**Adoption of Software Packages by Digital Open Access Repositories in the World: A Survey**

**Sheeja N.K.** Assistant Librarian Department of Ship Technology, CUSAT.

**Susan Mathew** Assistant Librarian University Library, CUSAT.

**Surendran Cherukodan** Junior Librarian School of Engineering , CUSAT

**Abstract**

The open access movement is gaining importance for several reasons. It has two dimensions. The first one is open access publishing and the other one is self archiving. Self archiving of scholarly materials is an important area where several issues coalesce. The self archiving process is fulfilled through the establishment and maintenance of digital repositories. Hence it is relevant to understand the functioning of digital repositories. Software is an important component of any digital repository. This study tries to understand the trend of adoption of software for digital repositories throughout the world. The study also attempts to understand the features of major repository software, the nature and types of repository software and the relationship between open access and open source software.

**Keywords**:Open Source Software, Digital library software, DSpace, Eprints and Open Access Repositories

1. **Introduction**

The growth in reach, importance and impact of open access (OA) publishing over the last few years has been dramatic (Jackson and Richardson, 2014). The global push for “free and unrestricted online availability” of scholarly research outputs first coalesced during the Open Society Conference at Budapest in December 2001 (Johnson, 2014). The term "open access" (OA) was first formulated in three public statements namely the [Budapest Open Access Initiative](http://en.wikipedia.org/wiki/Budapest_Open_Access_Initiative) in February 2002, the [Bethesda Statement on Open Access Publishing](http://en.wikipedia.org/wiki/Bethesda_Statement_on_Open_Access_Publishing) in June 2003, and the [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](http://en.wikipedia.org/wiki/Berlin_Declaration_on_Open_Access_to_Knowledge_in_the_Sciences_and_Humanities) in October 2003. OAmeans unrestricted online access to peer-reviewed scholarly research. Budapest Open Access Initiative (2002), states open access as: free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. (Budapest Open Access Initiative (BOAI), p. 3)

BOAI also recommends two strategies for open access. The first one is open access publishing and the other one is self archiving. Today, most large publishing houses have launched new experiments with OA publishing options (Rodriguez, 2014). Self archiving is the practice of depositing e-prints (published papers and pre-prints) into open electronic archives set up by the researchers' own institutions. These repositories are known in different terms such as digital repositories, institutional repositories, and digital libraries etc. The institutional repository (IR), a university-based digital-asset management system, is fast emerging as a key component of the current debate on OA and reform of the scholarly communication process (Chan, 2004). Open-access IRs are the most cost-effective and immediate route to providing maximal access to the results of publicly funded research, thereby maximizing the potential research impact of these publications (Harnad, 2001a; 2001b; 2003). These digital repositories provide alternative channels for the provision of scholarly open access literature (Suber, 2012).

The authors of this paper use the term digital repository to include digital libraries, institutional repositories, E-print archive, digital archives and other online digital systems designed for the archiving of scholarly materials. A digital repository is an online archive in which authors and academics can deposit their work, with the intention that it will be openly available in digital form. The benefits of establishing a digital repository includes the provision of easy access to scholarly material generated by research community, long-term archiving of information and research output thereby preserving it for the future, the possibility of information to be widely and quickly disseminated so that it achieves the highest impact and increases the academic reputation of your institution (Pappalardo, 2007). On the recognition of these benefits, institutions started establishing digital repositories. There are more than 2000 repositories across the globe now. The Directory of Open Access Repositories (DOAR) and the Registry of Open Access Repositories (ROAR) stand as the authoritative online source to list digital libraries in the world.

1. **Software for Digital Repositories**

Among other essential requirements, the process of selection and adoption of a software package is an important aspect of any digital repository. A group of software applications was created in the 1990s and early 2000s to simplify open access contributions and management. This includes the development of many open source packages for digital repositories. Now there are several software packages specifically designed to create and maintain a digital repository. The purpose of this study is to provide a broad outline of the availability and use of various digital repository software packages around the world. The study further goes to investigate the features of digital repository software packages, developers of these software packages, types of software packages, nature of availability and the distribution of these software packages among digital repositories. The study also compares the features of five major software packages: DSpace, EPrints, Fedora, Digital Commons and OPUS.

