Abstract

This study is an examination of features of the four most popular digital library open source software packages against a set of predetermined criteria that is deemed to be essential for the development of a digital library. While analysing and identifying the similarities, differences, strengths, and weakness of the open source software packages, it is indicated that open source digital library software still lacks certain functionalities that are perceived to be important. Each software package has its own individual strengths and weaknesses that will appeal to organizations and stakeholders with different needs. The study is expected to help those library professionals, seriously interested in implementing a digital library by providing a checklist to evaluate how well a particular software fits into their specific implementation requirements. Certain good features of the four open source software packages have also been pointed out via a comparison analysis.
Introduction

Open source software (OSS) refers to any software that is free, and hence, is often confused with freeware and shareware. In contrast to open source, freeware is software that is released free of cost in binary format only. Its licenses usually prohibit modifications and commercial redistribution. On the other hand, shareware is software that is released free of cost in binary format but only for a limited trial period, after which users are encouraged to purchase the software.

The availability of the source code in open source software allows users to modify and make improvements to it. Such contributions can originate from a diverse talent pool of programmers. Thus, open source software tends to have more functions—being developed by users of the software themselves—as compared to commercial software, where a vendor’s priority (usually profit generation) might not be in line with the needs of users. Further, because the source code is accessible and modifiable, contributions also lead to improvements in the functionality of the software. In addition, updates can be obtained at low or even no cost, and there are no royalties or license fees. Moreover, there is less likelihood of being dependent on a single software provider or being trapped into long-term software support contracts, which restrict flexibility in implementation (Surman and Diceman 2004).

However, open source software is not devoid of disadvantages. One of the common complaints is the lack of formal support and training that a commercial software package would offer (Caton 2004). Often, support is provided through mailing lists and discussion forums. In addition, open source software is also not known for ease of use as the focus is usually on functionality. Consequently, open source adopters will have to take greater personal responsibility, in terms of leveraging staff expertise to implement and maintain their systems, including hardware and network infrastructure (Poynder 2001).

Nevertheless, open source is increasingly considered as an alternative to commercial digital library systems, mainly due to dissatisfaction with functionality (Breeding 2002). Another factor is the increasing budget cuts that libraries face (Evans 2005). The cost of software production and maintenance is also rising dramatically. As a result, open source digital library software, with its free access and satisfactory level of functionality, are steadily gaining ground in both usage and interest.

The Open Source Software (OSS) model makes the source code available to users, who can then make the necessary changes to it to tailor it according to their own requirements. With many OSS applications now available for library and information management, organizations now have another option for acquiring and implementing systems, as well as access to opportunities to participate in OSS projects. Examples of such systems include Greenstone, DSpace, E-prints, and so on. OSS is extremely popular with technically sophisticated users, who are often also the software developers. This study highlights the comparison, features, function, and usability of OSS like Greenstone, E-prints, Fedora, and DSpace.

Over the last several years, the rapid growth of OSS has captured the attention of research librarians and has created new opportunities for libraries. OSS can benefit libraries by lowering both initial and ongoing costs, eliminating vendor lock-in, and by allowing greater flexibility (Naik and Shivalingaiah 2006).

Objectives of the study

The main objectives of this study are to:

- Determine the features that characterize ‘good’ open source digital library software and develop a checklist from these features. Here, we identified the essential categories of features a digital library software should possess. Examples include content
Comparison of open source software for digital libraries

management, searching, and metadata standards. These categories were then subdivided into supporting features. From these, a checklist that covered all such features was developed.

- Evaluate the features of four open source digital library software packages against the checklist. Scores were assigned to each software, depending on its adherence to the features mentioned in the checklist.
- Analyse the strengths and weaknesses of each of the four digital library softwares. After compiling the scores, a comparison of the features was conducted with the objective of identifying similarities, differences, strengths, and weakness of each (Goh D H, Chua A, Khoo D A, et al. 2006).

Four most popular digital library software packages were selected for the study. These were DSpace, Eprints, Fedora, and Greenstone (Jose 2007).

Digital library software packages

Many open source software packages are available for organizations and individuals alike to create digital libraries. However, an easy-to-use instrument to evaluate these digital library software packages does not exist. The present work attempts to develop a comprehensive checklist for assessing digital libraries. Its flexibility allows users to tailor it to accommodate new categories, items, and weightage schemes to reflect the needs of different digital library implementations (Goh D H, Chua A, Khoo D A, et al. 2006).

