Open Source Software for Creation of Digital Library:
A Comparative Study of Greenstone Digital Library Software & DSpace

Sukhwinder Randhawa

Abstract:

Softwares now-a-days have become the life line of modern day organizations. Organizations cannot think of doing their tasks effectively and efficiently without softwares. The extremely competitive environment, zero deficiency and enhanced productivity has made it mandatory for the organizations to carefully choose the appropriate software after comprehensive needs assessment. Softwares simply their tasks and saves a lot of precious time which can be utilized in managing other important issues. Libraries also need softwares if they want to create a parallel digital library with features which we may not find in a traditional library. There are several open source softwares available to create a digital library. For this, firstly the library professionals should be aware of the advantages of open source software and should involve in their development. They should have basic knowledge about the selection, installation and maintenance. Open source software requires a greater degree of computing responsibility than commercial software. Digitization involves huge money to create and maintain and the OSS appears to be a means to reduce it. Among these, DSpace and Greenstone are becoming more popular in India and abroad. This paper deals with the comparison of these two popular OSS from various points of view. The comparative table may help the professionals who are planning to create a digital library.

Keywords: Open Source Software; Digital Library; Greenstone Digital Library (GSDL); DSpace.

Introduction

Libraries do not have huge amounts of money to experiment, and they don't usually purchase additional resources. The need for software, its installation, training and the lack of money available to spend on it have forced many libraries to stand for themselves when it comes to staying up to date with the latest technology. Unless, of course, they adopt the open source movement and use a few of them available to overcome these problems. Most software that we all use everyday is known as "proprietary", which means that it costs money and that the actual code of the software is restricted, in that the code of the software

cannot be modified, copied, or changed from its original construction. The code is "unreadable" and pretty much is what it is. Open source software (OSS), on the other hand, is quite the opposite. The open source mentality revolves around sharing and collaboration, and these two important elements describe open source software perfectly. First and foremost, open source software is free for anyone to have; more importantly, not only is the software free, but it is also free for anyone to copy, hack, modify, etc. This increases the possibilities of a software program's potential because of this free-thinking model. There are many different kinds of open source software solutions out there today that could be embraced by the library. There's basic operating system, document processing programs, Library Management Software (LMS) and Digital Library software.

Open source software

Open source software is computer software whose source code is available under a license

Author's Affiliation: *Assistant Librarian, Center for Research in Rural and Industrial Development (CRRID), Sector 19-A, Madhya Marg, Chandigarh-160019

Reprint's request: Tridib Tripathi, Sukhwinder Randhawa, Assistant Librarian, Center for Research in Rural and Industrial Development (CRRID), Sector 19-A, Madhya Marg, Chandigarh-160019. Email: sukh020@gmail.com .

that permits users to study, change, and improve the software, and to redistribute it in modified or unmodified firm. It is often developed in a public, collaborative manner. It is the most prominent example of open source development and often compared to user generated content.

Advantages of Open Source Software

- The availability of the source code and the right to modify it is very important as enables in the improvement and extend the lifetime of a software product.

- Source code availability also makes it much easier to identify errors, and to fix them.

- The right to redistribute modifications and improvements to the code, and to reuse other open source code, permits all the advantages due to the modifiability of the software to be shared by large communities.

- Continuous improvement does not require users to pay for it. There is no single entity on which the future of the software depends. This is a very common concern with proprietary software.

- There are fewer conflicting priorities due to marketing pressures. Usually open source software is delivered "when it is ready", and when the development team feels that its quality is good enough. This means that software usually does not need as many "service packs", updates and such, reducing the maintenance cost.

- It provides a new forum for democratic action, collaboration, mutual benefit without geographical or any other barrier/ bias.

- It forces commercial software vendors to keep their product price at a reasonable level.

