
DIGITAL LIBRARY OPEN SOURCE SOFTWARE : A COMPARATIVE STUDY

Umesha Naik D. Shivalingaiah

Abstract

The Open Source Software (OSS) model makes source code available to users, who can change the software to tailor it more closely to their own requirements. With many OSS applications now available for library and information management, Organisations have a new option for acquiring and implementing systems, plus new opportunities for participating in OSS projects. Examples of such systems include Greenstone DSpace, and Ganesha. OSS is popular with technically sophisticated users, who are often also the software developers. OSS is becoming an increasingly popular software development method. This paper highlights the comparison, features, function and usability of OSS like Greenstone Digital Library, DSpace and Ganesha Digital Library.

Keywords: Digital Library, Open Source Software, Greenstone, DSpace, Ganesha

1. Introduction

OSS has grown tremendously in scope and popularity over the last several years, and is now in widespread use. The growth of OSS has gained the attention of research librarians and created new opportunities for libraries. OSS can benefit libraries by lowering initial and ongoing costs, eliminating vendor lock-in, and allowing for greater flexibility.

In the phrase *open source*, *source* refers to source code, the human-readable computer code which is the origin, or source, of the computer application. *Open* refers to the terms of access to that computer source code. So OSS is software for which the source code is freely available. The most important aspect of the open source movement is the participation of users. When a user or users what a feature or bug fix for a program, they have traditionally been at the mercy of the software vendor. However, with open source they can modify the program to their own needs or fix what is broken. Many users will help develop the program for free, simply to improve the product and benefit the community.

OSS is any software whose code is available for users to look at and modify freely. All Open Source projects have an owner, all Open Source projects are governed by some type of license agreement: GPL, BSD, etc¹

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
- **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. Some essentials are:

- The source is publicly available
- The software can be distributed freely
- You have the right to derive new works through modification

2. OSS and Rights

These rights fall into three broad categories²:

1. *Rights to use without discrimination.* Unlike commercial software, OSS may be used for any purpose, by anyone, at any time. For example, the same OSS used to run an academic website can also be used to run an e-commerce business. There are no annual license fees, restrictions on the numbers of users or systems, restrictions for noncommercial use, restriction to a particular country, expiration-dates, or other artificial limits on use.
2. *Full rights to create derived works.* OSS not only permits one use the software, but permits one to create new software from it.
 - a. *Source Code Availability.* The source code for the software is made on the same terms as the binaries used to run it.
 - b. *Free modification and redistribution.* Consumers not only have the right to examine the source, but to freely modify and redistribute modified (or unmodified) copies.
 - c. *Integrity of authorship.* OSS may require that previous authors be acknowledged, and that modifications be clearly labeled, and separately packaged and named from the original software when redistributed. This maintains integrity of software.
3. *No traps.* Modified copies of OSS must be redistributable under the same license as the original. The license cannot be restricted to a single product, and it must not restrict the distribution of other independently licensed software.

3. OSS development and librarianship

1. Both OSS development and librarianship put a premium on open access. Both camps hope the shared information will be used to improve our place in the world.
2. 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. Librarians understand that information seeking behavior is a human process. 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.
3. Third, 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.
4. Both institutions use peer-review, a process where “given enough eyeballs all bugs are shallow”.

-
5. DSpace - DSpace is a tool designed to allow institutions, such as libraries to collect, archive, index, and disseminate the scholarly and intellectual efforts of a community. Written with a combination of technologies by MIT, it is primarily used to capture bibliographic information describing articles, papers, theses, and dissertations. Once entered into the system, DSpace indexes the content and provides a way to link to the originals. DSpace plays well with open standards such as XML and OAI-PMH. If a large number of institutions of higher education were 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.
 6. Greenstone - Greenstone is a tool for creating and managing digital library collections. Running on Windows as well as various flavors of UNIX, 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. Thus the digital library collections can be distributed to people with poor or no 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 language³.

4. Open-Source and Usability

The problems we observed were typical of usability issues that frustrate novice users. Many of the identified issues had been present in the Greenstone software and documentation for some considerable time. They had been effectively invisible to the developers on the project and had not been reported by the users. Although Greenstone is not a completely typical open-source project we wondered whether there are structural reasons inside open-source projects that could cause these types of usability issues to persist.

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 software 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 users⁴.

5. Greenstone

The Greenstone digital library software is an open-source system for the construction and presentation of information collections. It builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt entirely automatically. The system is extensible: software "plugins" accommodate different document and metadata types⁵.

