Chapter 1

Introduction

1.1 Introduction

Information technology has played an important role in library and information science. Due to the developments in information technology, now, it is possible for libraries to provide several new services to the library users along with traditional services. Libraries are now able to provide information in print form as well as in digital form.

During 1980s libraries started automating their bibliographic databases and during 1990s digital library projects were initiated. As on today, lot of developments have taken place in digitizing print media. At national and international level several big funding projects have been initiated to digitize valuable material available within the libraries for the preservation as well as for providing wider access to the collections through latest technologies.

Digital libraries have been making their roots in the library profession as a separate discipline and many conferences, workshops and seminars are taking place in the area of digital libraries. These conferences are covering different topics under digital libraries such as collection development and organization, user studies, digital library architecture, usability studies, search and retrieval, digital library software or providing value added services to end users.

Digital libraries are becoming popular and are becoming one of the important activity of any organization. The rapid growth in computing networks, databases and public awareness have contributed to a hot topic of today such as digital libraries, digital archives, institutional repositories or digital repositories.
While using the technology, it has also placed couple of challenges in front of the librarians such as which hardware/software to be used for organizing scanned digital collections or born digital collections, how to maintain these collections over a long term, what are the other aspects which needs to be considered before bringing digital collections on the Internet/Intranet of Library Web Sites.

1.2 What is a Digital Library?

According to Deegan, M. and Tanner, S.(2002)[1], definition of a digital library is

1. A digital library is a managed collection of digital objects.

2. The digital objects are created or collected according to the principles of collection development.

3. The digital objects are made available in a cohesive manner, supported by services necessary to allow users to retrieve and exploit the resources just as in the case of other library materials.

4. The digital objects are treated as long term stable resources and appropriate processes are applied to them to ensure their quality and survivability.

Digital Library Foundation defined digital library(1998)[2] as “Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, 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.” This definition involves three key components, which constitute the theoretical framework of digital libraries such as: people, information resources and technology.

Another workable and widely used definition of digital library provided by Lesk(1997)[3] is “a collection of information which is both digitized and organised”.

A review of the above definitions indicate that digital libraries are organised in a special manner with the help of computers and are made available over a network, with different
procedures such as, to select the material for adding in the collections, to organize collections, to make collections available to end users with value added services and to archive the material for long term access[4].

During 1990s, the research development and practice related to digital libraries took off in a large scale. Researchers focused on digital libraries as contents collected on behalf of user communities, while librarians focused on digital libraries as institutions or services. Many new initiatives came through the research sponsored by the US National Science Foundation (NSF) and UK Joint Information Systems Committee (JISC).

Digital Library Initiative (DLI) projects in the USA and the eLib projects in UK have played a key role in the development of digital libraries. In addition many digital library projects are currently under way in Australia, Asia, Europe, Africa and Latin America. While some of them have their own funding, others are funded under digital library-specific funding initiatives.

Today many other groups as well are involved in the expansion of digital library technologies and techniques such as European Union, Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), The International Federation of Library Associations (IFLA), the American Library Association (ALA), the Coalition for Networked Information (CNI), and the Digital Library Federation (DLF).

During the past decade thousands of digital libraries in a variety of forms were built globally and are functioning operationally, with more to come. Hundreds of research projects were then devoted to many aspects of digital libraries in many countries, and more are reported each year.

A fully developed digital library environment involves the following elements[5]:

1. Initial conversion of content from physical to digital form.

2. The extraction or creation of metadata to assist in object viewing, management, and preservation.

3. Storage of digital content and metadata in an appropriate multimedia repository.

4. Client services for the browser, including repository querying and work flow.
5. Content delivery via file transfer or streaming media.

6. Patron access through a browser or dedicated client.

Digital libraries thus provide new technological platform for implementing functionality of traditional library systems by making them much more powerful. Digital libraries developed today are based on innovative web technologies such as Semantic Web, Ontology Specification, Database Technologies, XML databases, text retrieval in different languages etc.

1.3 Why Digital Libraries?

- The fundamental reason for building digital libraries is that they provide better delivery of information than was possible in the past.

