitely getting better all the time. It appears to be uncertainty that Microsoft believe at least some threat by this new developments.

6.3 MySQL (<http://www.mysql.com/>)

It is a relational database application, pure and simple. It is said to be the world's most popular open source database. MySQL undoubtedly has a wide support in the Internet community. MySQL runs on any computing platform and it has been used to deal with millions of records and gigabytes of data. Speedy and vigorous, it supports the majority of people's relational database needs.

6.4 Perl (<http://www.perl.com/>)

It is a programming language. Originally written to handle different systems administration tasks, Perl's strength lies in its ability to manipulate text. Perl matured through the era of Gopher but really started becoming popular with the advent to CGI scripting. Perl has been ported to just about any computer operating system, has one of the largest numbers of support forums. Perl is mature and very robust. Other very good programming languages exist and can do much of what Perl can do.

6.5 Swish-e (<http://www.swish-e.org/>)

It is a simple indexer/search engine. Swish-e indexes individual files on a file system, files retrieved by crawling a website, or a stream of content from another application such as a database. The indexing half of swish-e is able to index specifically marked up text in XML and HTML as fields for searching later. The same application that creates the indexes can be used to search the indexes. Swish-e supports relevance ranking, Boolean operations, right-hand truncation, field searching, phrase searching, free text searching, and nested queries. With swish-e librarians may perhaps create their own indexes.

6.6 Xsltproc and xmllint (<http://xmlsoft.org/XSLT/>)

Xsltproc and xmllint are very useful applications for parsing and processing XML file. By feeding xsltproc an XSL style sheet and an XML data file one can transform the XML data file into any one of a number of text files whether they be SQL, (X)HTML, tab-delimited files, or even plain text files intended for printing. Xsmllint is a syntax checker. Given an XML file, xmllint will check the validity of your XML files. With xsltproc and a plain o' text editor we can learn about XML.

The following is a list of more library oriented specific open source software distributions:

6.7 DSpace (<http://www.dspace.org/>)

It is a management system for institutional repositories created by the MIT Libraries (Massachusetts Institute of Technology) and Hewlett-Packard (HP) in 2002 and a tool designed to allow institutions, such a libraries to collect, archive, index, and disseminate the scholarly and intellectual efforts of a community. It is disseminated under an Open Source Licence (BSD Berkeley Software Distribution), an OSS licence with certain restrictions not included in the GNU GPL licence. This type of licence allows the use and modification of the source code and the commercial use of the product. It is primarily used to capture bibliographic information describing articles, papers, theses, dissertations, preprints, technical reports, images, videos, etc. and helps to digitalize an organization's scientific production. Once entered into the system, DSpace indexed the content and provide a way to link to the originals. It plays well with open standards such as XML and OAI-PMH. If a large number of institutions of higher education where to capture their intellectual output using DSpace or some other similar piece of software, then access to scholarly materials would be greatly increased and readily available. The main characteristics of Dspace are that it supports any type of material i.e. documents, theses, images, audio, video, etc.; the records are saved in Dublin Core format; able to interoperate with other systems in the organization; all records have a persistent identifier; allows format migration towards IFF, SGML, XML, AIFF, PDF.

6.8 Greenstone (<http://www.greenstone.org/>)

It is a tool for creating and managing digital library. Running on Windows as well as various Unix platforms, it provides the means to easily create searchable and browsable interfaces to digital library collections via the web. It also enables implementers to save their collections to CDs/DVDs. Thus the digital library collections can be distributed to people with poor or not Internet access. Greenstone knows how to create collections from standard file formats such as HTML files, email messages, PDF documents, JPEG and GIF images, Word documents, as well as plain text files. If the sets of files are well structured, then Greenstone will create things like A-Z list of resources, and field searchable interfaces. Greenstone's look and feel can be customized through an HTML-like template language. The mailing provides more than adequate support and the documentation is from beginning to end.

