

Issues and Challenges in Open Source Software Environment with Special Reference to India

Nabi Hasan

University Library

CSK H. P. Agricultural University, Palampur (HP), India

nabihasan@rediffmail.com

Abstract

The paper, while giving the introduction of the concept, describes the Open Source Software (OSS) and explains the meaning of the term OSS, giving some of the definitions, the typical terms used to explain it and also elaborates some of the important issues with reference to the explanation of the OSS. It discusses the reasons why the librarians and the libraries need the OSS model in current scenario, same. It further discusses in details, the important issues of OSS development and librarianship and Open-Source and Usability. The paper enumerates the important characteristics of the OSS and the criterion for selection of correct OSS as per the individual requirements. It describes important OSS, being presently used worldwide with special reference to the popular OSS used in Indian library environment, highlighting important characteristics/features, merits, etc. of the softwares. critical issues and challenges in OSS environment are also discussed. The paper further highlights the future of the OSS and conclude with an assessment of widely used OSS.

Keywords: Open Source Software, OSS, Free Software, Digital Library Softwares, Integrated Library Management System, Open Archive Initiative, General Public License.

Introduction

Open Source Software (OSS) came into existence with the development of ICTs. The term "open source" refers to software that includes the original source code, used to create it so that users can modify it to make it work according to their needs. It also includes the right of redistribution; therefore, there may be products that are based on other open source products. While the software may be free, a developer or distributor may charge for services including special programming, installation, training and technical support, etc. In general, the source code of OSS is widely accessible, freely available and reusable. The most popular source license, the General Public License (GPL), allows almost full use and re-use of source code.

Definition

OSS are computer programs in which the source code is made available to the general public for use and/or modification from its original design free of charge, i.e. open. Open source programs are typically created as a collaborative effort in which programmers offer the user a flexibility of use and share the changes within the community.

A certification standard is issued by the Open Source Initiative that indicates that the source code of a computer program is made available free of charge to the general public. The rationale for this movement is that a larger group of programmers not concerned with proprietary ownership or financial gain will produce a more useful and bug free product for everyone to use. The concept relies on peer review to find and eliminate bugs in the program code, a process which commercially developed and packaged programmes do not utilize. Programmers on the Internet read, redistribute and modify the source code, forcing an expedient evolution of the product. The process of eliminating bugs and improving the software happens at a much quicker rate as the information is shared through the open source community.

The term 'software' refers to two different but related things; *Source code*: a set of human readable and understandable instructions that comprise the 'recipe' from which an executable program can be made and *Object code*: the actual executable program which is compiled of machine readable source code. It is fed into a computer's microprocessor to perform various operations. The advocates of what we think of as the open source movements add further conditions before they regard software as open source.

Why Open Source Model for Librarians/ Libraries?

The basic idea behind open source is very simple; when programmers can read, redistribute, and modify

the source code for a piece of software, the software evolves. People improve it, people adapt it and people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing. OSS also helps in taking care of severe budget cuts, increased demand for services, lack of adequate staffing, etc.

The open source model offers librarians, the capability to create the software that we have always wanted - standards compliant, interoperable, extensible and scalable software that does what we want it to do: help customers find information quickly, conveniently, no matter where that information resides. To choose open source because it gives you the freedom to use, change or distribute the way you want. Remember, libraries are expected to stay much longer than the vendors. Vendors may not support a version which they sold you some time ago. Or they may go out of business. In that case, all your work and investment go waste. You may be forced to migrate to another version or software. With Open Source, you know what it all contains. You may tweak it yourself or hire people to do so, but then it is always with you. You can make it to evolve for your library's evolving needs.