- The digital library brings the library to the user.

- Computer power is used for searching and browsing.

- Information can be shared.

- It is easier to keep information current.

- Information is always available.

Digital Libraries are nothing but extension of traditional libraries. In traditional library activities librarian select, collect, organize, store information and make it available to users. In a similar way digital librarian have to carry out all these jobs. To carry out these tasks there is a need of skilled professionals who can create, organize and manage digital information.

Digital libraries (DL) are digitized versions of conventional media, such as text, images, audio/video, etc. DL hold any information that can be encoded as sequence of bits. Newspapers, scientific journals, books, question papers, thesis/dissertations, bibliographies, images, photos, audio/video resources, preprints/reprints of the faculty members etc. all these documents if converted from analog media to digital media or if created through digital technology fall under the category of Digital Libraries.
All over the world, libraries started realizing the fact that they have to preserve their valuable materials available with them with the help of new preservation techniques. This has helped librarians to scan any analog document (e.g. paper, microfiche, microfilm, negative films etc.) and convert those documents into digital format for long term preservation.

There are two groups which are creating digital libraries. During 1994, libraries started converting their historic collections into digital format and building digital libraries and during 1998, major scientific publications started becoming available online from both commercial and society publishers.

Technology is changing very rapidly hence it is very important for the librarians to keep on converting the existing digitized documents to the latest available formats. Though the print media is a stable media digital media is also supposed to be a parallel stable media and hence it is imperative to use this media for the long term preservation.

Researchers in various fields are interested in digital library technologies because DL provide the possibility of effective distribution of research results to the scientific community[6]. With the advent of the Internet, individuals expectations for access to information have increased considerably.

Though digital documents are created (either born digital or from analog media to digital media by digitization), it is as well required to manage these documents in a proper way, such as it should have proper Information Retrieval Systems or Digital Library Software.

Libraries can collect, organize and disseminate data and information in an effective manner if they have good information management systems support. Information systems require well-structured data and consistently applied vocabularies in order to be truly useful.

In order to access information systems they require elements of readability, browse ability, search ability and interactive assistance. As the size of an information system increases it requires ability to browse and search. This is necessary when the users seeks specific information and when users can articulate their information needs. Interactive assistance is also necessary when an information system becomes very large or complex.

For creating Digital Libraries, librarians started using either commercial software or Open Source Software (OSS). Many libraries realized importance of open source software and they took initiatives in using OSS tools for different applications in libraries.
1.4 Open Source Software

Since 1997, Open Source software (OSS) have taken up a good market in computer industry. Large number of Open Source Software are available on Internet. The world’s largest OSS development web site is SourceForge.net. It provides free hosting to OSS development projects with a centralized resource for managing projects, issues, communications, and code. Currently there are 1,66,993 registered projects and there are 1,771,097 registered users.

There is a general confusion, among the users about open source, freeware, shareware and public domain software licenses. Freeware is a software that is released free of cost in binary format only, usually prohibiting modifications and commercial redistribution on the part of the end user. Shareware is a software that is released free of cost in binary format only, usually allowed on a trial basis regarding time usage or functionality to encourage purchase. Public domain is a software whose copyrights has expired or has been released from copyright obligations by the author(s), rendering it free of restrictions on usage and redistribution.

The open source model on the other hand is a collaborative programming infrastructure that co-opts copyright law by freely releasing source code to the general public for any use, modification, and redistribution without licensing restrictions. The source code refers to instructions written by humans in a computer programming language to be compiled into a binary format that can run on a computer, carrying out the tasks outlined in the source code.

According to Raymond, the definition of Open Source Software (OSS) is Software that is freely redistributable and can readily be evolved and modified to fit changing needs. In the open source movement, openness implies on ability to access and change the source code, at any time, to support a desired capability.

OSS is a term to describe the tradition of open standards, shared source code and collaborative development. OSS programs are available for any user for use. OSS are becoming increasingly popular software development method. It is both a philosophy and a process. As a philosophy it describes the intended use of software and methods for its distribution. In case of proprietary software the software is not free nor the source code of...
the software is available to the end user.