6.9 Koha (<http://www.koha.org/>)

It is an integrated library system with a growing user community. It is written in Perl and using MySQL as the underlying database, Koha makes it simple to create and manage a small to medium-sized integrated library system. Equipped with acquisitions, cataloguing, circulation, and searching modules it provides much of the functionality of traditional online catalogues. With the recent implementation of its Z39.50 interface, it is easy to enter ISBN numbers into the system, locate MARC records, and have those records added. The user and system interfaces are simple and imaginative, but not very customisable. For many libraries, the catalogue is the centerpiece of the operation. Koha represents a major step in providing a catalogue that is functional and usable. As long as support continues, it is expected Koha to be more feasible option for big library collections. The obstacle is not technology but it is time and effort.

6.10 MARC Record (<http://marcpm.sourceforge.net/>)

This Perl module is the tool to use when reading and writing MARC records. It is very well supported on the Perl4Lib mailing list, and a testament to the module's abilities is its incorporation into things like Koha and Net: Z3950. On the other hand, learning to use MARC: Record will not only improve programming abilities but it will educate on the intricacies of the MARC record data structure, a structure that was designed in an era of scarce disk space, non-relational databases, and little or no network connectivity.

6.11 MyLibrary (<http://dewey.library.nd.edu/mylibrary/>)

It is a user-driven, customizable interface to sets of library resources - a portal. technically, MyLibrary is a database-driven website application written in Perl. It requires a relational database application as a foundation, and it currently supports MySQL and PostgreSQL.

MyLibrary grew out of a number of focus group interviews where people said they were suffering from information overload. To address this problem, MyLibrary takes three essential components of librarianship and tries to create relationships between them through the use of common controlled vocabularies. Like a library catalogue, MyLibrary provides the means to create collections of resources and classify them. Unlike a library catalogue, the system also allows librarians to be classified in the same manner. By sharing a common set of controlled vocabulary terms relationships between resources, patrons, and librarians can be made.

6.12 PMB (<http://www.sigb.net>)

It is a library automation system created in France in 2002 under the CeCILL licence. It is an OSS licence created by the French government which offers the same conditions and freedoms as the general public licence (GPL) for open source software. PMB is designed for medium-sized and large libraries, and can also manage networks of libraries integrated in a collective catalogue. The cataloguing format used is UNIMARC but it incorporates format conversion and introduction systems (USMARC and XML, among others). Some of the main characteristics offered by the system are management of authority data (by author, publisher, collection); management of a thesaurus of subject areas; use of the Z39.50 protocol; automation of the DSI; control of serial publications; management of journal summaries and storage; and system for producing back-up copies.

6.13 Openbiblio (<http://obiblio.sourceforge.net>)

Openbiblio or obiblio is an integrated library management system, created in 2002 under General Public Licence (GNU- GPL). Obiblio is designed for small libraries. The catalogue format is UNIMARC but records may be imported from other compatible formats. Some of the important characteristics are user alerts via pre designed messages; control of fines for late returns of documents; and statistical models for the use of the library and the material.

6.14 Potnia (<http://potnia.sourceforge.net>)

It creates thematic directories, for the management of scientific resources, journals, publications, and so on. The Directorate General of Universities of the Spanish Ministry of Education, Culture and Sport created it in 2003. The structure and description of the records in Potnia are compliant with the Dublin Core metadata set. Some of the important functions are search for records. In addition to basic searches, an advanced search option is included which allows definition of the fields to be searched by title, key word, description or subject and combinations using the Boolean operators and, or, and not. Administration of the system, for adding, modifying or deleting records in the data base, and for managing the list of

subjects. Home page, showing the list of subjects of the records in the data base. Under each subject heading, the records classified under the subject can be consulted, and details of each one may be accessed.

6.15 Alfresco (<http://www.alfresco.com>)

It is an ECM, an Enterprise Content Manager. Among its applications is document management. Members of Documentum and Interwoven created it in 2005. Alfresco is open code software, able to access the source code for the user community and provide high quality service for the development of the system. The commercial benefit is provided by the system's support package, including manuals, training and maintenance. There is also an area for the user community to contribute additional features to the program. These modifications are under the Alfresco Public Licence (<http://forge.alfresco.com/>). Some of the important characteristics are deposit of electronic documents in their original format like doc, xls, tif, html, jpg, pdf, video, audio, cad, etc.; document visualizes; groupings by folder and classification chart; system of information retrieval, basic and advanced search, file, folder and contents search; management and control of work flows; management of users and levels of access; and history of versions.