The aim of the Greenstone software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. The latest version of the software is 2.62 released on October 2005. Interfaces available for the Greenstone digital library software: the four "core" languages English, French, Spanish, and Russian. The Greenstone "Collector" is an interactive subsystem for managing and accessing collections.

5.1 The Collector can be used to:

- create a new collection with the same structure as an existing one;
- create a new collection with a different structure;
- add new material to an existing collection;
- modify the structure of an existing collection;
- delete a collection;
- write an existing collection to a self-contained, self-installing Windows CD-ROM.

5.2 Greenstone is

- **Widely accessible** : Collections are accessed through a standard web browser.
- **Multi-platform**: Collections can be served on Windows and UNIX, with an external Web server or (for Windows) a built-in one.
- **Metadata-driven**: Browsing (and, if desired, searching) indexes are built from metadata. Metadata may be associated with each document or with individual sections within documents. It must be provided explicitly (often in an accompanying XML or spreadsheet file) or derivable automatically from the source documents.
- **Extensible**: Plugins can be written to accommodate new document types. Classifiers can be written to create new kinds of browsing indexes based on metadata.
- **Multi-language**: Unicode is used throughout and is converted on-the-fly to an encoding supported by the user's Web browser. Separate indexes can be built for different languages: a plug-in allows automatic language identification for multilingual collections.
- **International**: The interface is available in multiple languages: new ones are easy to add.
- **Large-scale**: Collections containing millions of documents, and up to several gigabytes, have been built. Full-text searching is fast. Compression is used to reduce the size of the indexes and text
- **Z39.50 compatible**: The Z39.50 protocol is supported for accessing external servers and (under development) for presenting Greenstone collections to external clients.

5.4 Greenstone provides

- **Flexible searching**: Users can search the documents' full text, choosing between indexes built from different parts. Queries can be ranked or Boolean; terms can be stemmed or unstemmed, case-folded or not.
- **Flexible browsing**: Users can browse lists of authors, lists of titles, lists of dates, hierarchical classification structures, and so on. Different collections offer different browsing facilities, determined at build time.

-
- **Zero maintenance:** All structures are built directly from the documents themselves. New documents in the same format can be merged into the collection automatically. No links need be inserted by hand, but existing hypertext links in the original documents, leading both within and outside the collection, are preserved.
 - **Phrases and key phrases:** Standard classifiers create phrase and key phrase indexes of text — or indeed any metadata.
 - **Sustained operation:** New collections can be installed without bringing the system down. Even active users rarely notice when a collection is updated.

5.5 Greenstone enables:

- **Multimedia :** Collections can contain pictures, music, audio and video clips. Currently, non-textual material is either linked in to documents or accompanied by written descriptions to allow access.
- **CD-ROM option:** Collections can be published on a self-installing CD-ROM. A multi-disk solution has been implemented for larger collections.
- **Distributed collections :** Collections served by different computers can be presented to users as though they were part of the same library, through a flexible process structure.
- **Easy modify:** it is easily modified.

6. DSpace: Open Source Digital Library (DL) System

DSpace is a groundbreaking digital institutional repository that captures, stores, indexes, preserves, and redistributes the intellectual output of a university's research faculty in digital formats. It manages and distributes digital items, made up of digital files (or bit streams) and allows for the creation, indexing, and searching of associated metadata to locate and retrieve the items. DSpace design and developed by Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP) DSpace was designed as an open source application that institutions and organizations could run with relatively few resources. It is to support the long-term preservation of the digital material stored in the repository. It is also designed to make submission easy. DSpace supports submission, management, and access of digital content⁶.

6.1 DSpace support the types of content

DSpace accepts all manner of digital formats. Some examples of items that DSpace can accommodate;

- Documents, such as articles, preprints, working papers, technical reports, conference papers
- Books
- Theses
- Data sets
- Computer programs
- Visualizations, simulations, and other models

- Multimedia publications
- Administrative records
- Published books
- Overlay journals
- Bibliographic datasets
- Images
- Audio files
- Video files
- Reformatted digital library collections
- Learning objects
- Web pages