OSS place their source code in public domain, which allows third parties to contribute code and facilitates the creation of a self-organizing networked community of developers. There have been many successful open-source projects e.g. Linux Operating System, Apache Web Server, Perl, Sendmail, Bind, Tcl/tk and Python.

OSS guarantee free access to the programming behind the precompiled binary or source code. All OSS software are copyrighted and distributed with license terms and conditions designed to ensure that the source code is always available. The most popular Open Source license is GPL i.e. GNU Public License. Value of any OSS is measured in terms of its simplicity and connectivity.

Benefits of using OSS:-

1. Software does not depend on specific hardware or operating system platform in order to function.

2. With OSS, people can have any number of copies of programs on their machines, at home or at work.

3. Since source code is available one can customize the software as per the end user’s requirements.

4. It is possible to incorporate the software into the another program to perform new functions.

Drawbacks of using OSS:-

1. Lack of formal support and training that a commercial software package offer. For open source software often support is provided only through mailing lists and discussion forums.

2. Installing and maintaining OSS generally requires a higher level of technological sophistication than that required for commercial software.

3. OSS are also not known for ease of use as the focus is usually on functionality.
4. With no vendor responsible for the software, support for the OSS applications can vary and often depend on the user/developer’s communities commitment to the project.

5. Documentation manuals of OSS are not very simple.

In OSS several people are involved in developing OSS. There are different groups who are contributing to the development of OSS. First group is methodology gurus who spread importance of OSS, then product gurus who make changes in each OSS, then there are, contributors of OSS who are programmers who develop products and release them into the OSS product inventory and finally there is great mass of OSS readers who analyze each OSS, critique on the code of OSS, find its faults and propose changes and enhancements. Thus in this hierarchy the success of OSS depends on the readers who use these OSS[11].

The value of any OSS is measured in terms of its simplicity and connectivity. The simpler and more connective the software, the more it is used.

1.5 State of OSS in Libraries

Daniel Chudnov has done a lot to raise the awareness of OSS in libraries. The OSS4lib site http://www.oss4.lib.org lists all library related OSS applications including applications for document delivery, Z39.50 clients and servers, systems to manage collections, MARC record readers and writers, integrated library system, digital library software etc.

The OSS4lib site states that its mission is to "cultivate the collaborative power of open source software engineering to build better and free systems for use in libraries[12]." The site maintains listing of OSS designed for libraries and tracks news about project updates and related topics.

The open source movement and libraries have a common attribute such as free and open access to ideas and information[13]. Libraries have highly specialized software needs because the library community have developed its own complex standards and protocols to facilitate things like inter library loan, metadata sharing and federated searching. Until recently libraries relied on the commercial solutions for all their requirements due to unavailability of skilled IT staff as well as unavailability of user friendly open source solutions.
Hence open source solutions were not considered in libraries as a scalable or feature-rich solutions to handle most of the library requirements.

Today several companies/organisations all over the world are committed to support and develop solutions based on open source software for libraries as well as in other areas. They offer everything from hosting, installation to support and development services[14].

OSS presents many possibilities for libraries. It presents an opportunity for libraries to take control of library services and collections relying on computer software. The time and effort spent on buying commercial solution can be used to develop learning on how to use a particular OSS application.

The availability of the source code in open source software allow end users to modify and make improvements in any application. OSS tends to have more functions than commercial solutions as they are developed by the users of the software themselves, as compared to commercial software, where a vendors priority is in profit generation that may not be in line with the needs of users[15]. Since source code is available, accessible and modifiable, it is easy to make improvements in the functionality of the software. In addition to it, updates can be obtained at no cost and there are no royalties or license fees. There is less likelihood of being dependent on a single software provider or being trapped into long-term software support contracts, which restricts flexibility in implementation[16].

Many OSS tools are now available for use in libraries and librarians are making use of these tools. For creating digital libraries OSS tools are increasingly considered as an alternative instead of commercial digital library systems due to dissatisfaction with functionality[17]. Open source Digital Library software (OSS-DL), with its free access and good level of functionality, have been used by large community all over the world. Since there are no purchase costs involved for using OSS, libraries can divert such funds to reallocate those funds for training the staff or hiring additional programmers to customize software as per specific requirements.

