



Using open source software for digital libraries

A case study of CUSAT

Surendran Cherukodan

*School of Engineering, Cochin University of Science and Technology,
Cochin, India*

G. Santhosh Kumar

*Department of Computer Science, Cochin University of Science and Technology,
Cochin, India*

S. Humayoon Kabir

*Department of Library and Information Science, University of Kerala,
Thiruvananthapuram, India*

Using open
source software

217

Received 12 August 2010

Revised 14 January 2011

18 July 2011

Accepted 31 July 2011

Abstract

Purpose – The purpose of this paper is to describe the design and development of a digital library at Cochin University of Science and Technology (CUSAT), India, using DSpace open source software. The study covers the structure, contents and usage of CUSAT digital library.

Design/methodology/approach – This paper examines the possibilities of applying open source in libraries. An evaluative approach is carried out to explore the features of the CUSAT digital library. The Google Analytics service is employed to measure the amount of use of digital library by users across the world.

Findings – CUSAT has successfully applied DSpace open source software for building a digital library. The digital library has had visits from 78 countries, with the major share from India. The distribution of documents in the digital library is uneven. Past exam question papers share the major part of the collection. The number of research papers, articles and rare documents is less.

Originality/value – The study is the first of its type that tries to understand digital library design and development using DSpace open source software in a university environment with a focus on the analysis of distribution of items and measuring the value by usage statistics employing the Google Analytics service. The digital library model can be useful for designing similar systems.

Keywords Open source software, Digital libraries, DSpace, India, Libraries

Paper type Case study

Introduction

Digital libraries (DL) are an important part of modern information management. Along with the development and extensive application of information technologies and networks, digital libraries are the booming development in the world (Zhou, 2005). DLs combine technology and information resources to allow remote access to distributed information resources, thus breaking down the physical barriers between resources to become, in effect, a networked multimedia information system. The Digital Library Federation (1999) defined DLs 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 over time of collections of digital works so that they readily and economically available for use by a defined community or set of communities.

The design, implementation and running of DLs are extensively practised by libraries of all types for collecting, archiving and distributing born-digital and digitised items of information. DLs help the preservation of the intellectual content produced and required by a particular community.

Scholarly and professional interest in digital libraries grew rapidly from the 1990s onwards, with initiatives of digitisation and digital libraries in India embarked on in the mid-1990s. Presently there are various digital library working models such as Digital Libraries of India (DLI), Vidyanidhi, Traditional Knowledge Digital Library (TKDL), Gyandoot and Samadhan kendras (Bhatt, 2008). However, the application of free/open source software (F/OSS) is new to Indian research libraries. The Registry of Open Access Repositories provides the list of repositories in the world. It shows that USA has 340 repositories, followed by the UK with 183; Japan has 88 repositories, while India has only 61.

This paper describes the design and development of a digital library using DSpace open source software in Cochin University of Science and Technology (CUSAT), India. The data for the study was gathered from the experience of the authors with the digital library for the last seven years in installing, customising, creating communities, sub-communities, and collections corresponding to various departments and submitting items to these collections. Data was collected using the Google Analytics service for usage statistics of the digital library.

Overview of open source software

Free/open source software (F/OSS) has its roots from near the beginning of computing and is typically free while providing users with source code that is usually shared via the internet and can be adjusted for the users' own needs (Baytiyeh and Pfaffman, 2010). Open source software differs from proprietary software in many respects. The major difference between the two is the freedom to modify the software. OSS tools can provide considerable cost savings over proprietary tools (Morrissey, 2010). The implementation cost for a typical DL would include the cost of an entry-level server and costs incurred for system support. Since the mid-1990s, there has been a surge of interest among academics and practitioners in OSS (Lee *et al.*, 2009) and today a wide range of F/OSS are being applied in all fields of human activity.

F/OSS offer attractions for libraries, as a majority of libraries around the world, especially in developing countries, cannot afford costly commercial software (Rafiq, 2009). According to Chudnov (1999), founder of the Open Source Systems for Libraries project, there are three factors pushing the use of F/OSS in libraries:

- (1) F/OSS licenses allow libraries to cut their software budget and use it for other issues needing more funds.
- (2) F/OSS products are not locked into a single vendor. Thus even if a library buys an open source system from one vendor, it might choose to buy technical support from another company or get it from in-house experts.
- (3) The entire library community might share the responsibility of solving information system accessibility issues.