DSpace support the types of file formats

- Text
- Images
- Audio
- Video

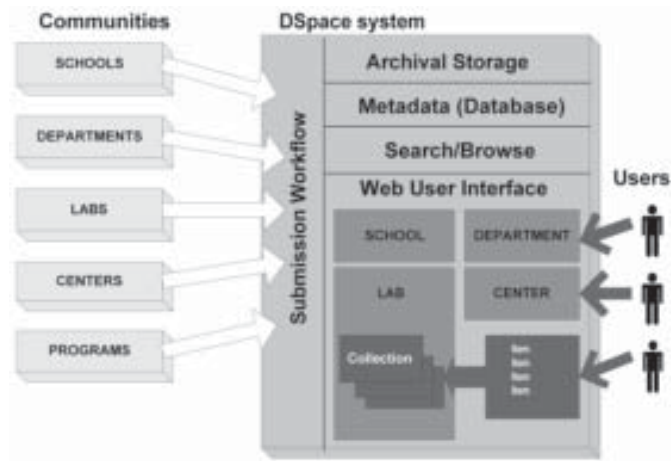
6.2 Institutional Repository

DSpace is a digital library system to capture, store, index, preserve, and redistribute the intellectual output of a university's research faculty in digital formats.

- DSpace is organized to accommodate the multidisciplinary and organizational needs of a large institution.
- DSpace provides access to the digital work of the whole institution through one interface.
- DSpace is organized into *Communities* and *Collections*, each of which retains its identity within the repository.

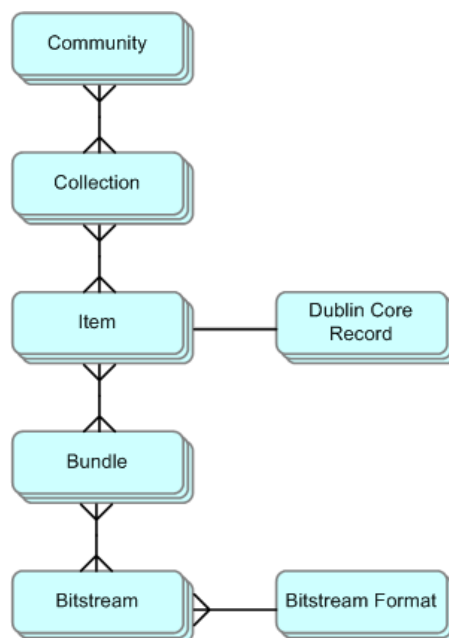
Customization for DSpace communities and collections allows for flexibility in determining policies and workflow.

6.3 DSpace information model



6.4 Data Model

Content in DSpace is at the highest level organized into *communities*.



6.5 DSpace Digital Preservation

Digital Preservation DSpace identifies two levels of digital preservation first one is Bit preservation ensures that a file remains exactly the same over time - not a single bit is changed - while the physical media evolve around it. - Functional preservation goes further: the file does change over time so that the material continues to be immediately usable in the same way it was originally while the digital formats (and physical media) evolve over time. One of the primary goals of DSpace is to preserve digital information.

- DSpace provides long-term physical storage and management of digital items in a secure, professionally managed repository including standard operating procedures such as backup, mirroring, refreshing media, and disaster recovery.
- DSpace assigns a persistent identifier to each contributed item to ensure its irretrievability far into the future.
- DSpace provides a mechanism for advising content contributors of the preservation support levels they can expect for the files they submit.

For all three levels, DSpace does bit-level preservation so that “digital archaeologists” of the future will have the raw material to work with if the material proves to be worth that effort.

- **Access Control:** DSpace allows contributors to limit access to items in DSpace, at both the collection and the individual item level.
- **Versioning:** New versions of previously submitted DSpace items can be added and linked to each other, with or without withdrawal of the older item. Multiple formats of the same content item can be submitted to DSpace, for example, a TIFF file and a GIF file of the same image.
- **Search and Retrieval:** The DSpace submission process allows for the description of each item using a qualified version of the Dublin Core metadata schema.

7. Ganesha Digital Library

Ganesha Digital Library enables institutions or personals to share their knowledge as well as simultaneously access and utilize knowledge. Ganesha Digital Library or GDL is a tool for managing and distributing digital collection using web-based technology. GDL enables institutions or personals to share their knowledge as well as simultaneously access and utilize knowledge in Indonesian “giant memory” in the form of network of Indonesia DLN digital libraries. The latest version of the software is 4.0 released on June 2004⁷.

7.1 Features of Ganesha digital library software

- **Distributed Knowledge Management:** Knowledge management done distributed, by partner in each digital library server.
- **Centralized Knowledge Distribution:** To make information closer to user, GDL Partner Server can benefit GDL Hub Server (Central Server) in order to disseminate metadata to all Digital Library Partner Server within IndonesiaDLN.