**3. Related Studies**

Baruah (2007) discussed the importance and utility of the three concepts i.e. open access, open source and open standards in libraries that support resource sharing in a far better way, limiting the expenses and standardizing measures to share resources. Meitei & Devi (2009) discussed the role of Open Source Software in digital repository development and digital preservation with the successful adoption of open source application for library and information management system on the global digital information environment. Payne (2010)discussed the existing presence of OSS in libraries and the functionality and variety of OSS products. The study observed that libraries were seeking alternatives to proprietary applications which require specialized support and services. Hanumappa and Dora (2014) explored OSS market relevant to Indian libraries and reviewed the existing library software solutions. The study found that there is considerable interest among Indian libraries to adopt or migrate to OSS. Singh, Witt and Dorothea (2009) proposed a comparative analysis of four institutional repository software packages viz., EPrints, DSpace, Fedora and Zentity . The repositories were assessed on the basis of execution of four common workflows: consume, submit, accept, and batch. The study helped to identify the strengths and weaknesses of each repository applications. Pyrounakis and Nikolaidou(2009) made a relative study of 5 broadly used digital library systems. A comparative study of DSpace, Greenstone, Fedora, Keystone and EPrints was made on the basis of essential characteristics and system features. The study observed that the comparison can be used as a guideline to select the software suitable for an organization to host its digital collections. Fay (2010) presented a comparison of repository software carried out at London School of Economics and Political Science to build a digital library. The repository versions tested included DSpace 1.6, EPrints 3.2.0 and Fedora 3.3. The study found that repository software are not equal. Different software had relative strengths and weaknesses in different functional areas. The study also found that software selections depended a lot on the institutional requirements.

The literature survey on the area of open access, open source and open repositories revealed that there is a gap in the knowledge on the issue of adoption and use of various software for open repositories. This paper is intended to fill the gap.

**4. Objectives**

1. To list out software packages being used for open digital repositories in the world

2. To study the proportion of open source software and proprietary software being used for open digital repositories in the world.

3. To find out the most used repository software in the world.

4. To identify and compare the features of four major repository software.

**5. Methodology**

Both the Directory of Open Access Repositories (Open (DOAR) and the Registry of Open Access Repositories (ROAR) list open digital repositories in the world. Since both systems provide similar contents, the study is limited to software packages listed by ROAR (http://roar.eprints.org/) hosted at the University of Southampton, UK. The content analysis of WebPages of each software was performed to gather essential data.

As per ROAR, there are 27 software packages available for creating open access digital repositories. The Table I provide the names of the software and the website address of the study population.

|  |  |  |
| --- | --- | --- |
| Sl.No. | Name of the Software | Website Address |
|  | ARNO | <http://www.uba.uva.nl/projecten/object.cfm/1A103F4F-A900-4FCF-9BA16965AAE3D75E> |
|  | Digital Commons | http://www.bepress.com/ |
|  | CDS Invenio | <http://invenio-software.org/> |
|  | CONTENTdm | http://www.contentdm.org/ |
|  | DIGIBIB | http://www.digibis.com/en/software/digibib |
|  | DiVA(Academic Archive Online) | http://www.diva-portal.org/smash/aboutdiva.jsf |
|  | DigiTool | http://www.exlibrisgroup.com/category/DigiToolOverview |
|  | DoKS | http://doks.khk.be/ |
|  | DSpace | http://www.dspace.org/ |
|  | EDOC | http://alternativeto.net/software/edoc-organizer/ |
|  | EPrints | [http://www.eprints.org](http://www.eprints.org/) |
|  | ETD-db | <http://scholar.lib.vt.edu/ETD-db/index.shtml> |
|  | Equella | www.equella.com |
|  | Fedora | <http://www.fedora-commons.org/> |
|  | FEZ | <http://sourceforge.net/projects/fez> |
|  | Greenstone | http://www.greenstone.org/ |
|  | Intralibrary | www.intrallect.com |
|  | Keystone DLS | http://www.indexdata.dk/keystone/  (not being actively developed) |
|  | MyCoRe | http://www.mycore.de/ |
|  | Opus(Open publications system) | http://elib.uni-stuttgart.de/opus/browsen\_uebersicht.php?la=en |
|  | Open journal system (OJS) | https://pkp.sfu.ca/ojs/ |
|  | Open repository | http://www.openrepository.com/ |
|  | PMB (PhpMyBibli) | www.sigb.net/pmb/ |
|  | SciX | http://www.scix.net/ |
|  | SobekCMI | http://sobekrepository.org/ |
|  | WIKINDX | http://wikindx.sourceforge.net/ |
|  | Zentity | http://research.microsoft.com/en-us/projects/zentity/ |

Table 1 List of Repository software taken for Study

**6. Analysis**

The 27 software packages were examined for their status as open source or proprietary software, the development environment of the software, the number of adoption of particular software, and the features of top four software packages.

