Open Source Software for Developing Digital Library: Comparative Study

Shalini R. Lihitkar* and Ramdas S. Lihitkar**

*Department of Library & Information Science, RTM, Nagpur University, Nagpur-440 010

**Govt. Institute of Science, Nagpur-440 001
E-Mail: rlihitkar1975@gmail.com

ABSTRACT

The open source software (OSS) makes source code available to users, who can change the software to modify it more closely to their own requirements. Now the OSS is available for library and information management Examples of such systems include GSDL, DSpace, NewGenlib, Fedora, and Ganesha digital library software, etc. The OSS is popular with technically sophisticated users, who are often also the software developers. The OSS is becoming an increasingly popular software development method. This paper highlights the comparison of features, functions and usability of OSS, i.e., GSDL, DSpace, E-Prints, Fedora, Ganesha, Invenio, XTS, Dienst, VuDL, and NewGenlib. Ranking of the software have been done based on the assigned points for each criteria. The GSDL scored maximum points, i.e., 47 and hence it is in rank first followed by VuDL which scored 43 points.

1. INTRODUCTION

Open source software (OSS) is software that includes source code and is usually available at no charge. There are additional requirements besides the availability of source code that a program must meet before it is considered open source including: the software must be free to redistribute; derivative works must be allowed; the license cannot discriminate against any persons; and the license cannot discriminate against any fields of endeavour. Software that is licensed under an open source license allows for a community of developers from around the world to improve the software by providing enhancements and bug fixes[1].

The services and the collection of the libraries and information centers are becoming global due to the application of information and communication technology. Now information can be accessed from the remote places with the help of internet. Due to shrinking budgets and the increasing prices of journals, librarians have to look forward to a new alternative by which they can collect, store, arrange, and disseminate information to the users. The concept of open access and institutional repository (IR) has evolved to find out the solutions. In building the IR the college libraries can take the help of the OSS[13].

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 program can be made.

Object code*:* Actual 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

- The sources may be studied and changed

Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve the software. More precisely, it refers to four kinds of freedom, for the users of the software to

- Run the program

- Study how the program works
- Redistribute copies so one can help neighbour
- Improve the program, and release improvements to the public, for the benefit of whole community.

2. FEATURES OF OSS

Open source doesn't just mean access to the source code. The distribution terms of OSS must comply with the following criteria:

- Free redistribution
- Source code
- Derived works
- Integrity of the author's source code
- No discrimination against persons or groups
- No discrimination against fields of endeavour
- Distribution of license
- License must not be specific to a product.
- license must not restrict other software

- No provision of the license may be predicated on any individual technology or style of interface[12]

3. AIMS AND OBJECTIVES

- To compare 10 OSS for building digital library.
- To find out and rank the more user-friendly OSS based on the comparative study.

4. METHODOLOGY AND SCOPE

Investigative and evaluative research methodology were used for the study. Data were collected:

- By surfing internet and downloading ten OSS such as Greenstone Digital Library (GSDL), Dspace, Ganesha, Fedora, E-Prints, Invenio, Dienst, VuDL, XTS, NewGenlib .
- Comparing the selected OSS.
- Preparing the worksheet by using different criteria of selected OSS.

Scope of the study is restricted to surfing internet and downloading the 10 OSS i.e. GSDL, DSpace, Ganesha Fedora, E-Prints, Invenio, Dienst, Vu-DL, XTF, and NewGenlib (Table 1).

Table 1. Selected open source software

Software	URL	Free downloadable	Developed by	Contact information
GSDL[2]	http://www.GSDL.org	✓	New Zealand Digital Library Project with UNESCO and the Human Info NGO	GSDL-users@list. scms waikato.ac.nz
DSpace[4-5]	http://www.dspace.org/	✓	Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard	sales@dspace.com.au
Ganesha[6]	http://gdl.itb.ac.id/	✓	Indonesian Digital Library Network	mrg@kmrg.itb.ac.id
Fedora	http://www.fedora-commons.org	✓	Cornell University Information Science and the University of Virginia Library	http://www.fedora-commons. org/contact-info
E-Prints	http://software.eprints.org	✓	University of Southampton	hans.falk.hoffmann@cern.ch
Invenio	http://invenio-software.org/	✓	CERN Document Server	http://invenio-software.org/wiki/ Community/MailingLists
Dienst	http://www.cs.cornell.edu/ cdlrg/dienst/DienstOvervi ew.htm	✓	CS-TR Project (Corporation for National Research Initiatives)	help@ncstrl.org
VuDL	http://vudl.org/	✓	Villanova University's Falvey Memorial Library	libtech@villanova.edu
XTF	http://xtf.cdlib.org/	✓	California Digital Library (CDL)	http://groups.google.com/group/xtf-devel
NewGenlib[9]	http://www.newgenlib.com	✓	Kesavan Institutes of Information & Knowledge Management	haravu@newgenlib.com siddharthe@newgenlib.com

5. ANALYSIS AND INTERPRETATIONS

There are various OSS but for the current study only 10 software were selected. The features of software were studied by applying some criteria like, searching facilities, platform to run software, associated software, languages included in software, searching parameters of software and after analysing grading of 10 OSS have been done based on the assigned points for each parameters.

