

# Open Source Initiative in Digital Preservation: The Need for an Open Source Digital Repository and Preservation System

L Shanta Meitei

Purnima Devi

## Abstract

*The paper discusses the Open Source Software in digital preservation for digital repository and preservation system. The paper highlights the need and features of Open Source Software in Digital Preservation with the successful adoption of open source application for library and information management system in this global digital information environment. The paper also discusses some of the important Initiatives of Digital Preservation and Repository System using Open Source Software in India.*

**Keywords:** Open Source, Digital Preservation, DSpace, Greenstone Digital Library, Eprints, Fedora, Koha

## 1. Introduction

Open source denotes the principles of promoting open access to a good's production or design process and the product itself. It is mostly used in the context of computer software, meaning that the knowledge assembled in software programs and operating systems is available. Open source is often mentioned in the digital preservation context for open standards play an important role here. File format specifications and document formats can be also open source, and related to open standards. Together they satisfy quite a number of preservation requirements but for a number of reasons they cannot be proclaimed as a one-fits-all solution for digital preservation [1].

The direction digital preservation must take is known, but the details and complex relationships which must be resolved are still being pursued. And it is a matter of record that the array of resources

and expertise are making inroads into the solution of all this digital complexity. Almost all of the digital preservation exist within established institutions and have dedicated staff with technological expertise up to the task of wrestling with aspects of these issues. These include, for example, National Libraries, National Archives, Institutional Repositories, Digital Libraries, Universities and other places of learning and research, and media and cultural museums and archives.

So the point of this paper, and UNESCO MoW report, "Towards an Open Source Repository and Preservation

System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development" on which it is based, is to imagine a scenario where there is a need to preserve a collection of simple digital objects, but where a digital preservation infrastructure has not yet been



developed. In other words; to develop a sustainable preservation standard digital management and storage system for a collecting institution that doesn't happen to be one of the world leaders in digital preservation [2].

## 2. What is Open Source Software?

Open-source software (OSS) is software for which the source code is freely available for anyone to see and manipulate. There are various licensing models to which the OSS label has been applied, but the basic idea is that the software's "license may not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs" and the working software must either be distributed along with its source code or have a "well-publicized means of downloading the source code, without charge, via the Internet." That is, anyone can access and manipulate the code that was used to write a program, as long as anything that person comes up with using that code is also offered to the public as OSS. This allows those who use the software to contribute to its further development, fix bugs and tinker with it as they please. This is contrasted with proprietary software, which is distributed as compiled object code or machine code, leaving the source code solely under the control of the individual software vendor [3].

## 3. Open Source and Digital Preservation

Open source is not necessarily confined to software. Open standards, for example, can also be regarded as open source, in the sense that they are freely available and open to the public. Assets conforming to open standards are more qualified for being preserved over a long period of time inasmuch as

they give access to the file format, making it easier to develop a tool which migrates this format should it become obsolete. In addition to this, many of the file format specifications like the OpenOffice.org spreadsheet and document formats are themselves open source. However, proprietary solutions can also provide satisfying results, having the advantages of, continuing and guaranteed customer support. Open standards for files to be preserved, and also the implementation of the preservation software and its parts under an open source licence brings advantages. Other institutions can use components developed using open software and adapt them to their needs. Furthermore, especially with respect to trust, open source software is much easier to evaluate than proprietary software [4].

## 4. Need & Features of Open Source Software in Digital Preservation

Open source software development model gives organizations a new option for acquiring and implementing systems, as well as new opportunities for participating in digital preservation projects. Five library related open source software are described to illustrate the need and practice in digital preservation system [5] with their comparative study as shown in Table – 1.

### 4.1. Koha

Koha is the first open-source Integrated Library System (ILS). In use worldwide, its development is steered by a growing community of libraries collaborating to achieve their technology goals. Koha's impressive feature set continues to evolve and expand to meet the needs of its user base. Koha is distributed under the open-source General Public License (GPL). Koha includes modules for

circulation, cataloging, acquisitions, serials, reserves, patron management, branch relationships, and more[6].

