

# Open Source Software: the savior for digital library development in Pakistan

Farzana Shafique<sup>1</sup>, Sundus Tariq<sup>2</sup>

<sup>1</sup>Deanship of Library Affairs University of Dammam, Saudi Arabia

<sup>2</sup>Department of Library and Information Science, IUB

e-mail: alakhdarem2001@yahoo.com

## ABSTRACT

*There are many open source software available for organizations and individuals to create digital libraries (DLs). Open Source Softwares (OSS) are a good choice for low income and developing countries such as Pakistan. However, the challenge is to select the most suitable software for developing the digital library. This carefully made choice can be a savior for digital library development in Pakistan. However, a simple instrument to evaluate these DL software packages does not exist. The objectives of the present work were not only to develop a checklist for evaluating the Digital Library Software packages but also to compare the features of two of the most used open source DL software against this checklist. This evaluation helped in recommending the most suitable DL software for libraries in Pakistan. To achieve these objectives, a comprehensive checklist was developed to assess the merits and demerits of digital library (DL) softwares. Features that were characterized as "good" open source DL software were determined after reviewing the relevant literature and the features of many state of the art DL software packages. Firstly, after the identification of essential feature categories of DL software, a checklist consisting of 12 main categories was developed. These 12 feature categories were further subdivided. The two most used DL software in Pakistan, i.e. GreenStone and DSpace were assessed and compared against this checklist. After evaluation against this checklist, GreenStone was found to be the better performer compared to DSpace. GreenStone was the only software package that held most of the features in the checklist.*

**Keywords:** Digital libraries, DL software, Open source software, Pakistan

## INTRODUCTION

The digital era has not only influenced the society but all the social institutions as well, libraries are no exceptions to this. This digital revolution has not only influenced the services of libraries but collection formats as well. New terms have emerged in recent years to represent this digital influence on libraries. The term of digital libraries was introduced not only to represent the new collection format but a new mode of accessing the library collection as well. According to Prytherch (2005), the digital library is an umbrella term for conceptual modes of libraries of the future that focuses on the provision of services associated almost totally with digital content and to describe those aspects of existing library services that have significant components. In a broader context, digital libraries can be defined as organizations that provide resources including specialized staff, to select, structure and offer intellectual access, to interpret, distribute, preserve the



integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities (Feather and Sturges 2003).

According to Hull (2001), libraries of all sizes and types are embracing digital collections, although most libraries will continue to offer both print and digital collections for many years to come. However, new purchases of journals, magazines, and abstracting and indexing services are heavily weighted toward digital. Digital books (e-books) are only beginning to make a presence in library collections. One must rely on hardware and software in order to view the digital contents of digital resources. Breeding (2002) believes that one of the key functionalities of a DL should be the matching of user work patterns. To achieve this, a thorough understanding of the users of libraries and the system itself should be obtained. Apart from the need for a deeper understanding of the users, the fit between the tools used to support the DL and the necessary requirements has to be ascertained.

There are wide varieties of digital library software; some are free or open source software and some are commercial ones. The concept of developing digital libraries and use of open source software has been widespread within the last few years. In developing countries, open source softwares are used more for developing digital libraries. Hussain (2006); and Hussain, Mahmood, and Shafique (2008) found that only few years back, there was little awareness among libraries and information professionals in Pakistan about the development of digital libraries. However, in recent years the concept of digital libraries has become pervasive in Pakistan. For example, Shafique and Tariq (2011) concluded, after a focus group discussions with information professionals in Pakistan that many library and information professionals are more aware of the concept and many libraries in Pakistan have begun to develop their digital repositories or are planning to develop them. They also mentioned that GreenStone and DSpace, the two open source digital library software were among the most used software. Thus keeping the findings of researchers in Pakistan (Hussain 2006; Hussain, Mahmood and Shafique 2008; Shafique 2009; Shafique, and Tariq 2010) in view, two open source digital library (DL) software i.e., Green Stone and DSpace were selected for evaluation against the checklist developed by these researchers.

