Proceedings of the National Seminar on "Preservation and Conservation of Information Resources in Knowledge Society: Issues, Challenges and Trends" held on March 3-4, 2009, at Manipur University, Canchipur, Imphal, Pages 238-258. Edited by Th. Madhuri Devi and Ch. Ibohal Singh

OPEN SOURCE DIGITAL LIBRARY SOFTWARE: A LITERATURE REVIEW

Hanadashisha Warr
Dr. P. Hangsing*

What is a Digital Library?

All conventional libraries basic functions focus on collection, organization and dissemination of information resources. Traditionally a "library is a place in which books, manuscripts, musical scores, or other literary and artistic materials are kept for use but not for sale". In effect, it is an institution oriented towards collections and custody, where people may make use of the facilities. Whereas a digital library is an assemblage of digital computing, storage and communications machinery together with the content and software needed to reproduce, emulate and extend the services provided by conventional libraries. In other words, a digital library is a computer–based system for acquiring, storing, organizing, searching and distributing digital materials for end user access. It is not just a collection of material in electronic form; it includes a browser interface and perhaps a virtual space and society. It requires less space and the data can be made available through communication networks to anyone anywhere, while facilitating searches with speed. The digital is not a single entity and as such is linked to the resources of many such collections.

The first use of the term *digital library* in print may have been in a 1988 report to the Corporation for National Research Initiatives. The term *digital libraries* were first popularized by the NSF/DARPA/NASA (National Science Foundation/Defense Advanced Research Projects Agency/ National Aeronautics and Space Administration) Digital Libraries Initiative in 1994.

Phrases like "virtual library," "electronic library," "library without walls" and, most recently, "digital library," can all be used interchangeably.

Definitions and Characteristics Digital Library

The definition of digital libraries varies depending upon various factors and perspectives. From the infrastructural and technical perspective, digital library is defined as a library that "basically store materials in electronic format and manipulate large collections of those materials effectively. Research into digital libraries is research into network information systems, concentrating on how to develop the necessary infrastructure to effectively mass-manipulate the information on the Net" (NSF, 1999). Seadle (2006) refused to accept it as a total technological change that would replace the traditional library in the statement "Digital libraries are not replacements for traditional libraries. They are rather the future of traditional libraries, much as medieval manuscript libraries simply became a specialized and much revered part of the larger print-based that we have today". This definition asserts that traditional libraries with printed collection will remain the core of a library collection in

* Department of Library and Information Science, North-Eastern Hill University, Shillong - 22

future and digital collection will be a section of the total collection. Though prediction of the future of libraries may be laden with imperfection, in the light of the drastic technological changes and the enormity of information in digital form the idea of printed documents not being replaced by digital collection is hard to swallow. Seadle's statement also assumes that digital collection is a single entity that would be surrogates or replacements of printed document but "the digital library is not a single entity; the digital library requires technology to link the resources of many services that are transparent to the end users; Universal access to digital libraries and information services is a goal; Digital library collections are not limited to document surrogates; they extend to digital artifacts that cannot be represented or distributed in printed formats" (ARL,1995).

Leiner (1998) deemed Digital library as "the collection of services and the collection of information objects that support users in dealing with information objects available directly or indirectly via electronic/digital means. Similarly, Arms (2000) also felt that digital library is a "A managed collection of information with associated services, where the information is stored in digital formats and accessible over a network". Leiner and Arms definition emphasized on access to services and on content, which is sometimes called collections or documents or information or information objects. Casting the scope wider Waters (1998) defined digital libraries as "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 overtime of collections of digital works".

Some of the features pointed out in the definitions of digital library may be listed as follows:

1. A library that served a defined community or set of communities
2. A conglomerate of multiple entities
3. Library that incorporate learning and access
4. Library that provide fast and efficient access, with multiple access modes
5. A library with a collection which
    - Are large and persist over time
    - Are well organized and managed
    - Contain many formats
    - Contain objects which may be otherwise unobtainable

In defining digital library, various aspects of its characteristics have been put to fore. However, these characteristics are further extended on the basis of various authors and it may be summarized as follows:

- Digital libraries will also include digital materials that exist outside the physical and administrative bounds of any one digital library
- Digital libraries will serve particular communities or constituencies, as traditional libraries do now, though those communities may be widely dispersed throughout the network.
- Digital libraries will require both the skills of librarians and well as those of computer scientists to be viable.