1.6 Literature Review

In the initial stage of research an extensive literature search was carried out on the following databases on key terms such as Digital Libraries, Evaluation of Digital Libraries, Evalua-
tion of Open Source Digital Library Software and Open Source Digital Library Software for understanding the research covered by earlier researchers. The searches were carried out on Internet as well as on Online bibliographic and full text databases. Contacts with the colleagues, email discussion lists and other appropriate sources were also used to gain the knowledge of actual work done so far in this research area. While reviewing the material, the emphasis was given on providing some insight into the current level of activity in the evaluation of open source digital library software.

### 1.6.1 Sources of Literature

1. **LISA:** Library and Information Science Abstracts is an international abstracting and indexing tool designed for library professionals and other information specialists. LISA currently abstracts over 440 periodicals from more than 68 countries and in more than 20 different languages. The abstracts are covered from 1969 till present. Over 296,127 records are there on CSA website as of March 2007[18].

2. **ACM Digital Library:** The ACM digital library covers major digital library science conference proceeding named as Joint Conference on Digital Libraries since 1997. Hence this database was used in depth to know the latest research carried out in Digital Libraries[19].

3. **Google Scholar[20] as well as General Google:** Google Scholar covers bibliographic as well as full text database of all articles published from all journals and google covers presentations, articles available from individual web sites for downloading.

4. **Science Direct[21]:** Covers major library science journals such as Information Processing & Management, Library and Information Science Research etc.

5. **Springer Link Services[22]:** Covers all articles from the journal International Journal on Digital libraries which is a European Journal.

While undertaking the review of the literature related to the key focus areas of the present study the researcher has concentrated on locating material to provide contextual information and background. Information has also been identified which examines the relationship between theory and practice in the areas of evaluation of open source digital library software.

### 1.7 Review of literature related to present study

#### 1.7.1 Evaluation of Digital Libraries

Research on the evaluation of digital libraries is still in its infancy. Very few studies have been carried out on evaluation of digital libraries.

Marchionini, Plaisant, and Komlodi (1998)[24] explored user needs in a digital library and developed an interface prototype for a digital library. This prototype was based on the principle of users should maximize their interactions with the information resources and minimize their attention to the system.

Saracevic and Covi (2000)[25] pointed out that, the evaluation of digital libraries is a complex and difficult activity. They enumerated challenges faced while evaluating digital libraries and suggested a conceptual framework for evaluation.

Hill [et al.] (2000)[26] developed three different user interfaces for Alexandria Digital Library and tested these user interfaces by different user groups. User feedbacks were also collected through various formal and informal approaches and the results were fed back to implementation of ADL.

Bishop [et al.] (2000)[27] presented the nature and extent of a digital library testbed application, which included extent of use of the digital library compared to other systems, nature of use, viewing behavior, purpose and importance of use and user satisfaction. Data were
collected from potential and actual users through focus groups, interviews, observations, usability testing, user registration and transaction logging and user surveys.

Meyyappan, N. [et al.] (2000)[28] carried out survey of 20 digital libraries around the world and reviewed various features of selected digital libraries and concentrated their study on reviewing the search features, output format and links to other Internet sources provided by these 20 digital libraries. The digital libraries chosen from the study includes ACM Digital Library, Alexandrian Digital Library, British Library, California Digital Library, National Library of Canada etc.

Cherry and Duff (2002)[29] conducted a longitudinal study of a digital library collection of early Canadian materials, focusing on how the digital library was used and the level of user satisfaction with different features of the digital library, such as response time, browsing capabilities, comprehensiveness of the collection, print function, search capabilities, and display of document pages.

Bertot and McClure (2003)[30] proposed a framework to relate traditional evaluation components and terminology of library services/resource assessment to the networked environment.

Saracevic (2004)[31] provided an overview of works on digital library evaluation. He tried to analyze and then isolate 80 evaluation studies on evaluation of digital libraries. He also found that evaluation of digital libraries is not yet widely practiced.