### **OBJECTIVES OF THE STUDY**

No comprehensive study has been conducted in Pakistan at a local or national level to find out the most suitable digital library software. Neither has any standard tool or directory been available for the selection and evaluation of the digital library software, nor has any comprehensive literature been written and published in Pakistani scenario discussing the technical and practical aspect of digital library software. In this perspective, the following are the objectives of the study:

1. To construct a comprehensive checklist of criteria for evaluating the digital library software.
2. To compare the two most used open source and free digital library software (i.e., GreenStone and DSpace) against this checklist.
3. To recommend the most suitable Open Source Digital Library Software for libraries in Pakistan.



## METHODOLOGY

The two most used DL software in Pakistan, GreenStone and DSpace are compared against the checklist. A checklist consisting of 12 categories of items was developed. In this paper only two open source digital library software were compared on the basis of checklist.

## FINDINGS AND ANALYSIS

This section compares the GreenStone and DSpace software against the checklist developed by the researchers. Table-1 provides a tabular view of the checklist. There are 12 main categories in the checklist i.e., *Content Management; Content Acquisition; Metadata; Search Support; Access Control and Privacy; Reports and Inquiry Capabilities; Preservation; Interoperability; User Interface; Standard Complaisance; Automatic Tools; and System Support and Maintenance*. Each category is further divided into sub categories. The details of all the categories covered in the checklist and an analysis of GreenStone and DSpace against this checklist is also presented. The description of that analysis is as follows:

### 1. Content Management

Content management involves procedure and tools pertaining to the submission of content in to digital library as well as management of the submission process

- 1.1 Submission Management: Only GreenStone allows multiple collections within same installation system, repository administration to set submission parameter and provides home page for each collection. These features are absent in DSpace
- 1.2 Submission Workflow: Both the software do not provide segregated submission workspace and configurable submission roles within collection, while submission roles is only available in DSpace
- 1.3 Submission Support: Facility of E- mail notification for users and administration is only available in DSpace
- 1.4 Submission Review: Feature of allowing users to review complete and incomplete contents and allowing content administration to review submission are available only in GreenStone digital library software, while DSpace does not have these features

### 2. Content Acquisition

Content acquisition refers to content import/export, versioning and support documents formats

- 2.1 Content Import / Export: Only GreenStone allows uploading of compressed files and uploading from existing URL. It also supports volume import of objects and metadata for existing volume
- 2.2 Document/Object Format: Administrator ability to limit approved file format: In GreenStone, the administrator has ability to limit approved file format, and submission items can comprise multiple files or file type
- 2.3 Text Format: Both the digital library software support ASCII as text format, while only GreenStone supports UNICODE and RTF as well
- 2.4 Image Format: GreenStone supports image formats such as TIFF, GIFF, and JPEG
- 2.5 Presentation Formats: Both the software support adobe post script, while only GreenStone supports PDF
- 2.6 Structured Format: HTML and XML as structured formats are supported by both the software, while SGML is only supported on by GreenStone



- 2.7 Audio and Video Formats: Only GreenStone digital library software supports Wave, MP3 as audio and video formats, while MPEG is supported by both the software
- 2.8 Version Control: Both the software allow past version of files to be retrieved, identification and composition of the change.

### 3. Metadata

Metadata support in DLs is vital in content indexing, storage, access, and preservation

- 3.1 Real time Updating and Indexing of Accepted Content Features Available: GreenStone and DSpace software support the following metadata standards; DUBLIN CORE, MARC21. On the other hand, none of the software supports EAD, LOM, MODS and VRA Core categories
- 3.2 Add /Delete Customized Metadata Fields: This feature is not available in any Software
- 3.3 Set Default Values for data: Only GreenStone allows to set default values for data.
- 3.4 Support Unicode Character set for Metadata: This feature is not available in any software

### 4. Search Support

- 4.1 Full Text: Only GreenStone supports full text search with Boolean logic. Other search options such as use of Truncation/wild cards, Phrased and Proximity search are not supported by both the software
- 4.2 Search all Descriptive Metadata: Searching of all descriptive Metadata with Boolean logic is available only in GreenStone, while use of Truncation/wild cards for this purpose is not supported by both the software
- 4.3 Search Selected Metadata: Browsing by author, title, issue date is supported by both the software, while searching by subject term, collection, and customized fields is supported only by GreenStone
- 4.4 Sort Search Results: Sorting of search results by author, title is possible in both the software, while sorting by issue date is only supported by DSpace. However, sorting by relevancy or by other criteria is not possible in both the software

### 5. Access Control and Privacy

Access control and privacy include the administration of passwords, as well as the management of the users' accounts and rights to specified location within digital library.