Table 2 shows the latest version of software that is readily available for the use and also gives the information about the year in which new version was released it also specifies, the site address from which the user can download the particular software. It is observed from Table 4 that maximum, i.e., 10 pre-requisite software have been required by E-Prints software followed by Fedora, i.e. 9. Dienst and NewGenlib required minimum three pre-requisite software.

Table 5 shows that the selected OSS run on different operating systems. Some of the software run on Windows version (Windows 95, Windows 98, Windows 2000, Windows NT and Windows XP, Windows 7), Linux, UNIX, and other. Maximum 9 points has been acquired by E-Print because it is window-based as well as it can be run on LINUX/UNIX based operating system and also run in any other operating system.

Table 6 shows languages that are supported by the selected OSS and are readily available for use. As per requirement, the user can choose and change the language according to his convenience for handling the software (Table 7). Table 8 shows the facilities provided by selected OSS such as provision of searching or browsing, multi-languages support multimedia. It also gives the information of other facilities such as web OPAC, metadata, catalogue, retro-conversion, internet, intranet, and extranet, etc. Maximum selected OSS provides all facilities' except Ganesha, XTF, and NewGenlib. Maximum numbers (10) of searching parameters have been found in GSDL followed by DSpace. A maximum point 14 has been acquired by GSDL followed by DSpace and Invenio.

Table 2. License, new version, downloaded site and size of the selected open source software

Name of software	License	Version used	Downloaded from
GSDL[2]	GNU	2.62 in 2005	http://prdownloads.sourceforge.net/GSDL
DSpace[4-5]	The BSD license	1.6.0 in 2010	http://sourceforge.net/projects/dspace/ 8.94
Ganesha[6]	GNU	4.0 in 2004	http://gdl.itb.ac.id/download/
Fedora	ECL (Apachi derived)	3.4.2 in 2011	http://sourceforge.net/projects/fedora-commons/files/fedora/3.4.2/fcrepo-installer-3.4.2.jar/download?use_mirror=space
E-Prints	GPL	3.3.1 in 2011	http://files.eprints.org/669/1/eprints%2D3.3.1.tar.gz
Invenio	GPL	v1.0.0-rc0 in 2010	http://invenio-software.org/wiki/Installation/Download
Dienst	Cornell University (CU) licence	3.0 protocol 2007	ftp://ftp.cs.cornell.edu/pub/Dienst/Dienst.tar.Z
VuDL	GPL open source license	0.4.1 Alpha in 2011	http://vudl.org/vudl/downloads/
XTF	BSD license	3.0 in 2011	http://xtf.cdlib.org/download/
NewGenlib[9]	GNU	2.2 in 2005	http://sourceforge.net/projects/Newgenlib/

Table 3. Pre-requisite/associated software for installation of selected open source software

Software	Apache		C++	Java		Oracle	MySQL	PHP	Perl	Postgre	SQL
	0.2	1.5		1.4	J2E					7.3	later
GSDL	✓	✓	✓	×	✓	×	✓	✓	✓	×	×
DSpace	×	✓	×	✓	✓	✓	×	×	×	✓	✓
Ganesha	×	✓	×	×	×	×	✓	✓	✓	×	×
Fedora	✓	✓	x	✓	✓	✓	✓	✓	×	✓	✓
E-Prints	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓
Invenio	✓	✓	×	×	✓	×	✓	×	×	×	×
Dienst	✓	✓	×	×	×	×	×	×	✓	×	×
VuDL	✓	✓	×	✓	✓	×	×	✓	×	×	×
XTF	✓	✓	×	✓	✓	×	×	✓	×	×	×
NewGenlib	×	×	×	✓	×	×	×	×	×	✓	✓

Table 4. Analysis of pre-requisite software for installation

Software	No. of pre-requisite software	Percentage
GSDL	7	63.63
DSpace	6	54.54
Ganesha	4	36.36
Fedora	9	81.81
E-Prints	10	90.90
Invenio	4	36.36
Dienst	3	27.27
VuDL	5	45.45
XTF	5	45.45
NewGenlib	3	27.27