Gonzalo, J. (2006)[32] proposed a new logging schema for logging and sharing a wide array of data about users, systems and their interactions. This service will allow diverse digital libraries to store their log data in a common repository using a common format.

Xie(2006)[33] carried out a survey and tried to find out how Usability is important in any digital library. It was found that out of the surveyed members, more than half of the participants (52%) discussed usability in general as the key evaluation criterion. Participants cared the most about interface usability. 54% of the participants emphasized the importance of interface design. Usability evaluation is important because if the users can not access DL effectively or finds it cumbersome, the information is not really accessible to them.
1.7.2 Evaluation of Open Source Digital Libraries (OSS-DL)

Trusted digital repositories: attributes and responsibilities, an RLG-OCLC report (2002)[34], identified the repository evaluation criteria that are essential from the digital preservation point of view.

Han, Y. (2004)[35], evaluated DSpace, Fedora and Greenstone software while identifying and defining few evaluation criteria. The authors examined a dozens of Open Source Software and Commercial Solutions in the content management area to match their requirements. After the analysis DSpace scored the highest rank among the selected three candidate software.


Technical Evaluation of Open Source Repository Solutions on behalf of CPIT ver. 1.3 (2006)[37], report was prepared on the basis of Open Access Repositories in New Zealand. The objectives of the study was to gain understanding of the design, architecture and implementation details of DSpace, Fedora and EPrints. The group tried to evaluate these three candidates on the basis of few evaluation criteria. The report found Fedora repository as a strong choice for building digital libraries as Fedora can handle millions of documents easily as per the stated report.

Goh, D. [et al.] (2006)[38], tried to develop a checklist for evaluating open source digital library software. The checklist was used to evaluate Fedora, Greenstone, EPrints and CDSware software. Greenstone performed best against the evaluation criteria applied on each selected software. It was also found that EPrints performed worst due to its poor support for certain features.

1.8 Need for the present study

Digital libraries give us opportunities we never had with traditional libraries or even with the web[39]. The advent of open source and proprietary software has enabled universities to develop web based digital repositories.
CHAPTER 1. INTRODUCTION

In the fast changing networked environment librarians have more responsibilities. Librarians have to tackle print media as well as digital media. Handling digital media demands more up to date knowledge of the subject as well as technology.

Though librarians try to digitize important collections available with them, the further biggest challenge is how to organize the born digital collections as well as digitized collections which are spread around in all organizations.

During 1998, open source trend came into libraries and libraries started making use of these OSS tools for different tasks carried out in libraries. OSS tools are helping libraries to overcome problem of high budget allocations for buying commercial solutions. Lot of research is going on in using OSS based applications which are useful in libraries. In recent years, there have been a steep rise in the number of applications that are available for the libraries to automate library catalogues, MARC editors, digitization work flow management software, web page designing software etc.

Paul Jones (2001) emphasized on a few key components of most open source projects such as open contribution, user centered feedback & distribution, and building of user communities. These trends have been identified as significant reasons to shift to open source initiatives.

A couples of digital library software are available for the librarians to build, maintain, organize and make their collections available over Internet for end users. These OSS-DL are available on Internet under Open Source License for helping librarians to manage digital collections and make them available on Internet/Intranet. To name a few Greenstone, DSpace, Fedora, EPrints, DoKS, MyCoRe, ARNO, CDSware are available on Internet under Open Source License form. OSS4Lib site maintains major open source software used in libraries for all applications that are carried out in libraries.

For years institutions have relied upon commercial vendors to create and maintain the software that they use – from small digital library projects to large enterprise financial systems but now OSS provide a viable, and preferable alternative to vendor provided software.

Many of the OSS-DL applications such as Fedora software, DSpace software and Greenstone software have come from grant funded projects. The academic community all over the world looked towards these open source software as an alternative to commercial solu-
All these software have their own advantages and disadvantages. Since there are couples of OSS available for building digital libraries it is difficult for the libraries to decide which software is most suitable for them in a given set of environment.