Digital Library Softwares

Digital libraries are defined in many ways, encompassing both analog material made available digitally and newly created digital content. It is an organized searchable collection in digital format. There are quite a few digital library software packages available in the market. One has to select the right kind of package depending on their specific need. There are a few library automation software packages, which have a separate 'digital library' module. Open source digital library software derives its strength from several enabling technology and metadata based inter operability protocols, which have become available recently. Examples of some of these are as follows:

Greenstone Digital Library

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Its developers received the 2004 IFIP Namur award for "contributions to the awareness of social implications of information technology, and the need for an holistic approach in the use of information technology that takes account of social implications."

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. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. This software is developed and distributed as an international cooperative effort established in August 2000 among three parties:

Sukhwinder Randhawa / Open Source Software for Creation of Digital Library: A Comparative Study of Greenstone Digital Library Software & DSpace

309

- New Zealand Digital Library Project At The University Of Waikato

- United Nations Educational, Scientific And Cultural Organization (Unesco)

The Human Info NGO, Based In Antwerp, Belgium

DSpace

DSpace is open source software for building and managing Digital repositories. Developed jointly by MIT Libraries and Hewlett-Packard (HP), is freely available to research institutions as an open source system that can be customized and extended. DSpace is a digital institutional repository that captures, stores, indexes, preserves, and redistributes content in digital formats. Institutional Repository is a set of services that a research institution/ organization/ university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members Typically, DSpace has been deployed for Institutional Repositories of publications, thesis and dissertations. There are several groups working on extending its capabilities such implementation of ontologies in search interface and for submission module, customization for management of electronic theses and dissertations and for localization and international of the package for the world languages. DSpace is designed for ease-of-use, with a web-based user interface that can be customized. The DSpace system provides a way to manage research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time.

Comparative Analysis

Both Greenstone and DSpace are considered effective software for building digital libraries. Both have a good number of successful installations. But there are some basic differences also that has been highlighted in the table given below. The objective of the present work, includes a comparative study of Greenstone and DSpace in order to understand their feasibility and effectiveness as tool for building sustainable digital libraries. Comparison between Greenstone and DSpace has been carried out on points like the availability, version, developer prestige, operating system, system requirement, license, language, technical and training support, examples, security, browse and search and other points which are necessary for comparative study. Greenstone emanates from the Department of Computer Science in the University of Waikato, New Zealand, and is developed and distributed in cooperation with UNESCO as part of its Information for All program. DSpace was conceived by Hewlett-Packard Labs as a noncommercial product, and has been developed in conjunction with MIT Libraries. Both systems have active open source developer communities.

Developers

Greenstone

Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. 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. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries.

DSpace

The MIT Libraries and Hewlett-Packard (HP) jointly developed DSpace. The system is now freely available to research institutions world-wide as an open source system that can be customized and extended.

Feature	Greenstone	DSpcae
Developer	New Zealand Digital Library Project at the University of Waikato, UNESCO and Human Info NGO	MIT Libraries and Hewlett-Packard Company
Platform	Windows 95/98/Me/NT/2000/XP Unix/Linux, and MAC OS-X	Windows (NT/2000/XP) and All POSIX (Linux/BSD/UNIX-like OSes), OS X
Open Source and Free	Yes	Yes
License	GNU General Public License	BSD License
Language	English, French, Spanish, Russian and Kazakhs. The reader's interface is available in many other including Indian languages Bengali, etc	English
Associated Software	Apache Web server, Java 1.4.0 or above, Image Magick Software and Web Browser	Java JDK 5 or later Apache Ant 1.6.2 or later, Apache Maven 2.0.8 or later Java 1.4 or later, PostgreSQL 7.3 or later, Apache Tomcat 4.x/5.x and Web Browser
URL	http://www.greenstone.org/	http://www.DSpace.org/
Download	https://sourceforge.net/projects /greenstone/files/Binary Installer-Windows/ Greenstone-2.83-windows.exe/ download	http://sourceforge.net/projects/ds pace/files/
Latest version	Greenstone 2.83	Space 1.6. 2
Metadata	Dublin Core (qualified and unqualified) RFC 1807 NZGLS (New Zealand Govt. Locator Service) AGLS (Australian Government Locator Service)	Dublin Core (qualified)
Protocols	OAI-PMH	OAI-PMH
Support: Documentation	Yes	Yes
WIKI	YES	YES
FAQ	YES	YES
MAILING LIST	YES	YES
Commercial Support	YES	YES
TRAINING	YES	YES
Supported File formats	MS-Word, PDF, HTML, PostScript, JPEG, GIF	MS-Word, PDF, PPTs, JPEG, GIF
Written in	Written in C++, Perl, & Java	Written in Java