6.16 E-Prints (<http://www.eprints.org>)

It is a system for the creation of institutional repositories of scientific production, theses, reports, etc. created in this community. The University of Southampton created it in 2000. E-Prints uses the XML and Dublin Core metadata formats based on the OAI-PMH. It is thus able to support different types of documents and data formats for storage and provides open access to their contents. Some of the characteristics are Dublin Core metadata format; repository of all types of material: text, image, etc.; content syndication; includes a bibliography manager; and basic and advanced search options.

6.17 Tematres (<http://www.r020.com.ar/tematres>)

Tematres is a web application for the management of thesauri, created in Argentina in 2004. This program is also designed for the creation of web browsers, digital library directories, and control of documentary languages. Some of the important characteristics are that the related terms by equivalence, hierarchy and/or association; unlimited quantity of related terms, hierarchical levels and non-preferred terms; multiple scope notes, historical notes and bibliographical notes per term; control of repeated terms; systematic and alphabetic browsing; exportation formats into XML, Dublin Core, RDF SKOS-Core, Zthes, etc; and general search engine.

6.18 Refbase (<http://refbase.sourceforge.net>)

It is a software management package for bibliographies and bibliographical quotations that works under a multi-user web interface. It was created in Germany in 2002 under General Public Licence (GNU-GPL). Refbase is designed to create collections of bibliographical references imported from different academic databases and to extract them in quotation form for text bibliographies. Some of the important characteristics are basic or advanced search; content syndication; quotation format: HTML, RTF, PDF, LaTeX; import formats: BibTeX, Endnote, RIS, RefWorks, PubMed, ISI Web of Science, CSA Illumina, Copac, MODS XML; and Export formats: BibTeX, Endnote, RIS, MODS XML, XML of the Library of Congress and Open Document.

7. OSS: Historical Perspectives

OSS describes the future use of software and methods for its distribution. Depending on the perspective, the concept of OSS is a relatively new idea, being only six to seven years old. On the other hand, the GNU Software Project- a project advocating the distribution of free software - has been operational since the mid '80s. Consequently, the ideas behind OSS may have been around longer than we think. It begins when a man named Richard Stallman who worked for MIT in an environment where software was shared. In the mid 1980s, Stallman resigned from MIT to begin developing GNU - a software project intended to create an operating system much like Unix. (GNU is pronounced guh-NEW and is a recursive acronym for GNU's Not Unix.) His desire was to create free software, but the term free should be equated with freedom, and as such people who use free software should be:

- 1) Free to run the software for any purpose,
- 2) Free to modify the software to suit their needs,
- 3) Free to redistribute the software gratis or for a fee, and
- 4) Free to distribute modified versions of the software.

The idea of free software was, and still is, a difficult idea for many people to understand. The process is similar to the scholarly communications process because open source software goes through a sort of peer review process. Fellow programmers examine a program's source code, find flaws, and suggest improvements.

8. Prospects of OSS in Assam

The state of Assam has largest number of educational institutions in the North East India with five universities *viz* Gauhati University, Guwahati; Dibrugarh University, Dibrugarh; Assam Agricultural University, Jorhat; Assam University, Silchar; and Tezpur University, Tezpur. It has one IIT at Guwahati; three Medical Colleges at Guwahati, Dibrugarh and Silchar; three Engineering Colleges at Guwahati,

Jorhat and Silchar (NIT); one each Pharmaceutical Institute, Nursing College, Ayurvedic College, Dental College in Guwahati. The state is also having number of Engineering, Polytechnic and other academic institutions and more than 300 degree colleges, some of which are pursuing Master Degree and Higher Secondary courses in Assam. Almost all the mentioned educational institutions are having libraries except few of them, which are High and Higher Secondary Schools. Besides these the state has 23 District Libraries, 14 Sub-Divisional Libraries and 4 Branch Libraries.