As the use of open source software (OSS) has taken off over the past decade, there has been increasing interest in the potential of open source to address longstanding concerns in the library and information science regarding the cost and performance of commercial software products. A common view is that existing proprietary options do not have the features required or allow for cost-effective customization[42]. There is an active open source community which has been constantly trying to provide a better solutions in all areas.

So far very few studies have been carried out in evaluating OSS-DL (as mentioned earlier) and these studies have not covered all software that are available under Open Source license terms and conditions as well as have not covered extensive evaluation criteria. Hence there was an urgent need to define extensive evaluation criteria and carry out in depth evaluation of all OSS-DL available free on Internet and preferably under open source terms and conditions and find different features supported by these software.

1.9 Objectives of the present study

The objective of the present work is to evaluate major Open Source Digital Library Software (OSS-DL) available free and preferably under Open Source License terms and conditions. The main objective or goal of the present work can be further divided into the following sub-goals or secondary objectives:

1. To identify a list of major Open Source Digital Library Software available for creating digital libraries on Internet.

2. To identify the evaluation studies carried out so far on Evaluation of Open Source Digital Library Software (OSS-DL).

3. To design a set of evaluation criteria for evaluation of selected Open Source Digital Library Software (OSS-DL).
4. To install selected candidate software for carrying out detailed evaluation study on a testbed environment.

5. To evaluate each selected software against the set of designed criteria such as content acquisition, content management, metadata submission and support, classification, information search/retrieval, access control, privacy and management, authentication and authorization, interoperability, ease of deployment of each software, security, scalability, backup/restore facility, user friendly interface, usability, transaction log analysis, copyright issues, load balancing, personalization, visualization, digital preservation etc.

6. To find out at what level digital preservation is supported in each software.

7. To enlist each criterion and against that evaluate each OSS-DL capabilities, along with its advantages and disadvantages which will help librarian to decide which software to be used for a given set of environment with a given set of features.

1.10 Scope of present work

All over the world libraries have started creating digital libraries of important documents available with them. For managing digital documents, libraries have either commercial information management systems or open source software.

Use of OSS for libraries has grown in the last decade. There are large numbers of OSS software available for carrying out different tasks in the libraries[43]. Since the source code is available it is easier to customize the software as per the end users requirements. Considering the fact that OSS are free for use, it is of great help to developing country libraries. The implementation of OSS in libraries helps to improve library services and collections.

There are couples of OSS-DL available today. Each software satisfies a particular need. Since each of the software has different configuration, have used different standards, satisfies different needs for creating DL, it was felt necessary to study all the major currently available Open Source Software for Building Digital Libraries with various factors which will help libraries especially from developing countries what features are supported by
each software, what are the installation steps, which software has sustainability to stay in future, what are the advantages/disadvantages of using each software, how many users are using the software etc.

The present study covers only those software which are available for creating digital libraries. This study does not cover any other library related open source software (for e. g. Library automation software such as Koha, NewGenLib, CDS-ISIS etc.) though they are available under OSS license.

The scope of the study is further extended to evaluating current stable versions of all software that are available on Internet during the course of work. The scope is also further restricted to successful installation of each software which are taken for the present study.

The primary objective of this study is not to present which is the best software as compared to the other one. It does not attempt to rank any software and indicate the best software. In fact, such a comparison many not lead us anywhere because each software has its own objectives behind developing it. However, what features are supported by each software may help us to understand range of facilities available in each software and which software best suits for a particular type of library will certainly be helpful.

1.11 Hypotheses

1. Many of the Open Source Digital Library software (OSS-DL) are not in conformity with standards.

2. Digital preservation is not really yet been addressed by many of the OSS-DL software.

3. Interoperability, scalability are highly questionable aspects of many OSS-DL software.

1.12 Research Methodology

The present work deals with the evaluation of Open Source Digital Library Software (OSS-DL). Over the last decade, there have been significant efforts to develop digital libraries.
Since the digital library community is largely busy in building digital libraries very little attention is paid on evaluating the contents added in digital libraries as well as evaluation of technology which is used for building digital collections as well as usability evaluation.