**6.1 Open Sources Vs Proprietary**

In the modern software development world, the two universally accepted and followed ways of software development are proprietary and open source (Isitan, 2011). Proprietary software is computer software that are exclusive property of their owners/creators and bear limits against uses, such as modification, sharing, studying, redistribution, or [reverse engineering](http://en.wikipedia.org/wiki/Reverse_engineering).  Generally, the [source code](http://en.wikipedia.org/wiki/Source_code) of proprietary software is closed. Proprietary software may also be called closed source software or commercial software. Open Source Software on the contrary, are owned by a community of users who belong to diverse locations in the world and provide access to source code for ensuring freedoms to run it, to study and change it, and to redistribute copies with or without changes. Currently, open source software products have started to become popular in the market as an alternative to traditional proprietary or closed source software (Sarrab and Rahman, 2014).

The analysis of the type of software being used for digital repositories revealed that open source software are the major players in the repository field. The Figure 1 portrays Open Source Vs Proprietary software distribution among digital repositories in the world.

Figure 1- Open Source Vs Proprietary Software use

Out of 27 software packages 23 (85%) belong to open source category. The share of proprietary software is 4(15%). Hence it can be concluded that open source software and the concept of open access go together.

**6.2. Leadership behind the repository Software development**

The leadership behind a software package is an important aspect on the reputation and maturity of the software. Hence the study examined the status of the developers of the repository software packages in the world. It was identified that half (51.8%) of the software packages were developed by universities/academic institutions around the world. 2 (7.4%) of them were developed by individuals and 40.7% of them by other firms/associations/public institutions. Figure 2 depicted the type and nature of software developers.

Fig. 2Leadership behind repository softwaredevelopment

**6.3 The Prominent Repository Software**

The study has examined the software that is being adopted by the majority of digital repositories in the world. It was found that DSpace software is top among the software being used for digital repositories. Out of 2680 repositories listed in the ROAR registry 1489 (55.5%) are using DSpace software. EPrints software has a share of 535(19.9 %) followed by Digital Commons 335(12.5%), OPUS 56(2.08%) and Fedora 43 (1.6 %). Figure 3 shows the major five repository software and their share of adoption.

Figure 3 Most prominent repository software

A further analysis of the five major repository software was attempted to provide details of the software. All the five software belong to open source category. Among them DSpace is the most widely used open source software for building digital repositories. It was developed by MIT Libraries and Hewlett Packard Lab in 2002 and is used by academic, non-profit, and commercial organizations. It is user friendly, easy to install and customizable to fit the needs of any organization.  DSpace preserves and provides open access to all types of digital content including text, images, moving images, mpegs and data sets.  It is being continuously updated and DSpace 4.2 is the latest release. DSpace 5.0 release has also been developed ready for testing. The documentation for the installation, customization and maintenance of DSpace is strong. It is optimized for Google Scholar. There are 1489 DSpace systems in the across the globe.

EPrints is the second major open source repository software. It is the first software platform for building high quality OAI-compliant repositories. EPrints was developed at the University of Southampton. It is the easiest and fastest way to set up repositories of open access research literature, scientific data, theses, reports and multimedia. EPrints 3 is the platform for a variety of open source repository solutions. EPrints enables time saving deposits and helps to import data from other repositories and services. It is optimized for Google Scholar and works with bibliography managers, desktop applications and new Web 2.0 services. EPrints software has 535 installations worldwide (<http://roar.eprints.org>).

Digital Commons is the software service developed by Berkeley Electronic Press, now simply called B**epress.** It is the third major repository software in the world and is being used by universities, colleges, law schools, and research centers. It has 335 installations worldwide and no installation in India. The software can be used to showcases the breadth of scholarship produced at an institution - everything from faculty papers, student scholarship, and annual reports to open-access journals, conference proceedings, and monographs. Scholarly material and special collections in Digital Commons repositories are highly discoverable in Google, Google Scholar, and other search engines. Additionally, articles in Digital Commons repositories are indexed in the [Digital Commons Network](http://network.bepress.com/), a free discovery tool for full text scholarly articles used by researchers worldwide.