-
- **Online Member Registration:** User registration can be done online on the web. Validation number sent by email to make sure to make any contact to user in the future.
 - **Roaming Membership:** Once user is registered in any GDL server, he/she can use his/her account in every online GDL server.
 - **Searching:** GDL 3.1 supports fast information searching and detail to all managed metadata.
 - **Category-Based Organization:** Organizing collection done with creating category and sub-category. This is make browsing easier.
 - **Upload Metadata and Files:** Every member can publish his/her digital collections by submitting metadata form and upload the file easily.
 - **Personal Directory:** Every member automatically posses personal directory in which he/she can freely manage.
 - **Review Forum:** Every uploaded article can be set whether ask for review by visitor or not. Visitor can post and read review. Contributor will receive email notification if there is any review posted.
 - **Access Restriction:** Uploaded articles can be arranged whether can only accessed by Intranet (any particular group) or open to the Internet. It help for work that not yet been finished or to be reviewed by the group.
 - **Image Thumbnail:** Image file (jpg and PNG) can be appeared in smaller size (thumbnail) at abstract page.
 - **Knowledge Organization:** Member, editor, and knowledge officer can organize where to put uploaded articles to appropriate categories in regard to their privileges.
 - **News:** Editor and knowledge officer can upload fresh news to be appeared in GDL News GDL easily.
 - **Synchronization:** GDL Partner Server can upload and download file and metadata to/from GDL Central Server through Synchronization facilities. Membership and publisher information can also be synchronized.
 - **Member and Group Administration:** Administrator can manage member data, create group, and regulate editor access right.
 - **Statistics:** Administrator can view statistics of knowledgebase content and its contributor.
 - **Advertisement:** Administrator can show advertisement banner that completed with keyword and subject matching facilities.




- **Dublin Core / IndonesiaDLN Metadata:** GDL utilize IndonesiaDLN Metadata Standard that based on Dublin Core metadata standard. It open possibilities of information exchange with other system on the Internet that also utilize Dublin Core.
- **XML Based Transaction:** Data transaction between client and server within GDL-Network using XML format. It is make possible for further development of GDL to become more extensive web-based networking application in the future.
- **CD-ROM Enabled:** GDL uses Apache, MySQL, and PHP free-software that can be run directly from CD-ROM to make easy information dissemination.
- **CD-ROM Enabled**

7.2 Ganesha Purposes:

- Managing scholar resources: theses, dissertations, research reports, journal, publication, etc.
- Promoting the SME's products: E-Mall (currently the e-transaction is not supported).
- Managing the art work and heritage resources: pictures, songs, videos, etc.
- Managing the expertise directory of people and organizations.
- Extend the metadata schema for other purposes easily.
- And the most important, develop distributed knowledge repository network.

8. Comparison of three popular digital library OSS

| Subject | Greenstone | DSpace | Ganesha |
|-------------------------|--|--|--|
| About OSS | It is a suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. | It is a groundbreaking digital library system that captures, stores, indexes, preserves and redistributes the intellectual output of a university's research faculty in digital formats. | Ganesha Digital Library (GDL) enables institutions or personals to share their knowledge as well as simultaneously access and utilize knowledge. |
| URL Developed by | http://www.greenstone.org/ New Zealand Digital Library Project at the University of Waikato, and developed & distributed in cooperation with UNESCO and the Human Info NGO | http://www.dspace.org/ Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP) | http://gdl.itb.ac.id/ Indonesian Digital Library Network, (IndonesiaDLN) |
| License | GNU General Public License http://www.gnu.org/copyleft/gpl.html | The BSD License http://www.opensource.org/licenses/bsd-license.php | GNU General Public License http://www.gnu.org/ |