For librarians, these characteristics, are the most logical because it expands and extends the traditional library, preserves the valuable work that they do, while integrating new technologies, new processes, and new media. (Arms,1995; Graham, 1995; Chepesuik, 1997; Lynch and Garcia-Molina, 1995).

Advantages of Digital Library:

Traditional libraries are limited by storage space; digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. As such, the cost of maintaining a digital library is much lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries do away with these fees.

- **No physical boundary**. The user of a digital library need not to go to the library physically; people from all over the world can gain access to the same information, as long as an Internet connection is available.
- **Round the clock availability**. A major advantage of digital libraries is that people can gain access to the information at any time, night or day.
- **Multiple accesses**. The same resources can be used at the same time by a number of users.
- **Structured approach**. Digital libraries provide access to much richer content in a more structured manner, i.e. we can easily move from the catalog to the particular book then to a particular chapter and so on.
- **Information retrieval**. The user is able to use any search term (word, phrase, title, name, subject) to search the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to its resources.
- **Preservation and conservation**. Another important issue is preservation - keeping digital information available in perpetuity. In the preservation of digital materials, the real issue is technical obsolescence. Technical obsolescence in the digital age is like the deterioration of paper in the paper age. Libraries in the pre-digital era had to worry about climate control and the de-acidification of books, but the preservation of digital information will mean constantly coming up with new technical solutions.
- **Space**. Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them. When a library has no space for extension digitization is the only solution.
- **Networking**. A particular digital library can provide a link to any other resources of other digital libraries very easily; thus a seamlessly integrated resource sharing can be achieved.
- **Cost**. In theory, the cost of maintaining a digital library is lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Although digital libraries do away with these fees, it has since been found that digital libraries can be no less expensive in their own way to operate. Digital libraries can and do incur large costs for the conversion of print materials into digital format, for the technical skills of staff to maintain them, and for the costs of maintaining online access (i.e. servers, bandwidth costs, etc.). Also, the information in a digital library must often be "migrated" every few years to the latest digital media. This process can incur very large costs in hardware and skilled personnel.

DIGITAL LIBRARY SOFTWARE

For building a digital library one needs digital library software. Digital library software is a complex computer system which handles all the processes expected from a digital library.

TYPES OF DIGITAL LIBRARY SOFTWARE

(a) Proprietary

1. Distribution of Software/License is prohibited
2. Usually available at a price

(b) Open Source

The Open Source Definition is referred by the *Open Source Initiative (OSI)* to determine whether or not a software license can be considered open source.

According to the OSI, Open Source Software must follow the following criteria:

1. The ability to distribute the software freely.
2. The availability of the source code.
3. The right to create derived works through modification.
4. The integrity of the author's source code must be preserved, making the source of changes clear to the community.
5. No discrimination against persons or groups both for providing contributions and for using the software.
6. No restriction on the purpose of usage of the software, providing no discrimination against fields of endeavour.
7. The rights attached to the software apply to all recipients of its (re)distribution.
8. The license must not be specific to a product, but apply to all sub-parts within the licensed product.

Many commercial digital library software packages are available today. But they are too costly to be afforded by average libraries. These proprietary software packages also require high installation costs followed by annual maintenance costs (AMC) and/or updating costs. A remedy for this situation is the adoption of open source software for digitization purposes. With the recent availability of several open source digital library software packages, the creation and sharing of information through digital library collections has now become an attractive and feasible proposition for library and information professionals around the world.

SELECTION AND REVIEW METHOD

Through reviews of Digital Libraries literature four Open Source Digital Library Software were selected on the basis of popularity and number of installed bases viz. DSPACE, GREENSTONE, EPRINTS and FEDORA. The features that are common to these four software are selected and a comparison of the selected software based on these features is made with the objective of identifying and analyzing the strengths and weaknesses of the software.

DSPACE

- DSPACE is a digital repository system, designed to capture, store, index, preserve and redistribute the intellectual output of an organization in digital formats.
- Developed jointly by Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP).
- DSPACE, originally released in November 2002, is built in Java.
- DSpace being an as open source software can be customized and extended. It runs on any UNIX-type OS, such as Linux, HP/UX, or Solaris.
- DSpace accepts all manner of digital formats. Some examples of items that DSpace can accommodate are: Documents, such as articles, preprints, working papers, technical reports, conference papers, Books ,Theses, Dissertations, Data sets, Computer programs Visualizations, simulations, and other models Multimedia publications, Administrative records, Published books, Overlay journals, Bibliographic datasets, Images Audio files, Video files Reformatted, digital library collections, Learning objects Web pages.