DSpace software

There are a number of F/OSS systems for capturing, preserving and distributing digital content. DSpace, E-Prints, Fedora and Greenstone are the most commonly used F/OSS platforms for this purpose. Among the various F/OSS systems, Greenstone and DSpace are the most widely used software for digital libraries (Witten, 2005). DSpace was developed to be open source jointly by MIT Libraries and HP Labs in 2002. The system is designed to run on the UNIX platform and all original code is in the Java programming language. It uses a PostgreSQL relational database, Apache web server and a Tomcat servlet engine, the Jena RDF toolkit, OAICat from OCLC and several other useful software libraries. DSpace has implemented the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) for supporting interoperability with other DSpace adopters and digital repositories (Smith *et al.*, 2003). DSpace allows the capture items in any format – in text, video, audio, and data. It distributes it over the web and indexes work, so users can search and retrieve items. Moreover, it preserves digital works over the long term.

DSpace is the most popular among the digital library solutions available in the open source domain (Jose, 2007). Presently DSpace is used by over 1,150 organisations in a production or project environment (see www.dspacedev2.org). Since DSpace is a fairly powerful software (Biswas and Paul, 2010), it has seen more installations over the world. In studies that attempted a comparative evaluation of open source digital library packages, DSpace emerged as a good option, having the best search and browsing support as well as good support for metadata and providing more power to administrators to put restrictions at collection level (Kumar, 2009). Moreover, a study that compared the DL software of DSpace, Fedora, Greenstone, KStone and Eprints recommended DSpace as the most appropriate system for a university environment (Pyrounakis and Nikolaidou, 2009). Online community support is more active for DSpace, which is evident from the DSpace mailing list and the DSpace wiki.

These features have influenced Cochin University of Science and Technology in selecting DSpace software for its digital library. The option for F/OSS is based on several factors including cost, facility to customise and modify the software, online support, and the lack of a digital library model using proprietary software in India.

Cochin University of Science and Technology

Cochin University of Science and Technology (CUSAT) is one of the top universities in India. CUSAT is organised academically into nine faculties (i.e. Engineering, Environmental Studies, Humanities, Law, Marine Sciences, Medical Sciences and Technology, Science, Social Science and Technology). CUSAT has at present 29 departments of study and research offering graduate and postgraduate programmes across a wide spectrum of disciplines in frontier areas of diverse faculties. According to a study carried out by Prathap and Gupta (2009) to rank the research performance of 67 Indian engineering and technological institutes during 1999-2008 using data from the SCOPUS database, CUSAT ranked tenth with a total number of 1,625 research papers published during the period. Among the universities, CUSAT came up third on the list.

CUSAT Digital Library

Along with the wide variety of educational practices, F/OSS implementation in administrative and teaching sectors is a priority area in CUSAT. The library system in

CUSAT has realised the concept and value of F/OSS. The University library and department libraries use the Koha open source integrated library system for automation. Workshops were offered to library staff on Linux, Koha and DSpace for imparting necessary education and training on applying F/OSS in CUSAT.

There was need for a digital library in CUSAT to organise, preserve and distribute the large amount of knowledge in the form of journal articles, research reports, dissertations, theses, images, teaching materials and other documents produced by the university in digital and analogue formats. The CUSAT Digital Library (CDL) was established in 2003 to fulfil these requirements and DSpace open source software was used for the purpose. The Centre for Information Resources Management (CIRM) at CUSAT allotted a server for the CDL. DSpace was reinstalled in 2007 to upgrade to the new version of the software. Apart from offering essential teaching and learning materials online to the CUSAT community, the CDL provides open access to the intellectual output of the University. The CDL can be accessed over the internet through <http://dspace.cusat.ac.in>. Figure 1 shows a screenshot of CDL.