Platform

Greenstone and DSpace both are available for Windows and Unix/Linux OS. The Windows version of Greenstone is available since long back, whereas the Windows version of DSpace has been released very recently. Though for Unix/Linux versions are more widely used for the production system, the Windows version encourages more number of users to try the softwares. Availability of these digital library softwares in both, Windows and OS-X, makes it immaterial for the institutions to choose any of them, from the operating system/platform point of view.

Open Source and Free

It is a matter of rejoice that both of them are free and open source. For such type of software a combined term 'F/OSS' or 'FOSS' is used. This means any institution interested in establishing a digital library will not be having financial burden on account of purchasing proprietary software. Open Source also guarantee continuous development on collaborative basis.

Licensee

Greenstone is available under GNU General Public License (GPL) whereas DSpace is available under BSD License. BSD licenses represent a family of permissive free software licenses. The licenses have few restrictions compared to the GNU General Public License or even the default restrictions provided by copyright, putting it relatively closer to the public domain. But as a end user of these digital library softwares, it does not pose any serious concern.

Language

Greenstone I available in multiple language pack including Indian languages like Hindi, Bengali, etc. DSpace is available in English only. For implementation at regional/local level, Greenstone may be good choice.

Associated Software

Greenstone requires Apache whereas for the installation of DSpace Apache Ant is needed. Apache Ant is actually a member of the Apache projects. Java 1.4 or later is required by both of them. For image handling, Greenstone prefers Image Magic software, which a image handling software. DSpace does not recommend any such image editing software. Any standard image editing software can work with these softwares. DSpace additionally requires PostgresSQL which is an open-source Object-Relational DBMS supporting almost all SQL and Apache Tomcat, that offers an environment for Java code to run in cooperation with a web server.

Metadata Standards

Metadata standards help in cross-domain information resource description. They also provides a simple and standardized set of conventions for describing things online in ways that make them easier to find. The metadata standards supported by Greenstone include: Dublin Core (qualified and unqualified), RFC 1807, NZGLS (New Zealand Government Locator Service), and AGLS (Australian Government Locator Service). DSpace supports Dublin Core qualified version.

Protocol

The OAI-Protocol for Metadata Harvesting (OAI-PMH) defines a mechanism for harvesting (gathering) records containing metadata (structured information about resources) from the digital libraries. Both Greenstone and DSpace follow OAI-PMH thus facilitating, metadata from many sources to be gathered together in one database, and services can be provided based on this centrally harvested, or "aggregated" data.

Support

Both Greenstone and DSpace offers comprehensive documentation for the

problem solving of the software users. The user support is provided in the form of documentation/ user manual/ guides; Wikis, FAQs, Mailing lists, training etc.

File format Supported

Both Greenstone and DSpace supports almost all of the standard formats including text, Image, audio, video etc.

Programming language

Greenstone has been written in C++, Perl, & Java. DSpace is written in Java.

Some other differences

Preservation

The act of creating any digital library collection based on open source software will contribute to the preservation of the material it contains. However, DSpace is explicitly oriented towards long-term preservation, while Greenstone is not. DSpace stores preservation metadata and includes a scheme where institutions commit to ensuring the continued availability of certain named formats.

Support infrastructure

DSpace is designed for institutional use, where there are centralized computing facilities and a competent infrastructure for software support. Greenstone is designed to be easy for anyone with basic computer-literacy skills to install, in a laptop, desktop, or institutional environment.

Author-oriented

DSpace incorporates an interface whereby users (typically authors, though some institutions choose to have librarians do this on behalf of the faculty) can submit documents to the system, and define metadata for them. Greenstone does not.