#### **4.2. Greenstone**

Greenstone is an open source suite of software issued under the terms of the GNU General Public License. It is a user-friendly, multilingual, multi-platform package for assembling electronic documents into digital collections

and for publishing these collections on the Web or on CD-ROM. It accepts documents in a wide range of proprietary and standard formats, supports numerous standards for document and metadata exchange, including compliance with the OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) and Z39.50 information retrieval standards, and readily converts bibliographic databases created under UNESCO's CDS/ISIS package into digital libraries, including the full texts of the related documents if available. Greenstone's flexibility, robustness, ease of use, and free availability make it a particularly useful resource for the development of a wide range of DL applications and for the training of librarians and information specialists in DL concepts [7].

#### **4.3. Eprints**

EPrints is free software developed by the University of Southampton, England. ePrints@IISc repository collects, preserves and disseminates in digital format the research output created by the IISc research community. It enables the Institute community to deposit their preprints, postprints and other scholarly publications using a web interface, and organizes

these publications for easy retrieval. While eprints@IISc can be accessed by anybody, submission of documents to this repository is limited to the IISc research community only. ePrints@IISc repository is running on EPrints open archive software, a freely distributable archive system available from eprints.org. ePrints@IISc complies with the Open Archives Initiative (OAI) framework allowing publications to be easily indexed by web search engines and other indexing services [8].

#### **4.4. DSpace**

Developed jointly by MIT Libraries and Hewlett-Packard (HP), DSpace is now freely available to research institutions worldwide as an open source system that can be customized and expanded [9]. DSpace is a digital asset management system. It helps create, index and retrieve various forms digital content. Dspace is adaptable to different community needs. Interoperability between systems is built-in and it adheres to international standards for metadata format [10].

#### **4.5. Fedora**

Fedora is a center for innovation in free and open source software, and creates a community where developers and open source enthusiasts come together to advance free and open source software. The Fedora community contributes everything it builds back to the free and open source world and continues to make advances of significance to the broader community. Fedora is a Linux based operating system that provides users with access to the latest free and open source software, in a stable, secure and easy to manage form [11].

	<b>GSDL</b>	<b>Eprints-II</b>	<b>DSpace</b>	<b>Fedora</b>	<b>Koha</b>
<b>Creator</b> & University	University of Waikato Ltd., New Zealand	University of Southampton	MIT libraries & Hewlett-	Cornell University & of Virginia	Katipo Communications Packard
<b>Open Source and Free</b>	Yes	Yes	Yes	Yes	Yes
<b>Operating System</b>	Unices, Windows	Unices	Unices	Unices, Windows	Linux, Windows
<b>Web-server</b>	Apache/ IIS	Apache 1.3	Apache 1.3/2.0 and/or Tomcat	Tomcat 1.4	Apache (2.0 is preferred)
<b>Language</b>	Perl	Mod-Perl 1.0	Java 1.3, JSP	J2SDK v.1.4	Perl
<b>Database</b>	Its own	MySQL	PostgreSQL 7.3	McKoi v.0.94 (uses by default MySQL//Oracle 9i (optional))	MySQL
<b>Resource Identifier</b>	No	OAI Identifiers (similar to URNs)	CNRI Handles	Uses own persistent	
<b>Dublin Core</b>	Dublin Core	Dublin Core	Qualified Core	Dublin Core Yes	Dublin Core
<b>METS</b>	No	No	NoTo be implemented in next Version 1.2		

Note: These features may change as newer versions of the software are made available

(Source of the Table: Madalli, Devika P. A Digital Library of Library and Information Science using Dspace. Statistical Institute, Bangalore)

### 5. Initiatives of Digital Preservation System and Digital Repositories in India

A digital preservation is a digital archive of the intellectual output of an organization/institution. It makes the quality and breath of scholarship produces at the organisation accessible to others world wide over the Internet. It is a set of services that a University/Organization offers to the

members of its community for the management and dissemination of digital material created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of the digital materials including long term preservation. An effective digital preservation and institutional repository of necessity represents collaboration among libraries, information technologies, archives and record managers, faculty and University administrators and policy makers [12].