OPUS—(Online Publications of the University of Stuttgart )—was developed in 1998 by the Stuttgart University Library and the Computing Center of the University of Stuttgart. OPUS is now developed by a consortium of German university partners in Berlin, Dresden, Saarbrücken, and Stuttgart. It is the fourth major repository software adopted in the world. It was designed to provide a platform for managing published and unpublished articles, theses and dissertations by the academic community of the university. The OPUS software is currently used by 48 German universities and 8 other institutions in Philippines, Poland, and United States etc. Main features for future development include digital signatures and multimedia documents.

Fedora, (Flexible Extensible Digital Object Repository Architecture) the fifth major repository software is mainly used in academic institutions (www.duraspace.org ). Fedora was originally developed by researchers at Cornell University as architecture for storing, managing, and accessing digital content in the form of digital objects. The Fedora team is now working on Fedora 4. Fedora provides a core repository service with web-based services .In addition, Fedora provides an array of supporting services and applications including search, OAI-PMH, messaging, administrative clients, etc. Fedora repository software enables long-term access to digital resources.  A key feature of Fedora is its flexibility and it can support all types of digital content. It is now widely used for digital collections, e-research, digital libraries, archives, digital preservation, institutional repositories, open access publishing, document management, digital asset management, and more. A list of the types of organizations in the Fedora user community includes Broadcasting and media, Consortia, Government agencies, Museums and cultural organizations, University libraries and archives and many more. Fedora has 54 installations worldwide and no installation was found in India. (<http://fedorarepository.org/>).

**6.4 Repository Software use in India**

The analysis of the adoption and use of repository packages in India revealed the following facts. Similar to the worldwide trend, Indian libraries and information centres prefer open source software for digital repositories. DSpace is the most preferred open source repository software in India. Around 140 Indian institutions use DSpace. (http://registry.duraspace.org/ registry/dspace). EPrints software is the second choice for repository software in India. It has 31 installations in the country. However, there are no instances for Fedora, Digital Commons and OPUS in India. The Figure 4 shows the distribution of software among the digital repositories in India.

Figure 4. Repository Software use in India

**6.5 Software Evaluation**

The survey of software being used for repository applications in the world brings several facts regarding the choice of software, dominance of open source software, and software development environment. Hence it is relevant to understand the features of some of the software that are prominent among digital repositories. The study has attempted to understand the features of four major repository software. The Table 2 shows the name of four software and their details.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Features** | **DSpace** | **EPrints** | **Fedora** | **Digital Commons** |
| 1 | License | Free | Free | Free | Commercial |
| 2 | Product type | Software | Software | Software | Hosted service |
| 3 | Update cost | With license | Free | Free | Free |
| 4 | Supported items | Documents, Images, Video, Audio, Learning objects | Documents, Images, Video, Audio, Learning objects | Documents, Images, Video, Audio, Learning objects | Stores and makes available for download any discrete file type |
| 5 | Metadata Formats | Dublin core,  Qualified DC  Variety of metadata formats | Dublin core,  BibTeX, EndNote, CSV, OpenURL, RDF, XML | Dublin core,  Qualified DC,  Any XML metadata format | Dublin core,  Qualified DC |
| 6 | User Interface | End user authentication, multi-language support | End user authentication, multi-language support | End user authentication, multi-language support | End user authentication, multi-language support |
| 7 | Format conversion to PDF | Current 3rd party | Future optional | Not applicable | Current Standard To pdf from  doc. .rtf |
| 8 | Advanced Searching | Field-specific, Boolean logic,  Sorting options | Field-specific,  Sorting options | Field-specific, Boolean logic,  Sorting options | Field-specific, Boolean logic,  Sorting options |
| 9 | Subject classes | Any administrator defined controlled vocabulary | LOC Classification | - | Custom vocabularies |
| 10 | Web 2.0 features | None.  Sharing only (Third party) | Tagging, comments (third party) | - | Bookmarks, reviews etc |
| 11 | Operating systems | Linux,  Windows, Solaris etc | Linux,  Windows  Solaris etc | Linux,  Windows, Solaris etc | Linux  - |
| 12 | Databases | Oracle, PostgreSQL | MySQL, Oracle, PostgreSQL,cloud | MySQL, Oracle, PostgreSQL,cloud | PostgreSQL |
| 13 | OAI-PMH | Current standard | Current standard | Current optional | Current standard |

The study of the four digital repository software show that all the major software share many common features. While the three major software solutions belong to open source category, Digital Commons is proprietary in nature. It is hosted repository software supported by Bepress. The institutions that adopt Digital Commons require an internet browser to access the service. A Digital Commons license includes setup, training, support, documentation, upgrades and hosting. The open source software solutions provide multiple options to customers. These software can be freely downloaded from respective websites or from sourceforge.net. It can be installed and customized either by the institutions themselves or by service providers. Moreover, DSpace and Fedora are now available as a hosted service by the Duraspace Foundation. Open access movement is an important area for Library and Information professionals as it accelerates the dissemination of more knowledge to more people. Software platforms fuel this movement in a number of ways.