Table 5. Operating system required for selected open source software

Name of software	Windows (4 Points)	Unix (2 Points)	Linux (2 Points)	Other (1 Point)	Points gained
GSDL[7]	✓	×	✓	×	6
DSpace[4-5]	✓	✓	✓	×	8
Ganesha[6]	✓	✓	✓	×	8
Fedora	×	✓	✓	✓	5
E-prints	✓	✓	✓	✓	9
Invenio	×	✓	✓	×	4
Dienst	×	✓	✓	✓	5
VuDL	✓	✓	✓	×	8
XTF	✓	✓	✓	×	8
NewGenlib[9]	x	x	✓	x	2
Total	**06**	**08**	**10**	**03**	**-**

Table 6. Common languages included in selected open source software

Name of software	Dutch	English	French	Languages Kazakhs	Russian	Spanish	Thai	Other	Points gained
GSDL[8]	✓	✓	✓	✓	✓	✓	✓	✓	9
DSpace[4-5]	✓	✓	✓	×	✓	✓	×	×	6
Ganesha	×	✓	✓	×	×	×	×	×	2
Fedora	✓	✓	✓	✓	✓	✓	×	✓	8
E-Prints	✓	✓	✓	✓	✓	✓	×	✓	8
Invenio (26 languages)	✓	✓	✓	✓	✓	✓	✓	✓	9
Dienst	✓	✓	✓	×	✓	✓	✓	✓	8
VuDL	✓	✓	✓	✓	✓	✓	✓	✓	9
XTF	×	✓	✓	×	×	✓	×	×	4
NewGenlib	×	✓	×	×	×	×	×	×	2

Table 7. Total number of languages included in the software

Software	Languages supported
GSDL[10]	8
DSpace[11]	5
Ganesha	2
Fedora	7
E-Prints	7
Invenio	8
Dienst	7
VuDL	8
XTF	3
New Genlib	1

share is that they are flexible, and can be customised and modified at many different levels–including the programming level since they are open source systems. This gives the ultimate flexibility and yields significant advantages over closed-source systems.

(a) All the OSS are freely available and some are under the GNU (General Public License) license.

(b) It is also observed that to run this software they need pre-requisite software. This software can be used to make sophisticated computational techniques accessible to everyone.

Table 8. Facilities provided by selected open source software

Name of software	Searching/ browsing	Multilingual support	Multimedia	Web OPAC	Metadata/ dublin core	Catalogue	Retro Conv.	Internet	Intranet	Extranet	Points gained
GSDL[3]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
DSpace[4-5]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Ganesha	✓	✓	X	✓	✓	✓	X	✓	✓	✓	14
Fedora	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
E-Prints	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Invenio	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Dienst	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
VuDL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
XTF	✓	✓	X	✓	✓	✓	✓	X	X	X	11
New Genlib	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	16

Table 9. Total number of facilities provided by open source software

Software	No. of facilities	Percentage (%)
GSDL[3]	10	100
DSpace[4-5]	10	100
Ganesha	8	80
Fedora	10	100
E-Prints	10	100
Invenio	10	100
Dienst	10	100
VuDL	10	100
XFT	6	60
NewGenlib	9	90

6. FINDINGS AND CONCLUSIONS

The study of OSS is restricted within the study of the features of the software. They represent rather different perspectives, and have different and in many ways complementary, goals and strengths. One goal they

(c) GSDL and DSpace software's metadata is difficult to create because cataloguer entry operator must know xml or html language, and only expertise in this language can do this work very easily, while in the other software there is no need to have the knowledge of xml or html languages.

(d) All the 10 selected OSS support internet, and intranet searching/browsing, multimedia, and web OPAC, etc., except XTF software. In this brief study, it is clear that this software would be applicable to fully digitised library and most of the library material can be expanded in the form of CDs, DVDs, etc.

(e) Among ten selected OSS, E-Print required maximum number of pre-requisite software for installing, i.e., 10 (90.90 %), followed Fedora (81.81 %) and GSDL (63.63 %).