Four digital library packages were selected for our evaluation:

- Greenstone (www.greenstone.org) has been developed by the New Zealand Digital Library Project at the University of Waikato, and is distributed in cooperation with UNESCO and the Human Info NGO. Greenstone is a tool for building libraries that aims to empower users—particularly those in universities, libraries, and other public service institutions—to build large distributed digital library collections. The current installed base is unknown, but the number of downloads of the software appears to be large. It is an open source software that can handle multilingual documents, and has a search and browse facility under GNU General Public License (Laxminarsaiah and Rajgoli 2005).

  Greenstone is a tool for creating and managing digital library collections. It runs on Windows as well as UNIX.

  The Greenstone digital library software builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt automatically. The system is extensible: software ‘plugins’ accommodate different document and metadata types.

- DSpace: Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP) have jointly created DSpace software as a digital repository to manage intellectual output of multidisciplinary research and development organizations. DSpace supports digital preservations and planning and managing institutional repository in a large institution. It also allows workflow and customization and supports community/collection-based content and submission by different user communities (Laxminarsaiah and Rajgoli 2005).

  It is based on a three-layered architecture, namely application layer, business layer, and storage layer. The application layer covers the interface to the systems, the Web and user and interface and batch loader, in particular. The business layer is where lie the DSpace specific functionality, workflow, content management, administration, and search and browse modules. The storage layer is implemented...
using the file system, as managed by PostgreSQL databases.

The system is primarily written in Java, and uses only free software libraries and tools, including the PostgreSQL, RDBMS, Java servlet, Apache and tomcat, Lucene search engines, XML tools, and RDF tool. Collections within communities consist of items, which are, in turn, composed of one or more bit streams, or physical files of digital materials. DSpace item is a single bit stream. For example, a digital image encoded as a TIFF file or a digital document encoded as a PDF file.

It is a groundbreaking digital institutional repository that captures, stores, indexes, preserves, and redistributes the intellectual output of a university’s research faculty in digital formats. It manages and distributes digital items made up of digital files (or bit streams), and allows the creation, indexing, and searching of associated metadata to locate and retrieve the items. DSpace also supports submission, management, and access of digital content (Naik and Shivalingaiah 2006).

- **EPrints** (www.eprints.org) is a generic archive developed by the University of Southampton. The first version of the software was publicly released in late 2000. The objective behind the creation of EPrints was to facilitate open access to peer-reviewed research and scholarly literature through OAI. However, EPrints also serves as an archive for other electronic documents such as images and audio (Goh D H, Chua A, Khoo D A, et al. 2006).

  EPrints is intended to create a highly configurable Web-based archive. Its primary goal is to be an open archive for research papers, but it could be easily used for other things such as images, research data, audio archives—in fact, anything that can be stored digitally by making changes in configuration. It works on Linux and needs MySQL, Perl modules and Apache Web server. The software can be installed by any institution across the world. By its integrated advanced search, extended metadata, and other features, the software can be customized to meet local requirements (Laxminarsaiah and Rajgoli 2005).

- **Fedora** (www.fedora.info) has been jointly developed by the University of Virginia and Cornell University. Its first version was released in 2003. The objective of Fedora 1.0 was to create a production quality system using XML and Web services to deliver digital content. Fedora supports digital asset management, institutional repositories, digital archives, content management systems, scholarly publishing enterprises, and digital libraries. The system is designed to be a foundation upon which full-featured institutional repositories and other interoperable Web-based digital libraries can be built. It currently has a distributed installed base of more than 360, with collection sizes of 10 million objects. As an open source software, Fedora offers organizations a flexible service-oriented architecture for managing and delivering digital content. At its core lies a powerful digital object model that supports multiple views of each digital object and the relationships among digital objects. All functions of Fedora, both at the object and repository level, are revealed as Web services. These functions can be protected with fine-grained access control policies. This unique combination of features makes Fedora an attractive solution in a variety of domains. Some examples of applications that are built upon Fedora include library collections management, multimedia authoring systems, archival repositories, institutional repositories, and digital libraries for education.