Librarian-oriented

Greenstone supplies an end-user interface with which collections can be designed, customized, and built. DSpace provides a generic design that can be tailored – but not by typical end users.

Distribution on removable media

Those who create Greenstone collections can write them to a self-installing CD-ROM that operates on all Windows systems (even obsolete ones right down to Windows 3.1/ 3.11, still in use in developing countries).

Dynamic collections (historically related to the above)

In Greenstone, adding documents to a collection normally involves rebuilding the full-text index and browsing structures (though rebuilding can be scheduled to take place automatically), whereas DSpace operates incrementally (though operations on recently-added documents like extracting text or producing image thumbnails are processed in batch mode).

International Users

Greenstone provides interfaces for readers in 40 languages, including many minority ones, and has a scheme that helps language maintainers keep the interfaces up to date when new interface features are added.

Both systems are continually evolving, and these features can change rapidly. For example, Greenstone can indeed accommodate dynamic collections by using a different search engine from the default one. Although this is probably beyond the technical capabilities of the librarian-level users that Greenstone targets, a user interface enhancement could easily rectify this. Conversely, although the default DSpace configuration is currently restricted to UNIX, it would not be hard to modify it for other operating systems. And there are some DSpace installations in languages other than English.

The difference between the two systems is largely explained by the environments in which they are designed to operate. DSpace is designed for the institutional setting, where members of faculty submit their documents to a common system that enforces common standards. Its model envisages "communities" (e.g., schools, departments, centers, labs, and programs) that contain one or more "collections" of digital "items". Greenstone is designed to allow non-specialist users to produce single, individualized, collections. Its model envisages a "librarian" who is creating collections from existing "resources" (comprising both "items" and metadata resources) and distributing them over the Web or on removable media, possibly in an international setting.

Conclusion

Digital libraries are a key technology for developing countries. They can assist human development by providing a non-commercial mechanism for distributing humanitarian information on topics such as health, agriculture, nutrition, hygiene, sanitation and water supply. Many other areas, ranging from disaster relief to medical education, also benefit from new methods of information distribution. Digital library can be created by using open source softwares available free of cost. Free and open-source software are not only "a useful and significant tool for the developing countries", but clearly have the potential to help democratization and help find solutions to the most pressing problems faced by the populations of developing countries. The growth of free, open-source software presents developing countries with an opportunity to escape from technological dependence on developed countries, but also a challenge to build up local expertise. The comparative study made in the present paper, which should help users to understand the differences and make an informed choice between the systems, indicates that both Greenstone and DSpace are appropriate softwares for creating digital libraries with

minor variation in the features and workflows. While DSpace enjoys more installations worldwide and being considered for creating large and complicated digital libraries, Greenstone is known for its simplicity and easy to use librarian and user interface. Users sometimes wonder which one to adopt. In fact, the aims of the two are very different, although their domains of application do overlap. Of the many open source systems for digital libraries, two of the most prominent are Greenstone and DSpace. Greenstone is older and more established internationally; DSpace has a more impressive institutional pedigree. There exists a • gdigital Divide• h among the countries of the worlds because of the disparity of available ICT infrastructure. But with the gradual transformation of this condition and availability of free & open source softwares for creating digital libraries, institution/ government can bridge this digital divide effectively with minimum investment. Organization also need to take steps in the areas of knowledge management and digital preservation. All these issues and areas can be taken care of by digital library softwares. Greenstone and DSpace digital library softwares cab be used for serving the people, society, nation and the world at large.