- 5.1 Password Administration: Both the software do not allow assigning of passwords by the system. However, both the software support the user selected passwords and retrieval of the forgotten software
- 5.2 User Management: The feature of adding and deleting of users' profile is not available in both the software. However, editing of users' profile is possible in both the software
- 5.3 Limits Access at Different Level: Features like File/object and collection levels are not available in both the software
- 5.4 Users' Roles: Both the software do not support features related to users' roles such as allowing definition of different user groups; limiting access by role; and allowing collection to be customized for each role
- 5.5 Access Management: Similarly the Access Management related features are also missing both the software such as IP source address and Proxy filtering
- 5.6 Credential-based Access: Credential-based Access features like encryption and digital signature are also not supported by both the software.



## 6. Reports and Inquiry Capabilities

This category is concerned with usage monitoring and reporting.

- 6.1 System Generated Usage Statistics: System generated usage statistics is possible only in GreenStone
- 6.2 Usage Report: The features related to Usage Report such as report timeline specification; report fields customization; and reports templates are available only in GreenStone.

## 7. Preservation

Preservation refers to preservation of metadata and quality control measure to ensure integrity, and persistence documentation identification for migration purposes

- 7.1 Persistence Documents Identification: System assigned identification and CNRI handle features are not available in both the software
- 7.2 Quality Control: This feature is not available in both the software
- 7.3 Prescribed Digital Preservation Strategy: This feature is not available in both the digital library software.

## 8. Interoperability

Interoperability is concerned with the benefits of integration distribution collection and system

- 8.1 OAI-PMH: OAI-PMH feature is available in both the digital library software
- 8.2 Z39.50 Protocol complaint: Z39.50 Protocol is only supported by GreenStone
- 8.3 Research Protocol: Dients and SDLIP features are not available in both the digital library software

## 9. User Interface

This category deals with support for multilingual access as well as the ability to customize the user interface to suit the needs of different digital library implementation

- 9.1 Modify Interface "Look and Feel": In DSpace and GreenStone the modification of interface according to users' needs is possible
- 9.2 Apply a customized Header /Footer to Statistics or Dynamic Pages: Both the digital library software support customized header/footer to statistics
- 9.3 Support Multi-lingual Interface: DSpace and GreenStone also support multilingual interfaces.

## 10. Standard Complaisance

Standards are important for sharing the digital contents and long-term digital preservation (Dawson 2004)

- 10.1 Structured Document Formats e.g. XML, SGML: DSpace and GreenStone support structured document formats
- 10.2 Metadata Formats e.g. Dublin core: Standard metadata formats are also supported by DSpace and GreenStone
- 10.3 Text Format e.g. UNICODE: Similarly the text formatting feature is also supported by both the software
- 10.4 Image Formats e.g. TIFF: Only GreenStone supports this feature on the other hand DSpace does not support it



### 11. Automatic Tools

This category refers to tools for automated content acquisition, harvesting, and metadata generation. In this context of DLs, automatic tools are useful for maintenance and can reduce labor cost, especially for large collection (Arms 2000)

11.1 Metadata entry system: This feature is available only in GreenStone

11.2 Generation of Search indexes: The features related to Generation of Search Indexes are available only in GreenStone, such as Search Indexes; HTMML; and Reports

### 12. System Support and Maintenance

Support and maintenance are important aspects in all software system, and open source software is often criticized to be lacking these. It contains documentation, manuals, mailing list, discussion forum, bug tracking, feature request system, and formal helpdesk.

12.1 Documentantation / Manuals: This feature is available only in GreenStone

12.2 Mailing list / Discussion Forums: Both the digital library software support this Option

12.3 Bug Tracks / Features Requested System: This feature is also supported by both the digital library software

12.4 Helpdesk Support: Helpdesk support is also provided by both the software.

## CONCLUSION

The checklist developed by the researchers can be helpful for evaluating the digital library software either commercial or open source. The checklist is consisting of 12 main categories. The 12 main categories i.e., *Content Management; Content Acquisition; Metadata; Search Support; ACCES Control and Privacy; Reports and Inquiry Capabilities; Preservation; Interoperability; User Interface; Standard Complaisance; Automatic Tools; and System Support and Maintenance* are further divided in sub categories. The two most used DL software in Pakistan, i.e. GreenStone and DSpace were assessed and compared against this checklist. The selection of these two open source software was in the light of reviewed literature (Hussain 2006; Hussain, Mahmood and Shafique 2008; Shafique 2009; Shafique, and Tariq 2011).