| | | | |
|----------------------------|---|---|---|
| Platform | Windows 3.1/ 3.11/ 95/ 98/ Me/ NT/ 2000, GNU/Linux, Darwin (Mac OS X), Solaris, and FreeBSD | UNIX or LINUX | Windows 98, NT, 2000, XP, Linux/Unix. |
| Latest Version | 2.62 October 2005 | 1.3.2 October, 2005 | 4.0 January 2004 |
| Languages | English, French, Spanish, Russian and Kazakhs | English | English |
| Software Size | Windows = 37 MB Unix = 44 MB Mac OS X = 35 MB | 8.94 MB | 13.90 MB |
| Associated software | <ul style="list-style-type: none"> • Apache Web server • PERL • GNU C++ Compiler • GNU Database Manger • Java runtime • Java compiler | <ul style="list-style-type: none"> • Java 1.4 or later (standard SDK is fine, you don't need J2EE) • Apache Ant 1.5 or later (Java make-like tool) • PostgreSQL 7.3 or later, an open source relational database, or Oracle 9 or higher. | <ul style="list-style-type: none"> • Apache • PHP • MySQL |
| Emblem |  |  |  |
| Download from | http://prdownloads.sourceforge.net/greenstone/greenstone-users@list.scms.waikato.ac.nz | http://sourceforge.net/projects/dspace/ | http://gdlib.itb.ac.id/download/ |
| Contact Information | greenstone-users@list.scms.waikato.ac.nz | sales@dspace.com.au | mrg@krmrg.itb.ac.id |
| Examples | New Zealand Digital Lib. Project at the University of Waikato. http://www.sadl.uleth.ca/nz/cgi-bin/library The Human Info NGO, based in Antwerp, Belgium http://humaninfo.org/home_flash.html | DSpace at MIT http://dspace.mit.edu/ ETD of Indian Institute of Science, Bangalore, India http://etd.ncsi.iisc.ernet.in/ | Institut Teknologi Bandung (ITB) http://digilib.itb.ac.id/ The Indonesia Digital Library Network http://www.indonesiadln.org/ |
| FAQ | http://www.greenstone.org/cgi-bin/library?e=p-en-home-utfZz-8&a=p&p=faq | http://wiki.dspace.org/TechnicalFaq | http://digilib.art.itb.ac.id/faq.php |
| Written in | Written in C++, Perl, & Java | Written in Java | |

9. Conclusion

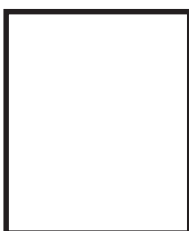
The Open Archives Initiative (OAI) has gained momentum since eprints.org was released in 2000. OSS incorporates an interface that makes it easy for people to create their own library collections. Collections may be built and served locally from the user's own web server, or remotely on a shared digital 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 a number of projects, including Greenstone, DSpace, and Ganesha 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. These concepts and their benefits and importance to libraries are examined. Benefits include lower costs, greater accessibility, and better prospects for long-term preservation of scholarly works

10. Reference

1. ACM Digital Library. Retrieved October 01 2005 from <http://www.acm.org/dl/>
2. David M. Nichols, Kirsten Thomson and Stuart A. Yeates . "Usability and open-source software development, New Zealand, University of Waikato, Retrieved September 20 2005 from <http://www.comp.lancs.ac.uk/computing/users/dmn/docs/oss.html>
3. Digital library standards and practices. Retrieved September 25 2005 from <http://www.diglib.org/standards.htm>
4. DSpace: open source Digital Library (DL) system. Retrieved October 20 2005 from <http://www.dspace.org>
Free Software / Open Source: Information Society Opportunities for Europe. Retrieved September 25 2005 from <http://eu.conecta.it/paper/Contents.html>
5. Ganesha: the first web-based digital library software in Indonesia. Retrieved October 20 2005 from <http://gdl.itb.ac.id>
6. Greenstone. Retrieved October 20 2005 from <http://www.greenstone.org/english/home.html>
7. IEEE Computer Society Digital Library. Retrieved September 30 2005 from <http://www.computer.org/publications/dlib/>
8. Open Source Initiative. Retrieved September 20 2005 from <http://www.opensource.org/docs/definition.php>
9. Open Source Systems for Libraries Retrieved September 25 2005 from <http://www.oss4lib.org>
10. Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers. Retrieved September 20 2005 from http://www.dwheeler.com/oss_fs_why.html
11. What is Open-Source Software? Retrieved November 10, 2005, from <http://www.darwinmag.com/learn/curve/column.html?ArticleID=108>

About Authors

Mr. Umesh Naik is a Lecturer in the Department Library and Information Science, Mangalore University, Mangalore. Prior to this he has worked 8 years at INFLIBNET Centre. He obtained his BLISC. degree from Mangalore University and MLIS from IGNOU. His areas of interest are Networking, Internet, Web Design, Digital and Electronic Libraries. He published eight articles in journals and seminar/conferences.
Email : umeshai@yahoo.com



Prof. D. Shivalingaiah is a Professor in Library and Information Science, Mangalore University, Mangalore. He holds M.A. in Rural Development and M.L.I.Sc. from Bangalore University and Ph.D. from Mangalore University. He has successfully guided two candidates for Ph.D. programme. Presently five candidates are working under him for Ph.D. programme. He has publications in Journals and Conference Proceedings and edited books. He is presently working as Special Officer (Administration) on deputation.
Email : d_shivaling@yahoo.com