OSS Development and Librarianship

Both OSS development and librarianship put a premium on open access. Both camps hope that the shared information will be used to improve our place in the world. Human interactions are a necessary part of the mix. Open source development requires an understanding of the problem, the computer application is trying to solve, and the maintainer must assimilate patches with the application. While databases and many "digital libraries" house information, these collections are really "data stores" until the data is given value and put to use whereby the stores become libraries. It has been stated that open source development will remove the necessity for programmers. Ironically, librarianship is flowering under new rubrics such as information architects and knowledge managers. Both institutions use peer-review, a process where "given enough eyeballs all bugs are shallow".

National Research Centre for Free/Open Source Software (*NRC-FOSS*) aims to contribute to the growth of FOSS in India through Research and Development, Human Resource Development, Networking and Entrepreneurship Development, as well as serve as the reference point for all FOSS related activities in the country including the creation and maintenance of this national FOSS Portal. Open Source Software Repository (*SourceForge*) – www.SourceForge.net is the world's largest Open Source software development web site. SourceForge.net provides free hosting to Open Source software development projects with a centralized resource for managing projects, issues, communications and code. UNESCO's Free and Open Source Software Portal - A gateway to resources related to Free Software and Open Source Technology movement is another milestone.

Open-Source and Usability

The problems if we observe, are typical of usability issues that frustrate novice users. Many of the identified issues are present in the software and documentation for some considerable time. The central mechanism for achieving software quality in open-source projects is extensive beta-testing. This 'bazaar-style' of development successfully encourages extensive functional testing of error-prone softwares to produce robust and reliable software such as the Apache web server. However, elements of usability may not be equally well-supported by open-source development - particularly when applied to software aimed at less technically-sophisticated users.

For a library professional like, open source software is a boon. It is difficult for library schools to purchase commercial software. Even if they are purchased, their maintenance and updating would be difficult as they involve financial commitment from the schools. Experience shows that commercial vendors were not very supportive for library schools in offering their software at nominal cost/free of cost. With the emergence of open source software, library schools now have options to include the software training in their curriculum. They can also now stay updated by using the latest version of the integrated library management system.

Characteristics

OSS has many characteristics. the important ones are:

- i) It is generally acquired freely
- ii) Manufacturer or developer has no right to claim royalties on the distribution or use
- iii) Source code is accessible to the user and distributed with the software
- iv) No denial to an individual or to a group to access source code of the software
- v) It has provision of modifications and derivations under the programme's original name
- vi) Rights of facilities attached to the programme must not depend on the programme's being part of a particular software distribution
- vii) Licensed software can not place restriction on other software that is distributed with it
- viii) Distribution of License should not be specific to a product and License should be technology-neutral, etc.

Criteria for Selection and Evaluation of OSS

Important points should be taken into consideration while choosing an OSS are:

- i) Reputation of the software
- ii) Monitor ongoing efforts and local usability

- iii) Support for Standards and Interoperability
- iv) User support
- v) Discussion Forums
- vi) Check versions, documentation available for the software
- vii) Skills of the workers
- viii) Availability and conditions of the license and the hidden cost involve
- ix) Commercial support for operability, etc.

Some of the popular OSS for Library and Information Management with special reference to India

Koha

It is the first Open Source Integrated Library Management System and is in use worldwide. Its development is steered by a growing community of libraries collaborating to achieve their library automation goals. Its important features meet the user needs including simple, clear interface for librarians and members (patrons), customizable search, circulation and borrower management, cataloguing module with integrated z39.50 client, Web 2.0, full acquisitions system including budgets and pricing information, ability to cope with any number of branches, patrons, patron categories, item categories, currencies and other data, serials system for magazines or newspapers, reading lists for members, etc. Its new features include a new user interface design, more advanced search functions, better multi-branch capabilities, user tagging and many generic enhancements.