**Evaluation of open source software**

To effectively evaluate digital library softwares, a framework is necessary to guide the planning,
controlling, and reporting of the evaluation. Common elements among software packages need to be examined so that suitable conclusions can be drawn. To accomplish this objective, evaluation instruments are needed, of which, several types are available including (Punter 1997):

1) static analysis of code: for structural measurement or anomaly checking;
2) dynamic analysis of code: for test coverage or failure data;
3) reference tools that compare the software product;
4) reference statistical data; and
5) inspection with checklists.

In designing a digital library, there is no decision more important than the selection of quality software that forms the platform for service delivery. The variety of choices available makes the selection somewhat daunting. The key is a careful definition of the nature of information in the library and how it will be used. In the present work, a review of the digital library literature yielded the following five broad requirements that were used as our evaluation criteria (Dobson and Ernest 2000):

- **Content management**: This requirement is related to the ease with which content is created, submitted, reviewed, organized, and versioned within a digital library. It also encompasses the provision for searching and browsing functions such as metadata search, full-text search, and hierarchical subject browsing. Additionally, content encoded in a variety of popular formats, including text (ASCII, UNICODE, RTF), image (TIFF, GIF, JPEG), presentation (Adobe PostScript and Adobe PDF), structured formats (HTML and XML), audio and video (Real, MP3, AVI, and MPEG) ought to be supported.

- **User interface**: The user interface requirement covers the flexibility in customizing the interface to suit the needs of different digital library implementations as well as the support for multilingual access. With multilingual access, the user is able to specify the language for the digital library’s user interface and the cataloging information stored within it (Witten and Bainbridge 2002).

- **User administration**: This requirement is concerned with the range of functions needed to manage users of a digital library. For example, access to content in a digital library needs to be restricted through password authentication, IP filtering, and proxy filtering. Also, usage patterns have to be monitored and reported. When usage patterns are analysed, the needs and interests of digital library users can be better understood (Jones, Cunningham, and McNab 1998).

- **System administration**: This requirement is related to the back-end maintenance of a digital library. Automatic tools are useful, particularly for large digital libraries where maintenance work is labour-intensive (Cordeiro 2004). Functions such as automated content acquisition, harvesting, and automatic metadata generation (including named entity recognition and automatic subject indexing/classification) make digital library maintenance much easier. Moreover, a digital library needs to support preservation standards, as well as persistent document identification so that the transfer of digital materials from one hardware/software configuration to another would not compromise reference citations and other links (Hedstrom 2001).

- **Other requirements**: A digital library needs to be interoperable with other systems to which it is connected. This allows each system to evolve independently without sacrificing the ability to communicate with each other. At least two basic interoperability protocols should be supported, namely, Z39.50 and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Additionally, the digital library must be compliant with standards established for digital library
collection and services. Examples of standards include XML for representation of information; XHTML for Web pages; GIF, TIFF, and JPEG for images; Unicode for multilingual support and information interchange; and Dublin Core and MARC 21 for metadata. Finally, a digital library needs to provide mechanisms through which digital library administrators and developers can obtain system support and help. Such mechanisms include documentation, manuals, mailing lists, discussion forums, bug tracking, feature request systems, and formal helpdesk support.

Digital library evaluation checklist

Due to the lack of a universally accepted definition of a digital library, there is no common methodology for the selection of a good digital library software. With this in mind, the present study aimed to develop a simple-to-use instrument for evaluating digital library software with the following characteristics:

- **Comprehensiveness** The evaluation criteria should cover all the key areas involved in selection of digital library software.
- **Useability** The instrument should be simple to understand, and more importantly, easy to employ by a variety of users with or without a background in software evaluation.
- **Flexibility** In choosing a digital library software, different stakeholders may place different emphasis on various criteria. For example, a library administrator may be more concerned with the ease of submission of material, whereas a researcher may be more interested in better search functionality. The weights associated with each criterion should be easily modifiable to reflect different stakeholder needs.
- **Expandability** The instrument should be easily expandable to include additional factors and criteria such as new and emerging standards.

Our digital library evaluation checklist consists of 12 categories of items, each with varying degrees of importance: content management, content acquisition, metadata, search, access control and security, report and inquiry, preservation, interoperability, user interface, standards compliance, automatic tools, and support (Goh D H, Chua A, Khoo D A, et al. 2006).