(f) It has been observed that E-Prints software runs on maximum number of operating system while

Table 10. Searching parameters provided by selected open source software

Software	Title	Author	Subject	keyword	Issue Date	Format	Publisher	Class no.	News paper clipping	Manuscripts/ Rare collection	Points gained
GSDL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14
DSpace	✓	✓	✓	✓	✓	×	×	×	✓	✓	10
Ganesha	✓	✓	✓	✓	×	×	×	×	✓	×	07
Fedora	✓	✓	✓	✓	×	×	×	×	×	×	05
E-Prints	✓	✓	✓	v	✓	×	✓	×	×	×	07
Invenio	✓	✓	✓	✓	×	✓	✓	×	×	✓	10
Dienst	✓	✓	✓	✓	×	✓	✓	×	×	×	08
VuDL	✓	✓	✓	×	✓	✓	✓	✓	×	×	08
XTF	✓	✓	✓	×	✓	✓	✓	×	×	×	07
NewGenlib	✓	✓	✓	✓	×	✓	×	×	×	×	07

Table 11. Total number of searching parameters in the software

S.No.	Software	No. of searching parameters	Percentage (%)
1.	GSDL	10	100
2.	DSpace	7	70
3.	Ganesha	5	50
4.	Fedora	4	40
5.	E-Prints	6	60
6.	Invenio	7	70
7.	Dienst	6	60
8.	VuDL	7	70
9.	XTF	6	60
10.	NewGenlib	5	50

NewGenlib runing only in Linux operating system.

(g) OSS is supporting different language, GSDL, Invenio and VuDL support maximum number of languages, i.e., 9 languages as compared to other software whereas Ganesha software supports only one language, i.e., English (Table 6).

(h) GSDL, DSpace, Fedora, E-Prints, Invenio, Dienst, and VuDL are providing maximum number of facilities, i.e., 10 (100 %) as compare to Ganesha software, XTF, and NewGenlib (Table 8).

(i) Searching parameters of each software were studied and it is observed that GSDL has maximum number of searching parameters, i.e., 100 per cent, 70 per cent of searching parameters are provided by DSpace, Invention, and VuDL whereas Ganesha, E-Prints, XTF and NewGenlib software provided less than 70 per cent of searching parameters (Table 10 and 11).

(j) It is also recorded that GSDL, DSpace, E-Prints,

Invenio and VuDL comes under excellent grades (40-50 points).

(k) GSDL scored maximum number of points i.e., 47 followed by VuDL, E-Print, DSpace and Invenio. (Table 12 and 13).

7. CONCLUSIONS

The implementation of OSS in library represents a method for improving library services. All OSS is governed by some type of license agreement. Library professionals should know how to set up and build digital library collection in their organisation.

The OSS interface makes it easy for people to create their own library collections. Collections may be built and served locally from the users 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

Table 12. Ranking of selected open source software

Parameters	Points assigned	GSDL	DSpace	Gnesha	Fedora	E-Print	Invenio	Dienst	VuDL	XTF	Newgenlib
Operating system											
Windows	4	4	4	4	0	4	0	0	4	2	
UNIX	2	0	2	2	2	2	2	2	2	2	0
LINUX	2	2	2	2	2	2	2	2	2	2	0
Other	1	0	0	0	1	1	0	1	0	1	0
Languages											
Dutch	1	1	1	0	1	1	1	1	1	0	0
English	2	2	2	2	2	2	2	2	2	2	2
French	1	1	1	1	1	1	1	1	1	1	0
Kazakhs	1	1	0	0	1	1	1	0	1	0	0
Russian	1	1	1	0	1	1	1	1	1	0	0
Spanish	1	1	1	0	1	1	1	1	1	1	0
Thai	1	1	0	0	0	0	1	1	1	0	0
Other	1	1	0	0	1	1	1	1	1	0	0
Facilities											
Searching/browsing	1	1	1	1	1	1	1	1	1	1	1
Multilingual support	2	2	2	2	2	2	2	2	2	2	2
Multimedia	2	2	2	0	2	2	2	2	2	0	2
Web OPAC	2	2	2	2	2	2	2	2	2	2	2
Metadata	2	2	2	2	2	2	2	2	2	2	0
Catalogue	2	2	2	2	2	2	2	2	2	2	2
Retro-conversion	2	2	2	0	2	2	2	2	2	2	2
Internet	2	1	1	2	2	2	2	2	2	0	2
Intranet	2	1	1	2	2	2	2	2	2	0	2
Extranet	1	1	1	1	1	1	1	1	1	0	1
Searching parameters											
Title	1	1	1	1	1	1	1	1	1	1	1
Author	1	1	1	1	1	1	1	1	1	1	1
Subject	1	1	1	1	1	1	1	1	1	1	1
Keyword	2	2	2	2	2	2	2	2	0	0	2
Issue date	1	1	1	0	1	1	0	0	1	1	0
Format	2	2	0	0	0	0	2	2	2	2	2
Publisher	1	1	0	0	1	1	1	1	1	1	0
Class no.	1	1	0	0	0	0	0	0	1	0	0
News paper clipping	2	2	2	2	0	0	0	0	0	0	0
Communities/rare collection	2	2	2	0	0	0	2	0	0	0	0
Total	**50**	**47**	**42**	**32**	**36**	**42**	**41**	**39**	**43**	**31**	**27**

collections can be updated and new ones brought online at any time. The OSS has much potential for libraries and information centers, and there are a number of projects, including GSDL, DSpace, Ganesha, E-prints, Fedora, etc., that demonstrate capabilities 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, suggesting enhancements, and testing new versions.