CDL – structure and policies

The structure of CDL is based on the default structure of DSpace involving Communities, Sub-Communities and Collections. A community represents a teaching department. There are 26 communities in the digital library. The various branches of a department, such as faculty, library, laboratory, etc., are brought under sub-communities. All items meant for a sub-community are ordered in separate collections. The CDL is controlled by an administrator who has powers to create, delete, edit or modify Communities, Sub Communities, Collections or Items. The collection development process is decentralised, where faculty members, librarians and students are permitted to add items to the Collection belonging to their Communities. Those who are authorised to upload items in a Collection are known as an “E-Person”, and their access to the system is controlled by a

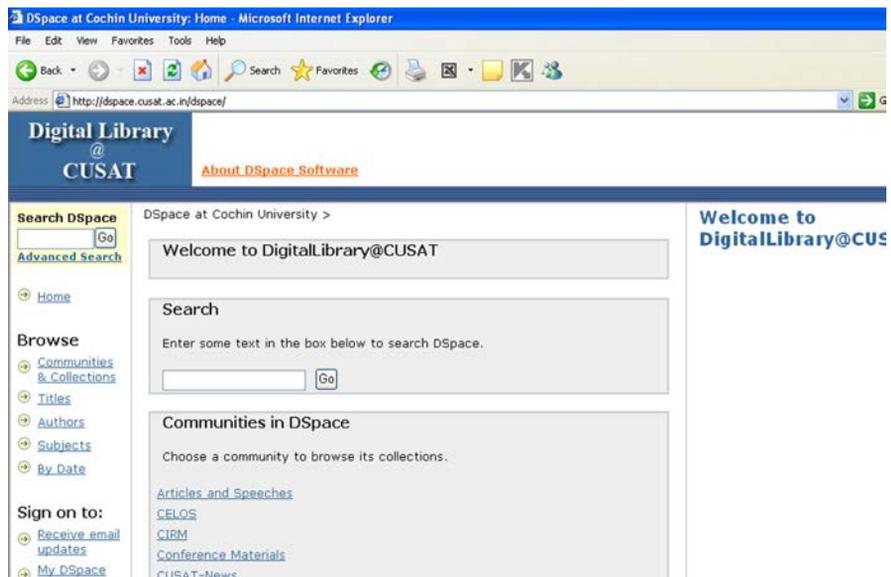


Figure 1.
CDL home page

user name and password. The CDL main server can be accessed over the web and the E-Person can add items from their location.

The respective teaching department determines the selection of material for the CDL. The content generated in the university is the basis of the CDL. The administrator has given necessary instructions to all E-Persons on the choice and uploading of items. PDF format is preferred over other document formats. The depositors have to agree that they are not depositing any copyrighted materials into the CDL. Even though there is no specific licence agreement, it is assumed that the materials deposited in the CDL are created by the CUSAT community and permission is given for their free use. The uploading process in CDL is composed of several steps including describe, upload, verify, licence and complete. There is provision for author, title, type of the item, language, identifiers, key words, abstract, file upload, verifying the submitted information and a non-exclusive distributive licence. DSpace supports the qualified Dublin Core metadata standard and a flexible framework to add user-defined metadata for localisation. While uploading an item into CDL each E-person is supposed to add certain mandatory fields apart from others as metadata. When an item is successfully uploaded, the system sends a message to the E-Person (Figure 2).

CDL – collection analysis

Analysis of the contents of a digital library is helpful for understanding the volume, type and distribution of documents in different categories. The data was collected by using “By Issue date” option available in the user interface of the CDL. Presently the CDL has around 2,312 items in it. Figure 3 shows the distribution of documents in the CDL. Out of the total, 1,875 (81 per cent) documents are past exam question papers, 162 (7 per cent) are seminar reports and 75 (3.24 per cent) are articles. The share of other documents such as presentations, news items, syllabi and journal contents pages, etc., is 147 (6.35 per cent).



Figure 2.
Submission approval message

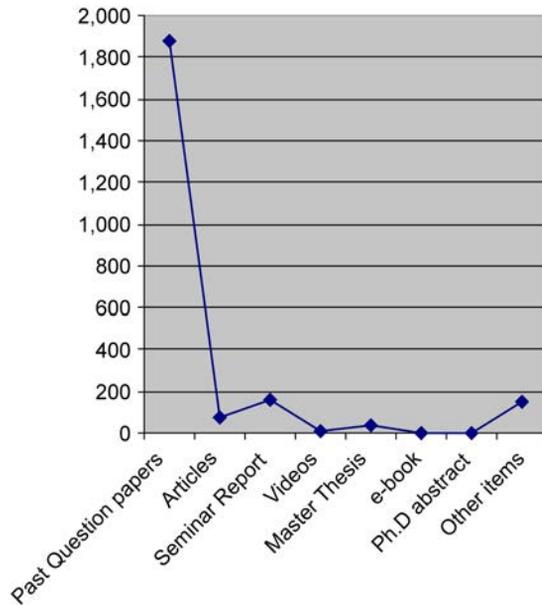


Figure 3.
Document distribution in CDL

CDL access statistics

The information on the use of a DL is an important element in measuring its value. A web-based DL will be visited by people from all across the world. The Google Analytics service can be employed as an important tool for obtaining data on the usage of DLs. The access statistics of CDL from January 2009 to September 2009 were collected using this service. The data is presented in Figure 4. In nine months, around 10,346 people visited the digital library, making 23,722 page visits. The country-wise