**Methodology for evaluation**

The method of assigning weights to evaluation criteria was adapted from Edmonds and Urban’s methodology, who recommended the use of the Delphi technique. In the original technique, a committee anonymously assigns weights to each criterion, usually through a questionnaire. It then reviews the results, and if there is no consensus, the steps are repeated until a consensus is reached. In the present study, we modified the Delphi technique by having a group of four people trained in information science and familiar with digital library concepts assign weights to each category and its respective items independently. The total sum of the category weights was 100, while the total sum of the items in each category was 10. Discrepancies were then resolved via face-to-face discussions, in which each person provided justifications for the reasons behind his/her decisions. Pairwise comparisons were also conducted as part of the process of formulating appropriate weights for the checklist. In pairwise comparison, the relative importance of each criterion against every other criterion is determined, often in a group session that is preceded by individual assessment (Koczkodaj, Herman, and Orlowski 1997).

Next, the four selected digital library software packages were evaluated using the checklist. Scores were computed by considering the software as a group, using the same four people that developed the checklist, but on a separate occasion. In cases where the evaluators disagreed over whether a particular criterion was met in a software package, a majority vote was taken. In case of a tie, a consensus was arrived through
emailing the digital library software developers or consulting other sources (Goh D H, Chua A, Khoo D A, et al. 2006).

**Findings and analyses**

Table I shows the scores for each of the sub-categories of the four digital library softwares evaluated, while Table IA depicts the consolidated score of all the categories for the four digital library softwares.

**Content management**

This category involves procedures and tools pertaining to the submission of content into the digital library as well as management of the submission process. As shown in Table I (row 1), all digital library software, with the exception of Fedora, satisfied most—if not all—of the criteria. Fedora only managed a score of 4.50 out of 10. This comparatively poor performance is due mainly to a lack of submission support and review. Fedora only provides capabilities to insert content, but not features such as notification of submission status or allowing users to modify submitted content.

**Content acquisition**

Content acquisition refers to functions related to content import/export, versioning, and supported document formats. Table I (row 2) shows that all the selected digital libraries managed to fulfill this criterion. EPrints, in particular, achieved a full score of 10.

**Metadata**

As mentioned earlier, metadata support in digital libraries is vital when it comes to content

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Category</th>
<th>E-prints</th>
<th>Fedora</th>
<th>Greenstone</th>
<th>DSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Content management</td>
<td>8.00</td>
<td>4.50</td>
<td>9.00</td>
<td>10.00*</td>
</tr>
<tr>
<td>2</td>
<td>Content acquisition</td>
<td>10.00*</td>
<td>9.44</td>
<td>9.17</td>
<td>8.67</td>
</tr>
<tr>
<td>3</td>
<td>Metadata</td>
<td>5.13</td>
<td>7.00*</td>
<td>5.75</td>
<td>6.25</td>
</tr>
<tr>
<td>4</td>
<td>Search support</td>
<td>3.40</td>
<td>4.96</td>
<td>7.25</td>
<td>8.71*</td>
</tr>
<tr>
<td>5</td>
<td>Access control and privacy</td>
<td>8.33*</td>
<td>6.17</td>
<td>7.83</td>
<td>7.67</td>
</tr>
<tr>
<td>6</td>
<td>Report and inquiry capabilities</td>
<td>0.00</td>
<td>6.00</td>
<td>10.00*</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>Preservation</td>
<td>5.50</td>
<td>10.00*</td>
<td>2.50</td>
<td>7.50</td>
</tr>
<tr>
<td>8</td>
<td>Interoperability</td>
<td>5.00</td>
<td>6.00</td>
<td>8.00*</td>
<td>5.00</td>
</tr>
<tr>
<td>9</td>
<td>User interface</td>
<td>10.00*</td>
<td>10.00*</td>
<td>10.00*</td>
<td>10.00*</td>
</tr>
<tr>
<td>10</td>
<td>Standards compliance</td>
<td>10.00*</td>
<td>10.00*</td>
<td>10.00*</td>
<td>10.00*</td>
</tr>
<tr>
<td>11</td>
<td>Automatic tools</td>
<td>5.00</td>
<td>5.00</td>
<td>10.00*</td>
<td>10.00*</td>
</tr>
<tr>
<td>12</td>
<td>Support and maintenance</td>
<td>6.00</td>
<td>10.00*</td>
<td>10.00*</td>
<td>10.00*</td>
</tr>
</tbody>
</table>

Note: * Indicates the highest category score

Method for computing sub-category score:

Number of check marks * sub-category weight ÷ No. of items in the sub-category = Sub-category score

indexing, storage, access, and preservation. However, the performance in this area was disappointing. As shown in Table I (row 3), most of the digital libraries in our study only supported a few metadata standards. While it is encouraging that at least core standards like MARC21 and Dublin Core were supported, emerging metadata schemas such as EAD and METS were missing in all, except Fedora (Guenther and McCallum 2003).