References

1. Baruah, B. G. (2007). Role of Open Access, Open Standards and Open Sources in Libraries: A Study, available at: http://ir.inflibnet.ac.in/bitstream/1944/1365/1/43.pdf
2. [Budapest Open Access Initiative](http://en.wikipedia.org/wiki/Budapest_Open_Access_Initiative) in February 2002, available at: www.budapestopenaccessinitiative.org/read
3. Chan, L. (2004). Supporting and enhancing scholarship in the digital age: the role of open access institutional repository. *Canadian Journal of Communication*, *29*(3).
4. Fay, E. (2010). Repository software comparison: building digital library infrastructure at LSE. Ariadne, 64. Available at: http://www.ariadne.ac.uk/issue64/fay
5. Hanumappa, A., Dora, M., & Navik, V. (2014). Open source software solutions in Indian libraries. *Library Hi Tech*, *32*(3).
6. Harnad, S. (2001a). Why I think research access, impact and assessment are linked. *Times Higher Education Supplement*, *1487*, 16.
7. Harnad, S. (2001b). The self-archiving initiative. *Nature*, *410*(6832), 1024-1025.
8. Harnad, Stevan. (2003). Online archives for peer-reviewed journal publications. In J. Feather & P. Sturges (Eds.), International encyclopedia of library and information. London: Routledge. URL: [http://www.ecs.soton.ac.uk/%7Eharnad/Temp/archives.htm](http://www.ecs.soton.ac.uk/~harnad/temp/archives.htm)
9. Isitan, K. (2011). An Approach to Establish a Software Reliability Model for Different Free and Open Source Software Development Paradigms.Available at: https://tampub.uta.fi/handle/10024/82597
10. Jackson, R., & Richardson, M. (2014). Gold open access: the future of the academic journal?. *The Future of the Academic Journal*, 223.
11. Johnson, P. C. (2014). International Open Access Week at Small to Medium US Academic Libraries: The First Five Years. *The Journal of Academic Librarianship*. Available at: http://www.sciencedirect.com/science/article/pii/S0099133314001475
12. Meitei, L. S., & Devi, P. (2009). Open Source Initiative in Digital Preservation: The Need for an Open SourceDigital Repository and Preservation System. CALIBER 7TH, Pondicherry University, Puducherry, February 25-27.
13. Pappalardo, K. M., Fitzgerald, A. M., Fitzgerald, B. F., Kiel-Chisholm, S. D., O'Brien, D., & Austin, A. (2007). A guide to developing open access through your digital repository. <http://eprints.qut.edu.au/9671/1/9671.pdf>
14. Payne, A., & Singh, V. (2010). Open source software use in libraries. *Library Review*, *59*(9), 708-717.
15. Pyrounakis, G., & Nikolaidou, M. (2009). Comparing Open Source Digital Library Software. Available at: <http://galaxy.hua.gr/~mara/publications/ideaDL09a.pdf>
16. Rodriguez, J. E. (2014). Awareness and Attitudes about Open Access Publishing: A Glance at Generational Differences. *The Journal of Academic Librarianship*. Available at: <http://www.sciencedirect.com/science/article/pii/S0099133314001852>
17. Sarrab, M., & Rehman, O. M. H. (2014). Empirical study of open source software selection for adoption, based on software quality characteristics.*Advances in Engineering Software*, *69*, 1-11. Available at: http://www.sciencedirect.com/science/article/pii/S0965997813001798
18. Singh, S. K., Witt, M., & Dorothea, S. (2010, July). A Comparative Analysis of Institutional Repository Software. In *Fifth International Conference on Open Repositories. Madrid, Spain*. Available at: http://www.or10.es/Resources/documentos/ GSabstracts/AComparativeAnalysisInstitutionalRepository Software.pdf
19. Suber, P. (2012). Open Access. Boston: MIT press,

# Available at: http://cyber.law.harvard.edu/hoap/Open\_Access\_(the\_book) (2004).