Considering this gap the present study proposes to define extensive evaluation criteria for evaluating OSS-DL and proposes to evaluate OSS-DL against set of defined evaluation criteria by installing each software on a test bed environment and actually evaluating each criterion.

In order to achieve the specified objectives and to test the hypotheses, extensive Evaluation Criteria were defined for evaluating each selected OSS-DL candidate.

Evaluation criteria were finalised after having review of earlier studies, few presentations and having discussions with digital library experts who are working in the digital libraries field.

The selection of OSS-DL software were made under the basic criteria of having the software available through Open Source License terms and conditions as well as available free on Internet. A list of such software is available on ROAR[44] as well as OSS4lib site to know the number of OSS-DL used for creating open access digital repositories/institutional repositories. An initial list was prepared which covered the following major OSS-DL. The following list also provide total number of institutions using these software.

- ARNO – 5
- CDSWare – 9
- DoKS – 5
- DSpace – 266
- Fedora –9
- Fez/Fedora –3
- GNU EPrints – 242
- MyCoRe – 4
• SciX–2 (Browsed as on 16th January 2007)

Since the study is especially confined to defining evaluation criteria for evaluating Open Source Digital Library Software and executing these criteria on the selected OSS-DL candidate, evaluation of only Open Source Software was carried out as the purpose of the study. Each software was thoroughly checked whether it has been made available under Open Source License or is available free on Internet and then it was selected for further evaluation.

To create list of evaluation criteria it was necessary to understand the digital library concept hence State-of-the-art and trend in Digital Libraries was carried out.

Since the present study falls under Evaluation part of the Research Methodology, extensive evaluation criteria were developed for assessing each OSS-DL software.

On the basis of the available software each software was evaluated against the defined broad and narrow evaluation criteria. Broad criteria were broken into sub-criteria with each sub-criteria having an importance rating.

A detailed evaluation criteria list was prepared covering the following broad evaluation criteria and further subdividing each of these broad criteria into several sub criteria such as:-

content acquisition, content management, metadata submission and support, classification, information search/retrieval, access control privacy and management, authentication and authorization, interoperability, ease of deployment of each software, General features related to technical aspect, Architecture of the software, Backup/Restore facility, User Friendly Interface, Usability, Transaction Log Analysis, Copyright/Policy Issues, Advanced Features, Digital Preservation.

The software were shortlisted on the basis of successful installation of each software. Installation of the software was essential factor as otherwise it would have been difficult to carry out further evaluation if the installation of each software was not successful. The repository systems that were not short-listed were not evaluated further due to the fact that the installation of those software were not possible due to different reasons.

The following candidate software were initially short listed for evaluation.
CHAPTER 1. INTRODUCTION

1. ARNO
2. CDS-Invenio
3. DSpace
4. DoKS
5. EPrints
6. Fedora
7. Greenstone
8. MyCoRe
9. SOPS

ARNO software was not considered for the further evaluation in the present study due to the fact that though the software is available under open source license the back end database which is used in ARNO is a commercial database i.e. ORACLE. Hence ARNO was not considered for further evaluation.

Further, the present study added Greenstone Digital Library Software for the present study. Greenstone is meant for building digital libraries. There is a large community which is using Greenstone and Greenstone allows to organise digital files and is meant for creating digital libraries. Hence it was necessary to add Greenstone in the present study.

The following list shows the total number of software selected for the present study:

1. CDS-Invenio
2. DSpace
3. DoKS
4. EPrints
5. Fedora
6. Greenstone
1.12.1 Installation of the Selected Software

A test bed environment was created to install all the selected candidate software. The short-listed candidate software were installed on a hardware with 160 GB hard disk space and 1GB RAM with installation of Linux Operating System having Fedora Core 7 version of Linux.

It was a challenging task to install all the above software on a single server. One of the challenging task was to configure different software on common web server such as EPrints/CDS-Invenio/Greenstone/SOPS run on Apache Web Server and DSpace/DoKS/Fedora/MyCoRe require Tomcat Web Server. It was then necessary to use different ports for installation of each software. The databases used at the back end with these software were also different such as DSpace uses PostGreSQL, DoKS/EPrints/CDS-Invenio/Fedora/uses MySQL, MyCoRe uses hsql database & SOPS uses WODA database. For each of these software it was necessary to create database connection with login name and password.