**Search support**

Search support refers to a range of searching and browsing functions such as metadata search, full-text search, and hierarchical subject browsing. Considering the importance of searching functionality, Table I (row 4) shows that performance in this respect across the four digital libraries was varied, ranging from a low of 3.40 to a high of only 8.71. In particular, E-prints’ poor performance was due to the absence of full-text search support, as well as metadata search. In addition, none of the software exhibited proximity searching capabilities.

**Access control and privacy**

Access control and privacy include the administration of passwords, as well as the management of users’ accounts and rights to specified locations within the digital library. Most of the digital libraries surveyed scored well for this indicator (see Table I, row 5), with E-prints being the best overall performer. DSpace obtained the best score in password

---

**Table IA Evaluation results of consolidated score of all categories**

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Category</th>
<th>E-prints</th>
<th>Fedora</th>
<th>Greenstone</th>
<th>DSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Content management (7)</td>
<td>5.6</td>
<td>3.15</td>
<td>6.30</td>
<td>7.00</td>
</tr>
<tr>
<td>2</td>
<td>Content acquisition (11)</td>
<td>11.00</td>
<td>10.38</td>
<td>10.08</td>
<td>9.53</td>
</tr>
<tr>
<td>3</td>
<td>Metadata (12)</td>
<td>6.15</td>
<td>8.40</td>
<td>6.90</td>
<td>7.50</td>
</tr>
<tr>
<td>4</td>
<td>Search support (14)</td>
<td>4.76</td>
<td>6.94</td>
<td>10.15</td>
<td>12.19</td>
</tr>
<tr>
<td>5</td>
<td>Access control and privacy (7)</td>
<td>5.83</td>
<td>4.31</td>
<td>5.48</td>
<td>7.50</td>
</tr>
<tr>
<td>6</td>
<td>Report and inquiry capabilities (3)</td>
<td>0.00</td>
<td>1.80</td>
<td>3.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>Preservation (5)</td>
<td>2.75</td>
<td>5</td>
<td>1.25</td>
<td>3.75</td>
</tr>
<tr>
<td>8</td>
<td>Interoperability (10)</td>
<td>5.00</td>
<td>6.00</td>
<td>8.00</td>
<td>5.00</td>
</tr>
<tr>
<td>9</td>
<td>User interface (8)</td>
<td>8.00</td>
<td>8.00</td>
<td>8.00</td>
<td>8.00</td>
</tr>
<tr>
<td>10</td>
<td>Standards compliance (10)</td>
<td>10.00</td>
<td>10.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>Automatic tools (4)</td>
<td>2.00</td>
<td>2.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>12</td>
<td>Support and maintenance (9)</td>
<td>5.50</td>
<td>9.00</td>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td></td>
<td>Consolidated score</td>
<td>66.49</td>
<td>74.98</td>
<td>73.16</td>
<td>72.33</td>
</tr>
</tbody>
</table>

Method for computing category score:

\[
\frac{\text{[(Sum of sub-category score) \times (Category weights)]}}{10} = \text{Category score}
\]

For example, Category score of ‘content management’ category of E-prints software is:

\[
(8 \times 7) \div 10 = 5.6
\]
administration, having not only system-assigned passwords, but also the ability to select passwords and retrieve forgotten passwords. Fedora scored well on access management in its support for IP address filtering, proxy filtering, and credential-based access.

Report and inquiry capabilities
This category deals with usage monitoring and reporting. Table I (row 6) shows that Greenstone was the only software that fulfilled all the requirements in this category. While Fedora provided usage statistics, it did not offer report generation tools. Both E-prints and DSpace lacked report and inquiry capabilities.

Preservation
It refers to preservation of metadata and quality control measures to ensure integrity, and persistent documentation identification for migration purposes (Hedstrom 2001). Fedora was a clear winner in this regard, with its support for CNRI handles, digital libraries, quality control, and provision for a prescribed digital preservation strategy (see Table I, row 7).