Above comparison shows that most of the features are available only in GreenStone. It is the only software package that holds most of the features included in the checklist. Literature findings also rate GreenStone as a better and most used software (Hussain 2006; Hussain, Mahmood and Shafique 2008; Shafique 2009; Shafique, and Tariq 2011). Following recommendations are made in the light of findings of the study:

- The librarians interested in developing the digital libraries should conduct a field survey before selecting any software. They should consult the librarians who are already successfully running their digital library projects.
- The checklist developed by the researchers can be used for evaluating any digital library software before selection.
- The open source digital library software can be a good and economical source in Pakistani perspective.

Table 1a: Comparison of GreenStone and DSpace

No.	Checklist Categories	GreenStone	DSpace
<b>1.</b>	<b>CONTENT MANAGEMENT</b>		
1.1.	Submission Admission		
	Allows repository administrator to set Submission parameters	Y	N
	home page for each collection	Y	N
1.2.	Submission Work Flow		
	Segregated submission workspace	N	Y
	Submission roles	N	N
	Submission roles	N	N
1.3.	Submission Supports		
	Email notification for users	N	Y
	Email notification for administrators	N	Y
1.4.	Submission Review		
	Allows users to review complete content	Y	N
	Allows users to review incomplete content	Y	N
	Allows content administrator to review submission	Y	N
<b>2.</b>	<b>CONTENT ACQUISITION</b>		
2.1.	Content Import/ Export		
	Upload compressed files	Y	N
	Upload from existing URL	Y	N
	Volume import of object	Y	N
	Volume import of metadata for existing collection	Y	N
2.2.	Document/ Object Formats		
	Administrator ability to limit approved file formats	Y	N
	Submission items can comprise multiple files or file type	Y	N
2.3.	Text Format		
	ASCII	Y	N
	UNICODE	Y	N
	RTF	Y	N
2.4.	Image Format		
	TIFF	Y	N
	GIF	Y	N
	JPEG	Y	N
2.5.	Presentation Formats		
	Adobe Post Script	Y	N
	Adobe PDF	Y	Y
2.6.	Structured Format		
	HTML	Y	Y
	SGML	Y	N
	XML	Y	Y
2.7.	Audio And Video Formats		
	Wave	Y	N
	MP3	Y	N
	A VI	Y	Y
	MPEG	Y	Y
2.8.	Version Control		
	Allow past version of files to be retried	Y	Y
	Change can be composed	Y	Y
	Change can be identified	Y	Y



Table 1b: Comparison of GreenStone and DSpace

No.	Checklist Categories	GreenStone	DSpace
<b>3.</b>	<b>METADATA</b>		
	Real time updating and indexing of accepted content	N	N
	Features available		
	Metadata scheme support	N	N
	Dublin core	Y	Y
	EAD	N	N
	MARC21	N	Y
	LOM	Y	N
	METS	N	N
	MODS	N	N
	VRA Core categories	N	N
	MPEG-7	N	N
3.1	Add/Delete Customized Metadata Fields		
	Feature available	N	N
	Set default values for data	Y	N
	Feature available	N	N
	Support Unicode character set for Metadata	N	N
	Feature available	N	N
<b>4.</b>	<b>SEARCH SUPPORT</b>		
4.1	Full Text		
	Boolean logic	Y	N
	Truncation /wild cards	N	N
	Phrase	N	N
	Proximity	N	N
4.2	Search All Descriptive Metadata		
	Boolean	Y	N
	Truncation/ wild cards	N	N
4.3	Search Selected Metadata		
	Feature available	N	N
4.4	Browse		
	By author	Y	Y
	By title	Y	Y
	By issue date	Y	N
	By subject term	Y	N
	By collection	Y	N
	By customized fields	Y	N
4.5	Sort Search Result		
	By author	Y	Y
	By title	Y	Y
	By issue date	N	Y
	By relevance	N	N
	By other	Y	N