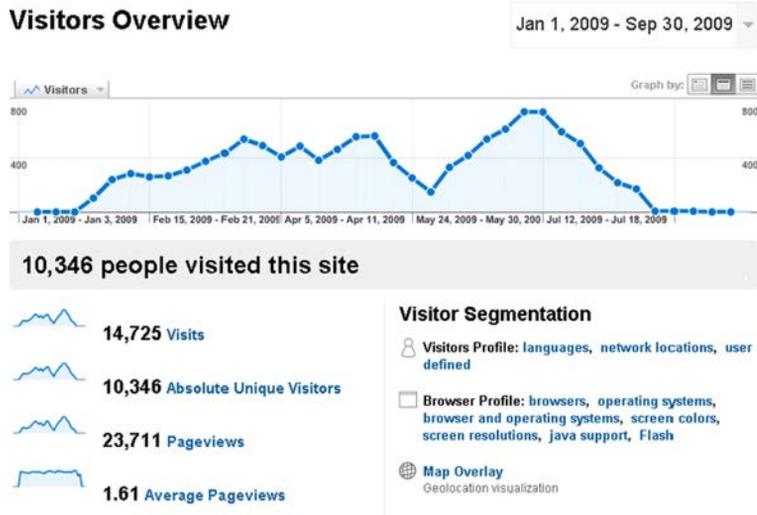


Figure 4.
Statistics on number of visitors

distribution of the usage of CDL is presented in Figure 5. It demonstrates that the digital library was accessed from all over the world – in fact from no fewer than 78 countries.

Out of the total page visits, 14136 (59 per cent) were from India with 142 page visits made from the USA. The Indian city-wise distribution of usage of CDL is presented in Figure 6 – visits were made from all the major cities in India. The home city, Cochin, recorded 8,890 (62 per cent) visits, followed by 1,285 (9 per cent) from Trivandrum, the capital city of Kerala. Cities in Kerala had a total share of 10,784 (76 per cent) visits.

The usage statistics of CDL shows that it contains information sought by people across the world. The majority of visits were recorded from the state of Kerala. However, the analysis is limited to page visits only, and we need further analysis to determine preferred collection that received most visits.

Conclusion

The CDL is an achievement for the academic community of CUSAT for storing relevant documents in an organised, secure, and searchable archive and preserving it for long-term use. From the data on usage statistics, it is clear that CDL is also providing a service to users outside CUSAT. The contents of CDL are getting top search results in Google and other search engines, leading to increased accessibility to

	Detail Level: Country/Territory ▾	Visits ↓
1.	India	14,135
2.	United States	142
3.	United Arab Emirates	79
4.	Saudi Arabia	36
5.	Qatar	36
6.	Brazil	18
7.	Singapore	15
8.	United Kingdom	15
9.	Spain	13
10.	Kuwait	12
11.	Malaysia	11
12.	Turkey	11
13.	Australia	10
14.	Hong Kong	10

Figure 5.
Statistics on number of
visitors by country

	Detail Level: <input type="text" value="City"/>	Visits ↓
1.	Cochin	8,890
2.	Trivandrum	1,285
3.	Bangalore	641
4.	(not set)	448
5.	Chennai	340
6.	Kottayam	319
7.	Kozhikode	290
8.	Delhi	277
9.	Coimbatore	225
10.	Patna	221
11.	Mumbai	158
12.	Pune	103

Figure 6.
Statistics on number of
visitors by Indian city

the documents. The use of F/OSS for the design and development of DL is the first instance of its kind in the state of Kerala, where seven other universities exist. The CDL can be viewed as a model based on F/OSS without any grant from a parent institution or other agency. The role and importance of CDL can be expanded by inviting more attention from the parent organisation towards creating digital content and archiving it in CDL to provide open access to the ideas and knowledge generated by CUSAT.