Interoperability
Interoperability is concerned with the benefits of integrating distributed collections and systems. Our study revealed that Greenstone was the best performer (see Table I, row 8) in this respect. All the software surveyed supported OAI-PMH. However, Z39.50 was only supported by Greenstone. This may be due to the protocol being much more complex to implement.

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 implementations. All the four digital library software surveyed obtained a full score (Table I, row 9), reflecting that these issues have been taken into consideration.

Standards compliance
Standards are important for the sharing of digital content and long-term digital preservation (Dawson 2004). Thus, this category was evaluated by looking for evidence of the usage of standards. As the only other category with a full score across all software (see Table I, row 10), there appears to be a demonstrated commitment to the use of standards. It should be noted, however, that such a commitment does not imply every conceivable standard should be adopted. The other evaluation categories should be consulted to determine which specific standards are supported by each digital library. For example, while most document and image format standards are supported by the four digital libraries, not all metadata formats are, with Dublin Core being the only one supported by all.

Automatic tools
This category refers to tools for automated content acquisition, harvesting, and metadata generation. In the context of digital libraries, automatic tools are useful for maintenance and can reduce labour costs, especially for large collections. Table I (row 11) shows that Greenstone and DSpace came up with full scores, while Fedora and E-prints did not fare that well.

Support and maintenance
Support and maintenance are important for all software systems. But open source software is often criticized to be lacking in these aspects. However, our results show that three out of the four digital libraries evaluated performed well in this category (see Table I, row 12) by offering documentation, manuals, mailing lists, discussion forums, bug tracking, feature request systems, and formal helpdesk support. Only E-prints fared relatively poorly, due to its lack of formal helpdesk support and documentation that was not updated.
Discussion

Consolidated scores

Figure 1 shows the consolidated scores of the four digital library softwares that were evaluated. Fedora emerged as the best performer (with a consolidated score of 74.98), followed closely by Greenstone (with a score of 73.16). These two were followed by DSpace and E-prints with scores of 72.33 and 66.49, respectively. It should be noted that the consolidated scores were obtained by summing all category scores after normalization by their respective category weights.

Fedora was the only software package that consistently fulfilled the majority of the criteria in many categories, and obtained maximum scores in four out of the twelve categories. These five indicators were content acquisition, metadata, standards compliance, user interface, and support and maintenance. In fact, Fedora secured full scores in four out of the twelve categories. Fedora’s key strength is its support for preservation and standards, in which full scores were obtained. It also ranked highest in the metadata category due to its support for many metadata standards. Other than the lack of Z39.50 support, Fedora appears to be a good candidate as far as long-term digital preservation needs are concerned. Fedora is also easily installed on a Windows 2003 server machine, although more configuration work is required as compared to Greenstone. However, Fedora has limited support for automated tools and content management features.

Greenstone places great emphasis on end-user functionality. For example, usage reports and statistics help a library administrator determine bottlenecks and identify popular files accessed. User interface customizability allows different digital libraries to create interfaces that suit the needs of their stakeholders, while automatic tools simplify content management and acquisition. In addition, Greenstone attained nearly perfect scores in content management and acquisition, implying that it considerably helps ease the task of managing content in a digital library.

Due to the ease in its installation, Greenstone is close to Fedora. Packaged in a single executable installation, the digital library becomes operational on a Windows 2003 server machine in less than an hour. Documentation for Greenstone is also extensive. There is a wealth of online documents and tutorials available on the Greenstone Website, and a number of organizations even offer training courses. In a nutshell, we believe that Greenstone is the most user-friendly software for creating digital libraries among the four softwares evaluated.

Similar to Greenstone, DSpace also secured maximum scores in five of the twelve categories. Although it was close to Greenstone in total score, DSpace performed slightly worse because of its lack of report and inquiry features. An issue with DSpace not reflected in the checklist was the difficulty in installing the software. As compared to the smooth installation of Greenstone, DSpace took time on a newly set up Linux machine, due to the number of other required software packages that needed to be installed and properly configured, and the extensive knowledge of Linux that was required.

DSpace is the most popular among the digital library solutions available in the open source domain. E-prints is also widely used. Educational institutions dominate in the use of these packages. Though many institutions have implemented digital libraries, only about
half of these are available online. Open access of knowledge is possible only if these repositories are made online. India is benefiting a lot from the open source movement (Jose 2007).