Currently available OSS projects cover application areas ranging from the traditional library management systems to innovations like GSDL and DSpace, which complement traditional systems. Benefits include low costs of maintenance, greater accessibility, and better prospects for long-term preservation of scholarly works. The GSDL and DSpace are recommended to build the digital libraries and make them accessible globally.

Table 13. Analysis of grading of selected open source software

S. No.	Software	Excellent	Very Good	Good	Average	Poor
1.	GSDL	✓	x	x	x	x
2.	DSpace	✓	x	x	x	x
3.	Ganesha	x	✓	x	x	x
4.	Fedora	x	✓	x	x	x
5.	E-Prints	✓	x	x	x	x
6.	Invenio	✓	x	x	x	x
7.	Dienst	x	✓	x	x	x
8.	VuDL	✓	x	x	x	x
9.	XTF	x	✓	x	x	x
10.	NewGenlib	x	x	✓	x	x

Grading based on the points:

(1) 40-50 ➜ Excellent; (2) 30-40 ➜ Very Good; (3) 20-30 ➜ Good; (4) 0-20 ➜ Average, (5) Below 10 ➜ Poor

REFERENCES

1. Biswas, G. & Paul, D. An evaluative study on the open source digital library software for institutional repository: Special reference to DSpace and Greenstone digital library. *Inter. J. Lib. Inf. Sci.,* 2010, **2**(1), 1-10. http://www.academicjournals.org/ijlis/PDF/pdf2010/Feb/Biswas%20and%20Paul.pdf

2. Building digital library collection. http://www.sivulile.org/workshops/uct_greenstone (accessed on 14 January 2011)

3. Digital libraries with Greenstone: An open source solution. http://www.iwu.edu/library/services/argus1.htm (accessed on 10 January 2012)

4. DSpace Foundation. http://www.dspace.org/ (accessed on 17 January 2012).

5. DSpace. http://www.dspace.org/conference/presentations/architecture.ppt. (accessed on 16 January 2012).

6. Free software Ganesha Digital Library (GDL). http://gdl.itb.ac.id/. (accessed on 6 January 2012)

7. GSDL information. http://www.asis.org/Bulletin/May-02/sigiii.html. (accessed on 20 January 2012)

8. GSDL collection. http://gsdlsupport.iimk.ac.in. Greenstone Documentation. http://www.iwu.edu/ckelch/ArgusProjectDoc12.pdf

9. http://NewGenLibCheckoufunctionsnewgenlib.html

10. http://www.dspace.org. (accessed on 5 January 2012).

11. http://www.GSDL.org. (accessed on 5 January 2012).

12. Lihitkar, S. Comparative study of GSDL, Dspace and Ganesha: Open source software. *IJLIS*, 2011.

13. Meitei, L.S. & Devi, P. Open source initiatives in digital preservations: The need for an open source digital repository and preservation system. *In* CALIBER 2009. http://hdl.handle.net/1944/996. (accessed on 6 January 2012).

About the Authors

Dr. (Mrs) Shalini R. Lihitkar is presently working as Assistant Professor (Sr. Grade) and Head in Dept of Library and Information Science (DLISc), Rastrasant Tukadoji Maharaj (RTM), Nagpur University, Nagpur. She has published more than 110 papers in seminars, conferences and journals of national and international repute. She has authored three books, edited 8 national conference proceedings and compiled various course materials for LIS professionals. She has completed Information Literacy Project with UNESCO and SALIS as a member, and UGC-Minor Research Project as Principal Investigator. She is Project Director of Research Project of ICSSR. She is an expert member in Definitional Glossary of LIS, project of Ministry of Scientific and Technical Terminology. She is a recognised research supervisor in LIS of RTM, Nagpur University, and S.G.B. Amravati University, Amravati.

Dr. Ramdas S. Lihitkar is presently working as Librarian, Govt. Institute of Science, Nagpur. He worked as a Librarian Govt. Rajaram College, Kolhapur from 2001-2003 and Govt. College of Education, Bhandara from 2005-2008. He has published more than papers in seminars, conferences and journals at national and international level. He is a recognised research guide for PhD. in LIS in the faculty of Social Science of RTM, Nagpur University, Nagpur.