Libraries in Assam face several shortcomings for which the modernization/ automation of library environment does not take place at a marked level. Though the reasons attributed to such a plight can be manifold, the primary reason is the inadequate provision of fund and lack of positive attitude towards library development. The main problems of library automation in Assam are trained manpower and to some extent the negative attitude of authorities also. Due to economic recession, libraries are hard-pressed in their budgets. However, in most cases the library professionals are not conversant with the environment of library automation. It may be mentioned that the prices of commercial library automation software are very costly that cannot be procured by all the academic libraries of Assam because of different reasons. And there is some locally developed software also, but the credibility of those packages cannot be determined because the local package does not fulfill all the required modules for library automation and are not following any internationally accepted format as well as guideline. Though these are easily and cheaply available in the markets of Assam by paying just Rs. 10,000/- to 15,000/- only. Thus saving some money of the library but not fulfilling the library requirements.

The solution for this problem is that most libraries and institutions of Assam do not have the necessary computing expertise to make the solution a reality. There does not seem to be a critical mass of people working in libraries who know how to write computer programs. Consequently library processes and computing environments are often held hostage by library-specific software vendors. At present only Guwahati Medical College Library at Guwahati is using **Koha** an OSS for library automation. This is the beginning of OSS movement in Assam. In near future other library may think to adopt OSS. So there is a bright prospects of OSS in Assam.

9. Suggestions

To overcome from the syndrome of confusion about which library automation software package will be useful for their library and which one to be procured, few suggestions are given below that may be useful for the library who are planning for automation.

1. Hardware and software including library automation software should be acquired as per the latest configuration and with maximum storage capacity.
2. Although the INFLIBNET is arranging workshop for library professional for implementation of library automation programme, opinion from the experts who have already computerized

the library activities and services may be taken for successful implementation of the computerization of library services.

3. Library professional at senior level should visit any computerized libraries to have better idea of computerization of library operation and services.
4. User-friendly good library application software should be procured either from the renowned company or from the INFLIBNET.
5. A common software SOUL may be used by all the libraries, which can be useful for resource sharing and are easily available through INFLIBNET.
6. As most of the library software are costly and beyond the reach of small and medium libraries. The library that cannot afford the costly software can download software available freely from the Internet. For minimization and cost effectiveness, such libraries can use OSS for managing the library.

10. Conclusion

A library and Information System cannot be said fully automated unless the emerging technology is not incorporated into the library services. Therefore many libraries are in queue for adopting such type of emerging technology, which will significantly improve the customer service and enhance patron satisfaction. Keeping in view the economic constraints of the majority of libraries, OSS has a very good prospect for automation of libraries and information centers in Assam. The open source movement is playing an important and vital role in software development in the 21st century. There are enormous benefits of OSS. People often advocate OSS because it is free. OSS is an umbrella term used to refer to a wide range of software products, services and communities, the most famous of these being the GNU/Linux operating system No doubt this OSS movement will certainly pave way for the libraries of this remote corner of India also.

References

1. <http://dewey.library.nd.edu/mylibrary/>
2. <http://edoc.hu-berlin.de/conferences/bobcatsss2008/>
3. <http://httpd.apache.org/>
4. <http://informations.com/musings/biblioacid/>
5. <http://marcpm.sourceforge.net/>
6. <http://obiblio.sourceforge.net/>
7. <http://potnia.sourceforge.net/>
8. <http://refbase.sourceforge.net/>

9. <http://www.alfresco.com/>
10. <http://www.dspace.org/>
11. <http://www.eprints.org/>
12. <http://www.greenstone.org/>
13. <http://www.koha.org/>
14. <http://www.mysql.com/>
15. <http://www.perl.com/>
16. <http://www.r020.com.ar/tematres/>
17. <http://www.sigb.net/>
18. <http://www.soft-libre.net/>
19. <http://www.swish-e.org/>
20. <http://www.ub.edu/biblio/>
21. <http://xmlsoft.org/XSLT/>

About Authors

Dr. Sanjay Kumar Singh, Lecturer, Department of Library and Information Science, Gauhati University, Guwahati-781014, Assam.

Mr. Prasanta Kumar Deka, Librarian, K. C. Das Commerce College, Guwahati.