NewGenLib

It is an Integrated Library Management System, is now freely available as an open source under the most widely used free software license, GNU General Public License (GNU GPL). Before becoming an open source product at the end of 2007, it was already in use in 122 libraries, mainly in India but also including installations in Syria, Sudan, Cambodia, etc. Its important features includes: completely web based, complies with international metadata and interoperability standards: MARC-21, MARC-XML, z39.50, SRU/W, OAI-PMH, uses chiefly open source components - scalable, manageable and efficient, OS independent - Windows and Linux flavours available, z39.50 Client for federated searching, Internationalized application (I18N)-Unicode 3.0 compliant, easily extensible to support other languages, data entry, storage, retrieval in any (Unicode 3.0) language, RFID integration, networking, automated email/instant messaging integrated into different functions of the software, form letters are configurable and use XML-based Open Office templates, extensive use of set up parameters enabling easy configuration of the software to suit specific needs.

Greenstone

It is a suite of software/tool for creating, building, managing and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet, means to easily create searchable and browsable interfaces to digital library collections via the Web, etc. It is open-source, multilingual software, issued under terms of the GNU General Public License. It 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 the software will create things like A-Z list of resources and field searchable interfaces. It builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use.

DSpace

DSpace is one of the first OSS platforms to store, manage and distribute the collections in digital format.

DSpace is the choicest software for academic, non-profit and commercial organizations, building open digital repositories. DSpace preserves and enables easy and open access to all types of digital contents including text, images, moving images, mpegs, data sets, etc. And with an ever-growing community of developers, committed to continuously expanding and improving the software, each new DSpace installation benefits from the previous. It is a groundbreaking digital repository system that captures, stores, indexes, preserves and distributes digital research material. As much of the world's contents are now being developed and disseminated in digital format, the software supports next generation digital archiving which is more permanent and shareable than current analog archives.

Fedora

The Fedora is based on the Flexible Extensible Digital Object and Repository Architecture (FEDORA). It can be used to develop institutional repositories and other interoperable web-based digital libraries and the system implements the Fedora architecture, adding utilities that facilitate repository management. It has been developed by the University of Virginia and Cornell University.

EPrints

The EPrints software has the largest and most broadly distributed installed base of any of the repository software systems described here. It is generic archiving software developed under University of Southampton. It is intended to create a highly configurable web-based archive and is a platform for building repositories of research literature, scientific data, student theses, project reports, multimedia artefacts, teaching materials, scholarly collections, digitized records, images, audio, exhibitions and performances, anything that can be stored digitally, etc. and offers integrating advance search, extended

metadata and other features such as Archive Documents, Multimedia and Data.

Ganesha

Ganesha Digital Library (GDL) enables institutions or personals to share their knowledge as well as simultaneously access and utilize knowledge. GDL is a tool for managing and distributing digital collection using web-based technology. GDL utilize Indonesia DLN Metadata Standard that is based on Dublin Core metadata standard. It opens possibilities of information exchange with other systems on the Internet that also utilize Dublin Core. Data transaction between client and server within GDL-Network using XML format is allowed. It makes possible for further development of GDL to become more extensive web-based networking application in the future.

CERN

CERN Document Server Software (CDSWare) is an integrated digital library management system which provides the framework and tools for building and managing an autonomous digital library server. It covers all aspects of digital library management and complies with the Open Initiative Metadata Harvesting Protocol (OAI-PMH) and uses MARC21. Its flexibility and performance make it a comprehensive solution for the management of document repositories of moderate to large size libraries.

CDS/ISIS for Windows

This is the most widely used software for Indian libraries, available freely from the UNESCO website. It has all the features of a complete Integrated Library Management Software and the support for this software is available locally as well as a number of professionals are trained in its use. It also has a very strong base for international discussion forum. Truly speaking, most of the libraries in India started computerizing their cataloguing activities with this software only and later on shifted to other commercially available softwares and OSS.