GREENSTONE

- Greenstone is developed by the New Zealand Digital Library Project in the Computer Science Department at the University of Waikato, and distributed in cooperation with UNESCO and the Human Info NGO.
- Greenstone is a suite of multilingual software for building and distributing digital library collections, including organizing information and publishing it on the internet or on a CD-ROM.
- Its aim is to empower users, particularly in universities, libraries and other public service institutions to build large distributed digital library collections.
- Greenstone was originally released in 2000.
- Greenstone is mainly programmed in C++ and Perl and runs on all versions of Windows, and Unix/Linux, and Mac OS-X.
- It is fully documented in English, French, Spanish and Russian.

EPRINTS

- EPrints was developed at the University of Southampton, U.K. with the first version of the software publicly released in late 2000. The objective behind the creation of EPrints was to facilitate open access to research and scholarly literature .
- In other words, EPrint is a digital archive of the research output created by the scientists, faculty, research staff and students of an institution and accessible over the internet to end-users both within and outside the institution. As a facility, it consists of hardware, software and procedures to capture, organize, archive, disseminate and manage digital research resources of an institution. It also serves as an archive for other electronic documents such as images and audio.
- Eprints is programmed on Perl and runs on any number of Linux distributions, and other UNIX-like systems including OS-X. It can also run on Windows Vista and XP.

FEDORA

- The Fedora project was funded by the Andrew W. Mellon Foundation to build a digital object repository management system based on the research paper, “Flexible Extensible Digital Object and Repository Architecture (FEDORA)”.
- Fedora is developed jointly by the University of Virginia and Cornell University.
- Fedora supports digital asset management, institutional repositories, digital archives, content management systems, scholarly publishing enterprises and digital libraries. The system, designed to be a foundation upon which interoperable web-based digital libraries, institutional repositories and other information management systems can be built, demonstrates how distributed digital library architecture can be deployed using web-based technologies such as XML and Web Services.
- Fedora was released in May 2003.
- Fedora, designed in Java, can be configured using other databases such as Oracle. It runs on Linux operating system.

SELECTED FEATURES

Based on literature reviews common features of the above mentioned software are selected. These features include:

- *content management*
- *user interface*
- *user administration*
- *system administration*
- *interoperable*
- *compliance with standards*
- *metadata*
- *system support and help*
- *types of contents*
- *total no. of installations*
- *types of organizations*
- *retrieval features (browsing & searching options).*

COMPARATIVE ANALYSIS OF THE FOUR SELECTED SOFTWARE

The following table shows the comparison of the four selected software based on the selected features.

A. CONTENT MANAGEMENT - Ease with which content is :-	**DSpace**	**Greenstone**	**Eprints**	**Fedora**
- created	√	Does not	Self-archiving	√
- submitted	√	have an in-	√	×
- reviewed	√	built work	√	×
- approved	√	flow process	√	users cannot modify the
- organised	√		√	submitted contents

B. USER INTERFACE				
- flexibility in customizing the interface	√	√	√	API
- multilingual access	√	√	√	√
C. USER ADMINISTRATION				
- password authentication	√	√	√	Does not support much. Access based on machine 's IP addresses
- usage patterns	√	√	×	√
D.SYSTEM ADMINISTRATION				
- content acquisition	√	√	√	limited
- harvesting	√	√	√	support
- metadata generation	√	√	√	
- preservation standards	MD5	Multiple plug-ins	Create its own	Handle System - CNRI
- persistent document identification	Handle System - CNRI	Non standard PURL	Persistent identifier	Unique PID
E. INTEROPERABLE				
Protocols followed – - Z39.50	×	compliant	×	×
- OAI-PMH	√	√	√	√
F.COMPLIANCE WITH STANDARDS FOR PRESENTATION OF INFORMATION	**DSpace**	**Greenstone**	**Eprints**	**Fedora**
- structured document format				
HTML	√	√	√	XML files
SGML	√		Specialized doc format- TeX or LaTeX	in
XML	√	√		FOXML
PDF	√	√	√	format
- image format				
GIF	√	√	√	√
TIFF	√	√	√	√
JPEC	√	√	√	√
- text format				
ASCII	√		√	
UNICODE	√	√	√	√
RTF	√	√	√	