Installation of MyCoRe was a challenging task as most of the documentation manual of MyCoRe was available only in German and there are no other detailed installation instructions available in English. Hence it took time to install MyCoRe successfully.

Initially 80 GB hard disk was used for installation of each software and successful functioning of each software but it was found that to upload audio/video files in each software there was a need to have higher disk space hence a new hard disk was used with 160 GB disk space where installation of all software was reinitiated.

Where ever possible latest stable versions of each software was installed for evaluation.

The following list provides installed versions of each software on a test bed environment.

1. CDS-Inveion version 0.92.1
2. DSpace version 1.4.2
3. DoKS version 1.2.4
4. EPrints version 3.0.3
5. Fedora version 2.2.1
6. Greenstone version 2.75
7. MyCoRe version 1.3.0
8. SOPS version 1.0

A small collection of each document type such as text, audio, video, image, etc. was added in each software to understand the basic functionality of each software and to find which document types are supported in each software. A small number of user community was also created in each software to know the user administration, authorization and authentication process supported in each software as well as to understand work flow process available in each software. After carrying out these tasks each software was evaluated against the set of defined evaluation criteria.

The objective of the present study as mentioned above is to understand how the functionality of each software as well as to know the different features supported by each software. The ultimate goal of any digital library is universal access, which is consistent with the traditional library goal of providing access to information to all. To accomplish this goal, digital libraries need to solve the problem of integrating distributed heterogeneous information and information sources have to support different standards and have to provide multilingual based information resources with search and retrieval facility.

1.13 Conclusion

The digital media as on today is considered as a stable media along with the print media. Libraries are creating DL of their valuable collections. Many live repositories are available today and libraries are using OSS-DL tools effectively to build institutional repositories/digital libraries/digital repositories. In developing countries where funds are less OSS have been found very useful. OSS has much potential in library science and there are number of projects that demonstrate its viability. It gives library staff an option to be actively
involved in OSS development projects and this involvement takes many forms, such as reporting bugs, suggesting enhancements, usability studies or testing new versions of OSS.

Considering these opportunities it was felt to evaluate available OSS-DL in detail and identify features supported by each software and evaluate them against a set of defined evaluation criteria. This study provides insightful information to the user community about which software best suits in a given set of environment. This analysis also finds out pros and cons of each OSS-DL which will give guidelines to the libraries for deciding what type of OSS-DL can be used for a specific type of digital collection in a given environment setting.

1.14 Chapterisation

The present study is divided into the following chapters:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Overview of Digital Libraries</td>
</tr>
<tr>
<td>3</td>
<td>Evaluation Criteria Defined for Evaluating Open Source Digital Library Software (OSS-DL)</td>
</tr>
<tr>
<td>4</td>
<td>CDS-Invenio Digital Library Software Evaluation</td>
</tr>
<tr>
<td>5</td>
<td>DSpace Digital Library Software Evaluation</td>
</tr>
<tr>
<td>6</td>
<td>DoKS Digital Library Software Evaluation</td>
</tr>
<tr>
<td>7</td>
<td>EPrints Digital Library Software Evaluation</td>
</tr>
<tr>
<td>8</td>
<td>Fedora Digital Library Software Evaluation</td>
</tr>
<tr>
<td>9</td>
<td>Greenstone Digital Library Software Evaluation</td>
</tr>
<tr>
<td>10</td>
<td>MyCoRe Digital Library Software Evaluation</td>
</tr>
<tr>
<td>11</td>
<td>SOPS (SciX Open Publishing Services) Digital Library Software Evaluation</td>
</tr>
<tr>
<td>12</td>
<td>Comparative Analysis of Software Evaluation</td>
</tr>
<tr>
<td>13</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Bibliography


[22] http://www.springerlink.com


[41] Ibid.


[44] [http://www.roar.org](http://www.roar.org) (Browsed on 16th January 2006).