With the emergence of successful digital library and preservation projects in more developed countries, the public institutions in the region opted for long-

term preservation of this wealth of knowledge through digitization projects and digital preservation initiatives. Diverse multi-cultural and multilingual contents are now being documented, preserved with the adoption of Open Source Software System and made available through the internationally acclaimed Digital Preservation and Repository initiatives such as [13]:

### 5.1. National Level Digital Preservations/Repositories:

- ◆ Catalysis Database [www.eprints.iitm.ac.in](http://www.eprints.iitm.ac.in)
- ◆ Software Used: EPrints
- ◆ Librarians' Digital Library (LDL) <https://drtc.isibang.ac.in/>
- ◆ Software Used: DSpace
- ◆ OpenMED@NIC <http://openmed.nic.in/>
- ◆ Software Used: EPrints

### 5.2. Institutional Repositories:

- ◆ Digital Archive of National Institute of Technology Rourkela <http://dspace.nitrkl.ac.in/dspace/>
- ◆ Software Used: DSpace
- ◆ Electronic Theses and Dissertations of Indian Institute of Science (ETD@IISc) <http://etd.ncsi.iisc.ernet.in>
- ◆ Software Used: DSpace
- ◆ Open Access Repository of IISc Research Publications (ePrints@IISc) <http://eprints.iisc.ernet.in/>
- ◆ Software Used: EPrints
- ◆ IDRC Digital Library <http://idl-bnc.idrc.ca/>
- ◆ Software Used: DSpace
- ◆ Digital Repository of IIT Bombay <http://dspace.library.iitb.ac.in/dspace/>
- ◆ Software Used: DSpace
- ◆ DSpace at National Centre for Radio Astrophysics <http://ncralib.ncra.tifr.res.in:8080/dspace/>

- ◆ Software Used: DSpace
- ◆ DSpace@IIMK <http://dspace.iimk.ac.in/>
- ◆ Software Used: DSpace
- ◆ DSpace at National Chemical Laboratory <http://dspace.ncl.res.in/dspace/>
- ◆ Software Used: DSpace
- ◆ DSpace@INFLIBNET <http://dspace.inflibnet.ac.in/>
- ◆ Software Used: DSpace
- ◆ University of Delhi EPrint Archive <http://eprints.du.ac.in/>
- ◆ Software Used: EPrints
- ◆ Raman Research Institute Digital Repository <http://dspace.rri.res.in:8080/dspace/>
- ◆ Software Used: DSpace
- ◆ One World South Asia Open Archive Initiative [http://open.ekduniya.net/SoftwareUsed: EPrints](http://open.ekduniya.net/SoftwareUsed:EPrints)

### 5.3. Digital Library:

- ◆ Archives of Indian Labour: Integrated Labour History Research Programme [www.indialabourarchives.org](http://www.indialabourarchives.org)
- ◆ Software Used: Greenstone Digital Library Software
- ◆ India Education Digital Library [www.edudl.gov.in](http://www.edudl.gov.in)
- ◆ Software Used: Greenstone Digital Library Software
- ◆ Vidyaidhi [www.vidyaidhi.org.in](http://www.vidyaidhi.org.in)
- ◆ Software Used: DSpace

## 6. Conclusion & Suggestion

It is largely achievable in a country where policy frameworks, institutional frameworks, information infrastructure, trained manpower, and financial resources are adequately available. The effect of focused capacity building programmers in the areas

of digital preservation, digital libraries and application of open source software is encouraging in country like India, where significant proliferation of digital preservations and digital repositories initiatives have been achieved in the last decade. A number of workshops and training events were organized in India during this period, where a few thousand libraries and computer professionals received training in open source software for building open access repositories and digital preservation initiatives. Library schools in India have since included open source digital archiving/preservation software in their curricula. Several national and international conferences, seminars, and symposia were also organized in India, where library professionals discussed methods and techniques of digitization, digital library development, institutional & digital repository development and digital preservation.