- *audio & video format*				
MP3		√		√
MPEG	√	√	√	√
G. METADATA				
- Dublin core	√	√	√	√
- Marc 21				√ + METS
				+ SOAP
H. SYSTEM SUPPORT & HELP				
- documentation	√	√	√	√
- manuals	√	√	√	√
- mailing lists	√	√	√	√
- discussion forums	√	√	√	
- bug tracking	√	√	√	√
- help desk support	√	√	×	√
I. TYPES OF CONTENTS	Theses, Dissertations lectures, speeches, conference, seminar proceedings, video, government publications, etc.	Newspaper articles to technical documents, educational journals to history, visual arts to videos, pop music collection to ethnic folk songs.	Theses, Dissertations, journal articles, project reports multimedia artifacts, teaching materials, digitized records.	complex, compound, dynamic content - audio, video, image, text, numerical data, dataset, maps, geospatial data, e-books etc.
J. RETRIEVAL FEATURES	**DSpace**	**Greenstone**	**Eprints**	**Fedora**
Browsing				
- by author	√	√	√	
- by title	√	√	√	√
- by date/by year	by date	by date	by year	by date
- subject		√		by object id
-keywords		√	√	
- communities/collection	√			
Searching				
- full text search		√	√	×
- Boolean	√	√	metadata	√
- truncation/wildcard	√	√	search	√
- phrase	√			
- proximity	√	√		

Table 1: Comparison of the four selected software based on the selected features

TOTAL NO. OF INSTALLATION AND TYPES OF ORGANISATION

Based on an Online survey (**Sanjo Jose, 2007**)

Among the 47 software installations included in the survey, DSpace with 20 installations seems to be the most popular choice among the digital library software packages. EPrints is at the second place with 18 installations. Greenstone has 7 installations and Fedora has got 2 installations.

Country	DSpace	EPrints	Fedora	Greenstone	Total
India	13	7	-	4	19
USA	4	2	1	2	9
UK	1	5	-	-	6
Germany	1	1	1	-	3
Argentina	-	-	-	1	1
Denmark	-	1	-	-	1
France	-	1	-	-	1
Netherlands	1	-	-	-	1
Spain	-	1	-	-	1
Total	20	18	2	7	47

Table 2: Country-wise Distribution according to the online survey (Jose, 2007)

From table 2, it is evident that DSpace is the most widely used digital library software in India, with 13 installations. EPrints and Greenstone are also used in India. But digital libraries based on Fedora could not be found. US institutions are using DSpace more, may be because of its American origin. Similarly, UK based institutions use EPrints more than any other software. Its cause may be attributed to membership of these institutions in various JISC funded consortia where they use Eprints for their repositories.

Among the 44 digital libraries identified in the survey, 31 are maintained by educational institutions like universities and colleges. Research institutions are at the second place with 10 digital libraries. Educational institutions seem to prefer DSpace and EPrints, may be due to its academic community specific architecture. Both are specially designed to preserve the scholarly
output of organizations. The only cultural organization responded to survey uses Greenstone. Category-wise distribution of the software packages is given in table 4.

Category	DSpace	EPrints	Fedora	Greenstone	Total
Educational institutions	13	11	2	5	31
Research institutions	6	3	-	1	10
Consortia	-	2	-	-	2
Cultural organizations	-	-	-	1	1
	19	16	2	7	44

Table 4: Category-wise distribution of software packages according to the online survey

(Jose, 2007)

DSpace

Content Management: DSpace allows the creation and granting privileges to E-people with the tasks of submitting, reviewing, approving the submitted documents in a digital repository. DSpace has an *excellent workflow process* and has been rated as one of the best amongst similar systems. The reason is that work flow process and the submission policy are configurable/ customized as per the institutions' requirements.
Each DSpace service is comprised of *Communities* - groups that contribute contents to DSpace. Communities in turn each have *Collections*, which contain the content items or files, e.g. in a university environment, Communities might be departments, labs, research centres, schools, etc. A Sub-community may be created under a community. e.g. Department of Library & Information Science is a sub-community of the community of School of Economics, Management & Information Science.

User Interface: DSpace being an open source technology platform, its ser interface can be customized and capabilities extended as per the requirements of institutions and individual departments.
DSpace is available over 20 languages and if English is not the local language, it can be customized.

User Administration: In DSpace, administration involves functions like

- create communities
- create collections
- create collection authorization

These functions can be performed by logging into DSpace as an administrator. The administrator is created at the time of installation of the software. It is here that the username and password for the administrator is registered. Thus logging into DSpace requires a username which is usually an e-mail ID along with a password.
DSpace has usage statistics and reports capabilities.