Critical Issues and Challenges

The library database should be published on Internet, i.e. all modules should be web based and network centric. Members can see the things from home or remote location. Library staff can make data entry from remote and different locations in India. If libraries are maintaining their database and dynamic website on LAN or single machine (unless mandatory), they are lagging behind others. In case of open source, the library will have to work hard to get it published through its ISP. Simply publishing static web site and hosting is different from hosting a dynamic website on some ISP. Reason is, understanding others' code is not so easy. Once one publishes one's data on Internet, authentication and authorization should be security audited. In case of open source softwares, extra precautions should be taken. A majority of the libraries in India is maintaining their database on LAN or local single PC. It is not the proper automated

system; you should switch over to latest technology. Here also, it is clarified that all the popular softwares in India are having web based architecture, and not only Open Source.

First of all going for any software, we should see our organization's computing infrastructure along with the ISP for publishing on Internet, accordingly one should select the software either Open source or commercial. Platform may be Linux or Windows. LMS may be anything; open source or proprietary. But certainly, one may not suggest using the OSS or any software in isolation like LAN or within four walls of library. Any solution (Open, Free or Closed) will require Servers, Network Infrastructure, Manpower to handle installation, alternation and alignment of System Processes (Library Process here) and training of the staff and users of the system. Now since, libraries are meant to survive over generations; the ideal software for libraries would be those, whose developer survives over generations. Not just survive but keep on developing the software with new requirements and in accordance to best technologies available in time. Total cost of ownership (TCO) is high in case of open source solution. If one considers the software installation, server setup, training, AMC, hosting, security, follow-up, customization the cost will be high. In relation to the total implementation, cost of software is negligible. One should not worry about the open or closed softwares; the motive should be Proper Information Systems Solution.

Open Source is the Difference

Open source has been a buzzword in the library community for several years now. We have heard the hype: open-source software is free, more reliable, more secure, boasts faster development cycles, and is just plain cooler than proprietary software. Here are just a few of the reasons why open source is an especially attractive solution for libraries. Open-source software is free; the library will pay only for the product support and training (if any) that it may need.

Software functions are paid for only once making open-source software extremely cost-efficient. Libraries using open-source software benefit from many advanced technology solutions that they otherwise could not afford to develop themselves yet they still have the option to steer development if they so desire. Open source empowers libraries to be innovate and collaborate. Not only can they download and use open-source software for free, they are free to alter it in any way they deem fit, provided the result are redistributed for free.

Choosing Support

In a proprietary software development model, one pays high license fees to use the software. If the vendor is not providing with adequate support or is not allowing client the freedom to customize and improve the software to meet their needs, switching vendors means

switching software. And then there is the matter of shifting the data from one vendor to the next: with open-source software, since all one is paying for is support, switching to another service provider or migrating to an in-house solution is simple. Further, an open-source software development model means unique identity and property of data.

The Future is 'open'

Open source in libraries has its challenges as well. Till now, library software vendors have built their businesses around a proprietary software development model, and, as a result, libraries have been slow to adopt open source. Many libraries simply do not have the in-house expertise to support open-source software development, and also don't have the ability to train their staff on the use of new technologies. Open source is here. It's growing. Anyone can be a part of it.

Conclusion

The Open Archives Initiative (OAI) has gained momentum since eprints.org was released in 2000. OSS incorporates an interface that makes it easy for people to create their own library collections. Collections may be built and served locally from the user's own web server, or remotely on a shared digital library host. End users can easily build new collections styled after existing ones from material on the web or from their local files (or both), and collections can be updated and new ones brought on-line at any time. OSS has much potential for libraries and information centres, and there are numerous projects, including Koha, NewGenLib, Greenstone, DSpace, Ganesha, etc. that demonstrate its viability in this context. It gives library staff an option to be actively involved in development projects, and this involvement can take many forms, such as reporting bugs, suggesting enhancements and testing new versions. Currently available OSS projects cover application areas ranging from the traditional library management systems to innovations like Greenstone and DSpace, which complement traditional systems. DSpace, EPrints, Greenstone and KOHA are among the top most OSS which are widely used in India and world over because of flexible searching, browsing, phrases and almost zero maintenance features. These concepts and their benefits and importance to libraries should be examined and explored for the wider audience and prospects for long-term preservation of scholarly works.