E-prints was the worst performer with a total score of 66.49. Its advantage is that the software, in the study, was the only one to obtain a full score in content acquisition, and that it supports the widest range of document formats. On the other hand, E-prints lacks in usage reporting and inquiry capabilities. It is also only available on the Linux platform, and therefore shares the same installation problems faced with Fedora. However, its greatest weakness is the low score (3.40) under the search category. Only metadata searching is supported and full-text search is not available in E-prints.

**Conclusion**

Although the checklist developed in the current study aims to be comprehensive, it is just the first step in the development of an exhaustive evaluation tool. The current version has some limitations in the assessment of digital library software. For example, the checklist does not take into account factors such as hardware, time, manpower, money, and other resources, as these may vary depending on the implementing organization or individual. The availability of application programming interfaces for implementing new features was also not considered. In addition, the weights given to the various categories were assigned through a consensus process among four evaluators. Other stakeholders with different viewpoints and needs may require different weightage schemes. However, as discussed previously, the checklist is flexible enough to accommodate new categories, items, and weightage schemes.

Extensive research was conducted to extract requirements for digital libraries, which led to the definition of criteria for digital library evaluation, and from these, a checklist was created. Assigning scores to each digital library software package against our checklist further reinforced the differences among these digital libraries in accommodating diverse needs. From the results of the evaluation, we have come to the conclusion that open source digital library softwares currently available still lack certain functionalities perceived to be important. However, among our four candidate digital libraries, Greenstone was able to fulfill most of the crucial requirements because of its strong support in end-user functionality. However, it must be noted that each software package has individual strengths and weaknesses that will appeal to different organizations and stakeholders with differing needs. Therefore, those interested in implementing a digital library can use the checklist to evaluate how well a particular software suits their specific implementation requirements (Goh D H, Chua A, Khoo D A, et al 2006).

Each of the above-mentioned software systems is designed to meet the original requirements of developing a digital library. DSpace supports community-based content policies and submission process, and accommodates various kinds of digital document formats. E-prints is a useful digital library system with a considerable user community. But when there is a need for technical support and training in using the software, Greenstone was found to be the most suitable. Though many libraries in India are using Greenstone and DSpace, some are also using E-prints because of its immense potential and the fact that it can support numerous forms and formats (Laxminarsaiah and Rajgoli 2005).

The Open Archives Initiative (OAI) has gained momentum since 2000 when eprints.org was launched OSS incorporates an interface that makes it easy for people to create their own library. Collections may be built and served locally from the user’s own Web server or remotely on a shared digital library host. End users can easily build new collections styled after existing ones from material on the Web or from their local files (or both), while collections can be updated and new ones brought online at any
time. OSS has a lot of potential for libraries and information centres, and there are a number of projects—including Greenstone, DSpace, and E-prints—that demonstrate its viability in this context. It provides library staff an option to be actively involved in development projects. This involvement can take many forms, such as reporting bugs, suggesting enhancements, and testing new versions. Currently available OSS projects cover application areas ranging from the traditional library management systems to innovations like Greenstone and DSpace, which complement traditional systems. These concepts and their benefits and importance to libraries are being examined. Benefits include lower costs, greater accessibility, and better prospects for long-term preservation of scholarly works (Naik 2006).

Traditional libraries are limited by storage space, while digital libraries/repositories have the potential to store massive amounts of scholarly information and requiring very little space in the process. Moreover, the cost of maintaining an institutional digital library is much lower than that of its traditional counterpart. Digital libraries can adopt innovations in electronic and audio-video book technology. Considering the advantages like absence of any physical boundary, round-the-clock availability of information, multiple access to information resources, faster information search and retrieval, preservation and conservation of exact copy of the original documents, and low costs, various organizations have opted for digital library using open source softwares. With the development of a digital library, information access and retrieval has been increased at campus levels and also at the global level. The aim is to deliver the right information to the right reader at the right time.

References


Caton M. 2004. Sugar sale: just the SFA basics. eWeek 21(50): 48

Cordeiro M. 2004. From rescue to long-term maintenance: preservation as a core function in the management of digital assets. VINE 34(1): 6–16


Comparison of open source software for digital libraries