System Administration: DSpace uses postgreSQL as its back-end. It is a database which stores and manages information about communities, collections, members, e-mail addresses, and passwords, e-groups namely reviewers, metadata of digital items.
Automated content acquisition, harvesting and automatic metadata generation functions are well performed by DSpace.
Preservation issues, standards persistent document identification: DSpace is the best among the four candidates as its theory is to make its contents outlast. DSpace keeps a file's original name, size and created date. It uses a built-in Message Digest (MD5) integrity check algorithm for security applications to ensure the correctness of each file. More importantly, it defines a migration strategy including introducing the concept of file formats as a hierarchy of "unknown", "known", and "supported".
Though the use of Digital Object Identifier (DOI) as a system for identifying content objects in digital environment is quite prevalent, DSpace uses the Handle System developed by the Corporation for National Research Initiatives (CNRI) as part of the Computer Science

Technical Reports (CSRT) funded by the Defense Advanced Research Projects Agency (DARPA), for assigning globally unique identifiers to digital objects.

Interoperable: DSpace implemented OAI-PMH to support interoperability with other digital repositories, systems, e-print servers. However it is not Z39.50 compliant.

Standard Compliance: DSpace accepts many file formats like PDF (Portable Document Format) MS-Word, Post Script, for text format. XML, HTML, SGML(Standard Generalised mark-up Language), for structured documents. TIFF, MPEG (Moving Pictures Expert Group), JPEG, GIF for images

Metadata: DSpace uses Dublin Core set of elements for furnishing metadata and describing items intellectually.

System Support & Help: One can find DSpace system documentation on the SourceForge project web site or on the DSpace Technology page. The DSpace Community of developers support one another and exchange ideas and solutions on the DSpace mailing lists. Before one post a question or problem, one has to check to see if the question has been answered already. It can be started by searching the DSpace mailing list archives then by reading the FAQ and checking the technical documentation. If one still has not found an answer or solution, one can post questions to DSpace-tech, where members of the DSpace community will offer their assistance. One can report bugs and suggest enhancements through the SourceForge system. Bugs will be fixed as soon as possible, within the limits of the DSpace team's technical support resources. The team considers all enhancements, and if an enhancement is accepted, adds it to the enhancement list for development as time and resources allow. Also, any users working with the open-source code are welcome to fix a bug or make an improvement to the system. By logging on to DSpace Community Development Guidelines one can learn how.

Types of Contents: Theses, Dissertations lectures, speeches, conference, seminar proceedings, video, government publications, etc.

Retrieval Features: (Browsing and Searching in DSpace)
Browsing :- Browsing allows the user to go through a list of items in some specified order. DSpace allows browsing through:

- By Community/Collection
- By Title
- By Author
- By Date
- By Subject
- By Keywords

Searching: - DSpace allows a user to conduct a search in two ways. A search could be done through all communities and collections of the repository or it could be restricted to a specific community or collection. DSpace supports the following types of search.

- Exact term/Phrase Search
- Field Search
- Wild cards & Stemming

- Proximity Search
- Boolean Search
- Field Grouping
- Advanced Search
- Full Text

GREENSTONE

Content Management:
Greenstone does not have an built-in work flow process. Therefore users are unable to customize the work flow process to meet their unique workflow needs while building various collections. Greenstone envisages a ”librarian” who is responsible for creating collections from existing resources and distributing them over the Web or on CD-ROMs.

USER INTERFACE
Greenstone’s user interface is specifically designed to be highly extensible and customizable. It allows users to gather together set of documents, enrich it by importing or assigning metadata, build them into Greenstone collection and serve it from their web site.
Extensible : software “plug-ins” accommodate different documents and metadata types ie. software- plug –ins processes different text format.
Greenstone is multilingual in nature. It is fully documented in English, French, Spanish, and Russian and multiple Asian languages. As a result, Greenstone is popular in European and Asian countries.

USER ADMINISTRATION
Greenstone includes an “administrative” function whereby specified users can examined the composition of all collections; protect documents so that they can be accessed only by registered users on presentation of password. As an additional safe guard against misuse, the server administrator is automatically notified of each new collection. However, Greenstone was not rated well on authorization issues, as it only defines three kinds of users: general users, collection builders and administrators, which are not enough to meet the needs for the management of the content and the system.
Usage pattern -Greenstone places great emphasis on end-user functionality such as monitoring and reporting. e.g. usage reports and statistics help a library administrator to determine training of staff and identify popular file accessed.