References

1. Altman, Micah. Open Source Software for Libraries: from Greenstone to the Virtual Data Center and Beyond. *IASSIST Quarterly* 2001; Winter: 5-11. Retrieved June 17, 2010, from Web site: http://www.iassistdata.org/downloads/iqvol254altman.pdf

2. Bretthauer, David. Open Source Software in Libraries. *Library Hi Tech News* 2001; 18(5): 8-9. Retrieved January 25, 2010, from Web site: http://www.emeraldinsight.com/0741-9058.htm

3. Das, Jaba. Installing Greenstone Digital Library on Windows Platform. *Workshop on Digital Libraries: Theory and Practice* 2003. Retrieved March 11, 2010, from Web site: https://drtc.isibang.ac.in/bitstream/handle/1849/142/Q_NTgreen.pdf?sequence=2

4. DSpace. http://www.DSpace.org/

5. Free Software Foundations. http://www.fsf.org/

6. Goh, Dion Hoe-Lian. A Checklist for Evaluating Open Source Digital Library Software. *Online Information Review,* 2006; 30(4): 360-379. Retrieved March 23, 2010, from Web site: http://www.emeraldinsight.com/1468-4527.htm

7. Greenstone Digital Library Software. http://www.greenstone.org/

8. Madalli, Devika P. Digital Libraries: Study into the Features of the DSpace Suite. *DRTC-HP International Workshop on Building Digital Libraries Using DSpace* 2005. Retrieved June 30, 2010, from Web site: https://drtc.isibang.ac.in/bitstream/handle/1849/238/ C _ Devika _ Dsp .pdf ? sequence=1

9. Naik, Umesha and Shivalingaiah, D. Digital Library Open Source Software: a Comparative Study. Proceedings of the *CALIBER 2006: 4th International Convention on Automation of Libraries in Education and Research Institutions* 2006; 27-39.

10. Rao, M. Koteswara. Digital Libraries: Challenges, Opportunities & Implications. *SIS 2004 Conference* 2004. Retrieved April 6, 2010. from Web site: https://drtc.isibang.ac.in/ bitstream/handle/1849/174/digital-libraries .pdf ? sequence=2

11. Singh, S.N., Ngurtinkhuma, R.K. and Singh, Pankaj Kumar. Open Source Software: a Comparative Study of Greenstone and DSpace. In Kumar, Manoj K., Eds. Proceedings of the *CALIBER 2007: 5th International Convention on Automation of Libraries in Education and Research Institutions* 2007; 89-97.

12. Smith, MacKenzie. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine* 2003; 9(1). Retrieved July 2, 2010, from Web site: http://www.dlib.org/dlib/january03/smith/01smith.html

13. SourceForge. http://sourceforge.net.

14. UNESCO. Digital Libraries and Virtual Libraries: Definitions, Concepts and Goals. *Workshop on Technical Aspects of Building Digital Libraries and Electronic Information Networks* 2003; 1-8. Retrieved July 11, 2010, from Web site: http://www.uneca.org/library/avlin/documents/digital_libraries_and_virtual_libraries.pdf

15. Witten, Ian H. Examples of Practical Digital Libraries: Collections Built Internationally Using Greenstone. *D-Lib Magazine* 2003; 9(3). Retrieved June 15, 2010, from Web site: http://www.dlib.org/dlib/march03/witten/03witten.html

16. Witten, Ian H. and Bainbridge, David. Creating Digital Library Collections with Greenstone. *Library Hi Tech* 2005; 23(4): 541-560. Retrieved July 15, 2010, from Web site: http://www.emeraldinsight.com/0737-8831.htm

17. Witten, Ian H. and others. StoneD: a Bridge Between Greenstone and DSpace. *D-Lib Magazine* 2005; 11(9). Retrieved March 5, 2010, from Web site: http://www.dlib.org/dlib/september05/ witten/09witten.html

18. Witten, Ian H., Bainbridge, David and Boddie, Stefan F. Greenstone: Open-Source Digital Library Software with End-User Collection Building. *Online Information Revie* 2001; 25(5): *288-298.* Retrieved March 12, 2010, from Web site: http://www.emeraldinsight.com/1468-4527.htm

19. Witten, Ian H., Bainbridge, David and Boddie, Stefan F. Greenstone Open-Source Digital Library Software. *D-Lib Magazine* 7(10). Retrieved March 27, 2010, from Web site: http://www.dlib.org/dlib/october01/witten/10witten.html.