Table 1c: Comparison of GreenStone and DSpace

No.	Checklist Categories	GreenStone	DSpace
<b>5.</b>	<b>ACCESS CONTROL AND PRIVACY</b>		
5.1	Password Administration		
	System-assigned password	Y	Y
	Users selected passwords		
	Forgotten password retrieval	Y	Y
5.2	User Management		
	Add user profile	N	N
	Edit user profile	N	N
	Delete user profile	N	N
5.3	Limit Access at Different Levels		
	File/object level	N	N
	Collection level	N	N
5.4	User Roles		
	Allows definition of different user groups	N	N
	Limits access by role	N	N
	Allows collection to be customized for each role	N	N
5.5	Access Management		
	IP source address	N	N
	Proxy filtering	N	N
5.6	Credential-Based Access		
	Encryption	N	N
	Digital signatures	N	N
<b>6.</b>	<b>REPORTS AND INQUIRY CAPABILITIES</b>		
6.1	System Generated Usage Statistics		
	Feature available	Y	N
6.2	Usage Reports		
	Report timeline specification	Y	N
	Report fields customization	Y	N
	Report templates	Y	N
<b>7.</b>	<b>PRESERVATION</b>		
7.1	Persistent Document Identification		
	System assigned identification	Y	Y
	CNRI Handel	N	N
7.2	Quality Control		
	Feature violable	N	N
7.3	Prescribed Digital Preservation Strategy		
	Feature available	N	N
<b>8.</b>	<b>INTEROPERABILITY</b>		
	OAI-PMH	Y	Y
8.1	Z39.50 Protocol Complaint		
	Feature available	Y	N
	Rehears protocols	N	N
	Dienst	N	N
	SDLIP	N	N

Table 1d: Comparison of GreenStone and DSpace

No.	Checklist Categories	GreenStone	DSpace
<b>9.</b>	<b>USER INTERFACE</b>		
9.1	Modify Interface "Look And Feel" Feature available	Y	Y
9.2	Apply A Customized Header/Footer To Statistic Or Dynamic Pages Feature available	Y	N
9.3	Supports Multi-Lingual Interface Feature available	Y	Y
<b>10.</b>	<b>STANDARDS COMPLAISANCE</b>		
10.1	Structured Document Formats e. g XML, SGML Feature available	Y	N
10.2	Metadata Formats E. G Dublin Core Feature available	Y	Y
10.3	Text Formats E. G Unicode Feature available	Y	N
10.4	Image Formats e. g. TIFF Feature available	Y	N
<b>11.</b>	<b>AUTOMATIC TOOLS</b>		
11.1	Metadata Entry System Feature available	Y	N
	Generation of	Y	N
11.2	Search Indexes HTMML	Y	N
	Reports	Y	N
<b>12.</b>	<b>SYSTEM SUPPORT AND MAINTENANCE</b>		
	Documentation / manuals	Y	N
12.1	Mailing Lists / Discussion Forums Feature available	Y	Y
12.2	Bug Track / Feature Request System Feature available	Y	Y
12.3	Help Desk Support Feature available	Y	Y

## REFERENCES

- Breeding, M. 2002. An update on open source ILS. *Information Today*, Vol. 19no. 9: 42-43.
- Feather, J. and Sturges, P., Eds. 2003. *International Encyclopedia of Information and Library Science*. (2<sup>nd</sup> ed.). London: Routledge.
- Hull, Barbara. 2001. Can libraries help to overcome the social barriers to access? *New Library world*, Vol. 102no. 10: 382-388.
- Hussain, A. 2006. *Perceptions of LIS professionals about Digital Libraries*. Unpublished master's thesis, the Department of Library Science, University of the Punjab, Lahore.
- Hussain, Ashfaq, Khalid Mahmood, and Farzana Shafique. 2008. Perceptions of LIS professionals about digital libraries in Pakistan, *The Electronic Journal of Academic and Special Librarianship*, Vol. 9 no. 1. Accessed March 23, 2012, [http://southernlibrarianship.icaap.org/content/v09n01/hussain\\_a01.html](http://southernlibrarianship.icaap.org/content/v09n01/hussain_a01.html)
- Prytherch, R. J. 2005. *Harrod's librarians' glossary and reference book*. (10<sup>th</sup> ed.). Hants: Ashgate.