SYSTEM ADMINISTRATION
Greenstone perform the functions of automated content acquisition, harvesting and automated metadata generation well which in turn makes DL maintenance much easier and can reduce labour cost.
Preservation :- With Greenstone, a file loses its original name after the submission process but keeps its other identities such as size and created date. This is a problem that Greenstone needs to fix in future releases. Moreover, Greenstone does not check data integrity via general checksum techniques, which may not guarantee that an imported digital object is the same as the original one. Regarding the strategy for long-term preservation, Greenstone has some features for long-term preservation as its multiple plug-

ins automatically convert files in common formats (e.g. Word, PDF, PS) to their corresponding HTML documents and keep the files in the original formats at the same time. The use of non-standard persistent URL allows users to access a digital object, but is unable to resist an object's changes in location and state.

INTEROPERABLE

Greenstone is highly interoperable using protocols such as OAI-PMH. It is also Z39.50 compatible for accessing external servers and for presenting Greenstone collection to external client.

STANDARD COMPLIANCE

For presentation of *structured document*, the most common format managed with Greenstone environment are : Extensible mark-up language (XML), Hyper Text mark-up language (HTML), PDF.

For *images* :- GIF,TIFF, JPEG
For *text format :* UNICODE, RTF (Rich Text Format),PDF, Post Script, Word.
For *audio & video:* MPEG(Moving Pictures Experts Group), MP3.

METADATA

Greenstone is very flexible in metadata support. It can support any metadata sets, if the desired schema for the metadata set is provided. The original package include Dublin Core Metadata schema.

SYSTEM SUPPORT & HELP

Extensive *Documentation* for Greenstone is available.
Mailing lists intended primarily for discussions is also available.
To *report any bugs*, one can address it to greenstone@ cs.waikato.ac.nz
It also provides *Manuals and Help Desk Support*.

TYPES OF CONTENTS

Newspaper articles to technical documents, educational journals to history, visual arts to videos, pop music collection to ethnic folk songs.

RETRIEVAL FEATURES

Greenstone's extensible search module supports Wildcard, Boolean and full-text searches and to some extent supports Proximity commands. Search results can be displayed by author, subject, title, dates, keywords, hierarchical classification and collection.

EPrints

CONTENT MANAGEMENT

In EPrints authors/submitters must create a repository (register with the archive) in order to be able to submit documents. The repository administrator controls what information is requested and what information to be used for creating metadata for submitted documents.

The EPrints movement has traditionally been associated with the so-called 'self-archiving' where authors themselves format their documents and submit them which will be reviewed and approved / rejected by the Eprints administrator.

USER INTERFACE

User interface in EPrints is easily customizable as required. The 'look and feel' of EPrints repository site can be customized by adapting scripts that control the presentation.
With version 2, users can easily identify the kind of content which is displayed and its status e.g. Published, In Press or Unpublished. The archive can be searched by publication or document type but this information is not displayed in the record screen. The flexibility of this software is that it enables authors to add fields to ensure that they are displayed.
Eprints is Unicode compliant, so that repositories' content and user interface can be in any language that has a script, i.e. it supports multilingual collections.

USER ADMINISTRATION

Users or readers are encouraged to create an account. Registered users may set up an e-mail alerting service for new content in their subject area.
The current version of EPrints (3) assigns each user a password which is somewhat annoyance to them since they cannot assigned their own password, they are likely to forget them. EPrints lacks report and inquiry capabilities i.e. generation of usage report and usage statistics is not very straight forward.

SYSTEM ADMINISTRATION

EPrints perform the functions of automated content acquisition, harvesting and automated metadata generation well which in turn makes DL maintenance much easier and can reduce labour cost.
Users can create their own persistent identifiers independent of the EPrints system.

INTEROPERABLE

EPrints is highly interoperable using protocols such as OAI-PMH.

STANDARD COMPLIANCE

EPrints support any type of document format including HTML, PDF, Post Scripts, ASCII, RTF. Possible additions maybe specialized document formats such as TeX or LaTeX, used by mathematicians or physicists.

METADATA

EPrints supports Dublin Core metadata.

SYSTEM SUPPORT & HELP

Extensive *Documentation* for EPrints is available on Wiki. EPrints Wiki pages have been set up for sharing experiences, scripts and tools related institutional repositories.
Discussion lists intended primarily for discussions are very active. Many times most of the queries are answered very quickly by the developers themselves which makes the solutions very reliable.
Eprints *lacks* formal help desk support.

TYPES OF CONTENTS

Scholarly work including Theses, Dissertations; journal articles, project reports, multimedia artifacts, teaching materials, digitized records.