The management and parent organization never stop if a good proposal with time bound implementation is put for financial approval. Government of India has allocated Rs. 23000 crore for e-Governance during 11th Plan. Many of the organizations are not able to spend their ICT budgets or sometime manage to spend somehow. Community backed softwares are the answer for long term continuity and continuous updates. These are nothing else but Open Source

Softwares which are developed and maintained by 'community spirit'. Since communities last longer than individuals and businesses, the open source softwares backed by strong communities last much longer. The community spirit does wonders and cannot be explained in simple business models – but it works! Well, who owned Internet anyway? How Wikipedia came to a level that it could be compared with Encyclopaedia Britannica. How Linux keeps itself going? And what is this Web 2.0? What the hell makes sites like Facebook / Orkut (and many others) so popular. Well, there is this 'community spirit' in every such wonder. If youngsters are following this; follow the advice – go for Open Source Softwares.

Bibliography

1. Caplan, Priscilla (2007). The Florida digital archive and Daitss: A working preservation repository based on format migration, *International Journal on Digital Libraries*, 6(4), 305-311.
2. Chaoji, Vineet, et al. (2008). An integrated, generic approach to pattern mining: Data mining template library, *Data Mining and Knowledge Discovery*, 17(3), 457-495.
3. Deb, Subrata (2006). *TERI integrated digital library initiative*, *Electronic Library*, 24(3), 366-379.
4. Edwards, Kasper (2005). *An economic perspective on software licenses: Open source, maintainers and user-developers*, *Telematics and Informatics*, 22(1-2), 111-133.
5. Hansen, Marit et al. (2002). The open source approach: Opportunities and limitations with respect to security and privacy, *Computers and Security*, 21(5), 461-471.
6. Jesus M, et al. (2009). *Macro-level software Evolution: A case study of a large software compilation*, *Empirical Software Engineering*, 14(3), 262-285.
7. Kai Pan et al. (2009). *Toward an understanding of bug fix patterns*, *Empirical Software Engineering*, 14(3), 286-315.
8. Krogh, Georg von & Spaeth, Sebastian (2007). The open source software phenomenon: Characteristics that promote research, *Journal of Strategic Information Systems*, 16, 236-253.
9. McDonald, Clemant J. et al (2003). Open Source Software in medicinal informatics: Why, how and what, *International Journal of Medical Informatics*, 69(2-3), 175-184.
10. Reilly, Colleen A & Williams, Joseph John (2006). *The price of free software: Labor, ethics and context in distance education*, *Computers and Composition*, 23(1), 68-90.
11. <http://blog.sukhdevsingh.com/2008/02/why-open-source-model-for-librarians.html>
12. <http://cdsware.cern.ch>
13. <http://digilib.itb.ac.id>
14. [http://dspace.inflibnet.ac.in/bitstream/1944/529/1/4\(cal+06\).pdf](http://dspace.inflibnet.ac.in/bitstream/1944/529/1/4(cal+06).pdf)
15. <http://en.wikipedia.org>

16. <http://eprints.rclis.org/archive/00012791>
17. <http://gdl.itb.ac.id>
18. <http://www.drupal.org>
19. <http://www.dspace.org>
20. <http://www.gnu.org>
21. <http://www.greenstone.org>
22. <http://www.indonesiadln.org>
23. <http://www.infomotions.com/musings/ossnlibraries-workshop>
24. <http://www.joomla.org>
25. <http://www.koha.org>
26. <http://www.liblime.com/open-source/why-your-library-needs-open-source>
27. <http://www.nrcfoss.org.in>
28. <http://www.oss4lib.org>
29. <http://www.sourceforge.net>
30. <http://www.unesco.org/isis>
31. www.diglib.org/architectures/ossrep.htm