References

- Baytiyeh, H. and Pfaffman, J. (2010), "Open source software: a community of altruists", *Computers in Human Behavior*, Vol. 26 No. 6, pp. 1345-54.
- Bhatt, R.K. (2008), "March towards digitization of information resources in India: issues and initiatives", *World Digital Libraries*, Vol. 1 No. 2, pp. 147-64.
- Biswas, G. and Paul, D. (2010), "An evaluative study on the open source digital library softwares for institutional repository: special reference to DSpace and Greenstone Digital Library", *International Journal of Library and Information Science*, Vol. 2 No. 1, available at: www.academicjournals.org/ijlis/PDF/pdf2010/Feb/Biswas%20and%20Paul.pdf (accessed 3 August 2010).
- Chudnov, D. (1999), "Open source library systems: getting started", available at: www.oss4lib.org/readings/oss4lib-gettingstarted.php (accessed 23 July 2003).
- Digital Library Foundation (1999), "A working definition of digital library", available at: www.diglib.org/about/dldefinition.htm (accessed 3 August 2010).

-
- Jose, S. (2007), "Adoption of open source digital library software packages: a survey", paper submitted to the Convention on Automation of Libraries in Education and Research Institutions (CALIBER 2007), available at: <http://eprints.rclis.org/8976/1/Sanjojose.pdf>
- Kumar, V. (2009), "Comparative evaluation of open source digital library packages", available at: https://drtc.isibang.ac.in/bitstream/handle/1849/441/comparative_evaluation_DL_vinit.pdf?sequence=1
- Lee, S.-Y.T., Kim, H.-W. and Gupta, S. (2009), "Measuring open source software success", *Omega*, Vol. 37 No. 2, pp. 426-38.
- Morrissey, S. (2010), "The economy of free and open source software in the preservation of digital artefacts", *Library Hi Tech*, Vol. 28 No. 2, pp. 211-23.
- Prathap, G. and Gupta, B.M. (2009), "Ranking of Indian engineering and technological institutes for their research performance during 1999-2008", *Current Science*, Vol. 97 No. 3, pp. 304-6.
- Pyrounakis, G. and Nikolaidou, M. (2009), "Comparing open source digital library software", *Collection of Handbook of Research on Digital Libraries*, pp. 51-60, available at: www.dit.hua.gr/~mara/publications/ideaDL09a.pdf (accessed 17 July 2011).
- Rafiq, M. (2009), "LIS community's perceptions towards open source software adoption in libraries", *The International Information & Library Review*, Vol. 41 No. 3, pp. 137-45.
- Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G. and Stuve, D. (2003), "DSpace: an open source dynamic digital repository", *D-Lib Magazine*, Vol. 9 No. 1, available at: www.dlib.org/dlib/january03/smith/01smith.html (accessed 11 January 2011).
- Witten, I. (2005), "A bridge between Greenstone and DSpace", *D-Lib Magazine*, Vol. 11 No. 9, available at: www.dlib.org/dlib/september05/witten/09witten.html (accessed 12 November 2009).
- Zhou, Q. (2005), "The development of digital libraries in China and the shaping of digital librarians", *The Electronic Library*, Vol. 23 No. 4, pp. 433-41.

About the authors

Surendran Cherukodan has been working as Junior Librarian in the School of Engineering, Cochin University of Science Technology since 2000. He acquired a Master's Degree in Library and Information Science from the University of Calicut, Kerala, India. He has passed the UGC Test for Lectureship and Junior Research Fellowship. He has published several papers in national and international seminars. His research interests include digital library, application of open source software in libraries, etc. He is a member of Kerala Library Association. Surendran Cherukodan is the corresponding author and can be contacted at: csura@cusat.ac.in

G. Santhosh Kumar has been working as Assistant Professor in the Department of Computer Science at Cochin University of Science and Technology since 2001. He acquired a Master's Degree in Physics and an MTech degree in Computer and Information Science from Cochin University. His research interests are in networked embedded systems, software architectures, e-learning and free/open source software systems. He is a Professional Member of ACM and IEEE.

S. Humayoon Kabir started his career as a librarian in Cochin University and has been a Library and Information Science teacher at Mahatma Gandhi University and Pondicherry University. He is currently working as a Reader in the Department of Library and Information Science, University of Kerala where he is an Associate Professor. He has authored several papers and is one of the Editors of *KELPRO Bulletin*, published from Kerala. He has a PhD in Library and Information Science and he is a research guide. His research interests include open access, open source, digital library software, etc.