RETRIEVAL FEATURES

EPrints search module supports metadata searching. Browsing views are generated for author, subject, title, year and country metadata. The software can easily be customized to include any other metadata fields as needed for browse views. E.g. users can add "keyword" and "documents" for browse views. Hierarchical subject browsing provides meaning browsing options to users.

FEDORA

Content Management

Fedora lacks submission support and review, where features such as notification of submission status or allowing users to modify submitted content are not available. It only provides capabilities to insert content.

USER INTERFACE

Fedora provides public Application Programming Interface (API) which allows users to access and customize the system.
Supports multilingual access.

USER ADMINISTRATION

Fedora does not support authorization much and as such it is not sufficient to manage the content of the system. In addition it has built-in access restriction features which restricts access based on machines' IP (Internet Protocol) addresses.
Fedora provides usage statistics.

SYSTEM ADMINISTRATION

Fedora has limited support for automated tools.
Preservation standards :- Fedora supports CNRI Handles.
Fedora's key strength is its support for preservation and standards since it allows multiple versions and formats for a digital object. This versioning feature is hardly found in the other candidates.
However Fedora neither checks data integrity nor describes the migration strategy.
Persistent document identifier: - To access a digital object, Fedora provides a unique persistent identifier (PID) for each digital object.

INTEROPERABLE

Fedora supports OAI-PMH. But it is not Z39.50 compliant.

STANDARD COMPLIANCE

Fedora does not provide a simple way to deposit digital objects. Therefore the user must create a Fedora compatible XML document containing reference to digital objects for ingest and export.
So, digital objects are stored as XML files in FOXML (Fedora XML) format.

METADATA

Fedora uses standards like METS (Metadata Encoding and Transmission Standard) and SOAP (Simple Object Access Protocol) in order to flexibly manage digital objects within a repository and to exchange them among repositories.
It also supports various metadata sets including Dublin Core and Marc.

SYSTEM SUPPORT & HELP

Currently two *mailing lists* serve the Fedora community –

- Fedora –commons-users for general discussions on technical support relating to Fedora.
- Fedora –commons-developers for technical discussions related to the development of the Fedora Repository software.
  For Fedora bugs, they are reported to Bugs report.

TYPES OF CONTENTS

Fedora can manage multiple digital contents including complex, compound, dynamic be it audio, video, image, text, numerical data, dataset, maps, geospatial data, e-books etc, used in a range of repository scenarios from digital asset management to institutional repositories to digital libraries.

RETRIEVAL FEATURES

Fedora supports Wildcard and Boolean searches but has no full-text search capability.
Fedora allows browsing through date of creation, title and object id.

Conclusion

Although the comparative study aims to be comprehensive, certain features such as hardware, time for installation, cost, manpower and other resources are not taken into account as these vary depending upon the implementing organization or individual and task involved. After analyzing the four selected software, it is found that some of the software still lacks certain functionalities perceived to be important, as gathered from literature. Among the four candidates, DSpace is able to fulfill almost all of the selected features. It is one of the software which has an excellent workflow process, whereas Greenstone does not have an in-built workflow process. It however places great emphasis on end-user functionality like usage reports and statistics. EPrints is a powerful and expensive solution for sharing scholarly works with the world, and has been associated with “self-archiving”. However it lacks usage patterns. The current version of EPrints (3) assigns each user a password which is somewhat annoyance to them since they cannot assigned their own password, they are likely to forget them. Fedora‘s key strength is its support for preservation and standards since it allows multiple versions and formats for a digital object. This versioning feature is hardly found in the other candidates. But it does not support password authentication much. Access to its contents is based on the machines’ IP (Internet Protocol) addresses. Further, based on the study, it can be concluded that each software has its own strengths and weaknesses that will appeal differently to organizations and individuals depending on their needs.

References:

1. **Arms**, W.Y. (1995). "Key concepts in the architecture of the digital library "available at: http://www.dlib.org/dlib/July95/07arms.html, (accessed on 10/10/2007)
2. **Association of Research Libraries** (1995), "Definition and purposes of a digital library", available at www.arl.org/sunsite/definition.html, (accessed on 10/10/2007)
3. **Bell**, Jonathan & Lewis, Stuart,(2006), Using OAI-PMH and METS for exporting metadata and digital objects between repositories, available at : http://www.emeraldinsight.com/10.1108/00330330610681349 (accessed on 12/03/2008).
4. Building Digital Library Collections With Open Source Software (2005) available at : http:// www.asis.org/Conferences/AM05/Digital_libraries.html (accessed on 29/02/2008).
5. **Cleveland,** Gary,1998, Digital Libraries: Definitions, Issues and Challenges, available at : http://www.ifla.org (accessed on 29/09/2007)
6. DSpace FAQs (2007), available at: http://www.dspace.org/FAQs.htm (accessed on 9/10/2007).
7. DSpace software website (2007), available at: http://www.dspace.org (accessed on 9/10/2007).
8. Dublin Core website (2008), available at :http://www.dublincore.org (accessed on 29/02/2008).
9. EPrints for Digital Repositories, (2008) available at : http://www.eprints.org (accessed on 01/04//2008).
10. Fedora (Software)(2008), available at : http://en.wikipedia.org/wiki/fedora_software (accessed on 10/03/2008).
11. Fedora website (2008), available at : http://www.fedora-commons.org (accessed on 01/04/2008).
12. **Goh,** Dion Hoe-Lian et al.(2006), A checklist for evaluating open source digital library software, available at http://www.emeraldinsight.com/10.1108/14684520610686283. (accessed on 19/12/2007).
13. Greenstone Digital Library Software website (2006). http://www.greenstone.org/ (accessed on 20/12/2007).
14. Greenstone User Survey (2004). http://www.ils.unc.edu/~sheble/greenstone/survey.html (accessed on 17/12/2007).
15. Greenstone website (2007), available at : http://www.greenstone.org (accessed on 10/11/2007).
16. Introducing DSpace (2007), available at http://www.dspace.org/introducingdspace.htm (accessed on 9/10/2007).
17. **Jayakanth**, Francis, et al. (2008), ePrints@IISc: India's first and fastest growing institutional repository, available at : http://www.emeraldinsight.com/10.1108/10650750810847260 (accessed on 12/03/2008).
18. **Jeevan**, V.K.J.,(2003), Digital Libraries, New Delhi, Ess publications, p. 1,3-4.
19. **Jose,** Sanjo (2007), Adoption of Open Source Digital Library Software Packages: A Survey, available at : http://eprints.rclis.org/archive/00008750/01/sanjojose.pdf (accessed on 17/12/2007).
20. **Krishnamurthy**, M.(2003), Digital Library Using Greenstone :A Study, *SRELS Journal of Information Management,* V. 40, No. 4, p. 313-320.
21. **Lagoze**, Carl, et al. (2005), Fedora: an architecture for complex objects and their relationships, available at : http:// springerlink.com/content/x7224797g8703g30/ (accessed on 10/03/2008).

22. **Leiner**, B.M. (1998), "The NCSTRL approach to open architecture for the confederated digital library", D-Lib Magazine, available at www.dlib.org/dlib/december98/leiner/12leiner.html, (accessed on 10/10/2007)
23. **Muir**, Scott P.,(*2005),* An introduction to the open source software issue, *Library Hi Tech*, V. 23 No.4, p.465-468.
24. **National Science Foundation**(1999), Digital Libraries Initiative : Available Research, US Federal Government, available at http://dli2.nsf.gov/dlione/,(accessed on 10/10/2007)
25. Open Source Initiative (2005), *The Open Source Definition*, available at: www.opensource.org/docs/definition.php(accessed on 15/01/2008).
26. Open Source website (2007), available on http://www.opensource.org (accessed on 20/12/2008).
27. **Seadle**, Micheal & Greifeneder, Elke(2007), Defining a Digital Library, *Library Hi Tech*, V. 25 No. 2, p.169-173.
28. **Sharma,** R.K. & Vishwanathan, K.R.(2001), Digital Libraries: development and challenges, Library Review, V.50 No. 1, p.10-15.
29. **Sponsler**, Ed & Velde, Eric F. Van (2001), EPrints.org Software: A Review, available at : http://resolver.library.caltech.edu/caltechLIB:2001.004 (accessed on 012/04/2008).
30. **Stanger,** Nigel & McGregor, Graham (2007), EPrints Makes its Mark, available at : http://www.emeraldinsight.com/10.1108/10650750710748432 (accessed on 9/10/2008).
31. **Waters**, D.J. (1998), "What are digital libraries?" available at: http://www.clir.org/pubs/issues/issues04.html (accessed on 2/10/2007).
32. **Witten**, Ian H. et.al. (2002), The promise of digital libraries in developing countries, *The Electronic Library*, V. 20, No. 1, p. 7-13. also available at : http://www.acm.org/pubs/articles/journals/cacm/2001-44-5/p82-witten/p82-witten.pdf