Therefore, (i) Governments should encourage providing adequate open source access through various communication resources, notably the Internet, to public official information. Establishing legislation on digital access to information and the digital preservation of public data, notably in the area of the new technologies, is encouraged; (ii) Develop policy guidelines for the development and promotion of digital preservation system as an important international instruments promoting public access to information; (iii) Encourage initiatives to facilitate open access and digital preservation including journals and books, and archives for scientific information; and (iv) Promote research and development of digital preservation initiatives projects and digital repositories with open source software and ICTs for all, including

disadvantaged, marginalized and vulnerable groups.

### References

1. **Neumayer, Robert.** Open Source in Digital Preservation. Available at [http://www.digitalpreservationeurope.eu/publications/briefs/open\\_source.pdf](http://www.digitalpreservationeurope.eu/publications/briefs/open_source.pdf) (Accessed on 05/01/2009).
2. **Bradley, Kevin.** Digital Preservation: the need for an open source digital archival and preservation system for small to medium sized collections. Available at [http://www.amw.org.au/mow2008/mow/speakerpapers/Bradley paper.pdf](http://www.amw.org.au/mow2008/mow/speakerpapers/Bradley%20paper.pdf) (Accessed on 07/01/2009).
3. **Lee, Cal.** Open-Source Software: A Promising Piece of the Digital Preservation Puzzle. Available at [http://ils.unc.edu/Callee/oss\\_preservation.htm](http://ils.unc.edu/Callee/oss_preservation.htm) (Accessed on 05/01/2009).
4. **Neumayer, Robert.** Open Source in Digital Preservation. Available at [http://www.digitalpreservationeurope.eu/publications/briefs/open\\_source.pdf](http://www.digitalpreservationeurope.eu/publications/briefs/open_source.pdf) (Accessed on 05/01/2009).
5. **Chawner, Brenda.** Free/Open Source Software: New Opportunities, New Challenges. Available at <http://www.vala.org.au/vala2004/2004pdfs/33Chawn.PDF> (Accessed on 07/01/2009).
6. **Koha** : The First Open Source ILS. Available at <http://www.koha.org/about-koha/> (Accessed on 13/01/2009).
7. About Greenstone. Available at <http://www.sagreenstone.unam.na/aboutgsdl.html> (Accessed on 09/01/2009).
8. About ePrints@IISc. Available at <http://eprints.iisc.ernet.in/information.html> (Accessed on 13/01/2009).

**9.** About DSpace. Available at <http://dspace.udel.edu/AboutDSpace.html> (Accessed on 13/01/2009).

**10. Madalli, Devika P.** A Digital Library of Library and Information Science using Dspace. Documentation Research and Training Centre, Indian Statistical Institute, Bangalore.

**11.** Available at <http://fedoraproject.org/wiki/Overview> (Accessed on 13/01/2009).

**12. Patel, Yatrik, Vijayakumar, J K and Murthy, T A V.** Institutional Digital Repositories/e-Archives: INFLIBNET initiatives in India. Published in Digital Libraries in Knowledge Management: Proceedings of the 7<sup>th</sup> MANLIBNET Annual National Convention. pp. 312-318, edited by M G Sreekumar [et al]. New Delhi: Ess Ess, 2006.

**13. Das, Anup Kumar.** Open Access to Knowledge and Information. Published in Open Access to Knowledge and Information in South Asia: Scholarly Literature and Digital Library Initiatives-The South Asian Scenario. Edited by Bimal Kanti Sen and Jocelyne Josiah. New Delhi: UNESCO, 2008.

#### **About Authors**

**Dr. Th. Purnima Devi,** Reader, Department of Library and Information Science, Manipur University.

E-mail: [puni\\_thoidingjam@yahoo.co.in](mailto:puni_thoidingjam@yahoo.co.in)

**Dr. Lairenlakpam Shanta Meitei,** Library Assistant, College of Agricultural Engineering & Post Harvest Technology (Central Agricultural University), Ranipool, Gangtok, Sikkim.

E-mail: [shanta\\_lairen@yahoo.co.in](mailto:shanta_lairen@yahoo.co.in)