Reshaping the World of Scholarly Communication : Open Access and the Free Online Scholarship Movement

Usha Mujoo Munshi

1. Introduction

The proliferation of scholarly materials is acknowledged by all stakeholders. The tremendous growth in the size is coupled with cost escalation of the primary journal literature. The rise in size and price has subsequently resulted in the current scholarly communication system becoming more and more unaffordable and has forced academic libraries over the last several decades to cancel existing serial titles and buy fewer books. The situation attributes that the crisis in the scholarly communication system not only intimidates the welfare of libraries, but also it perils the academic community's aptitude to do world-class research.

The emerging technologies have, however, shown us ways through obtainable tools, necessary to change the current scholarly communication system and promote open access to scholarly resources.

The digital libraries, particularly the digital preservation have become a much-debated topic over the years. Increasing focus on content as institutional asset, where substantial proportion of this content is now born-digital or re-born digital. This digital content creation has further got buttressed with wide uptake of open source software (OSS) such as Dspace and eprints.org and the like.

The present document keeping the above in view is going to focus on various facets of open access movement. It may be equally important and appropriate to include the requisite OSS that facilitates the creation of digital open access content and as such, example of DSpace is being included here. But before addressing open access let us first have a look at open access movement and some endeavours to reform the scholarly communication system.

2. Open Access Scholarship Movement

Open access scholarship movement is gathering strength worldwide. The recent statements & declarations are promoting establishment of institutional research repositories

and mandating deposit of institution's intellectual output . The statements and declarations like Budapest Open Access Initiative (http://www.soros.org/openaccess/read.shtml), Berlin Declaration on Open Access to Knowledge in the Sciences & Humanities (http://www.zim.mpg.de/openaccessberlin/berlindeclaration.html), Bethesda Statement on Open Access (http://www.earlham.edu/~peters/fos/bethesda.htm) , NEAR [Network For Education And Academic Rights] (http://www.nearinternational.org), OECD Final Communique (http://www.oecd.org/document/47/0,2340,en_2649_201185_34668207_1_1_1_1,00.html) , Tempe Principles (http://www.lib.ncsu.edu/scc/events/tempe.pdf) , Washington D.C. Principles for Free Access to Science (http://www.dcprinciples.org/), Wellcome Trust Position Statement on Open Access Publishing (OA and Unrestricted access to Published Research) (http://www.wellcome.ac.uk/doc_WTD002766.html) , World Summit on the Information Society - Declaration of Principles and Plan of Action are some landmarks in the history of promoting open access culture and many more. You may however have the Timeline of the Open Access (OA) Movement, written and revised by Peter Suber a crusader of OA movement available at http://www.earlham.edu/~peters/fos/timeline.htm

2.1. Organizations, Projects, Campaigns, and Initiatives to Reform the Scholarly Communication System

Several endeavours by many organizations, individuals, groups and the like have been made and are being made to reform the scholarly communication. Some such endeavours are: The Alliance for Taxpayer Access (http://www.taxpayeraccess.org/)- an "alliance of organizations representing varied groups that support open public access to taxpayer-funded research; American Council of Learned Societies (ACLS) (http://www.acls.org/) - with main mission to "advance humanistic studies in all fields of learning in the humanities and the social sciences and to maintain and strengthen relations among the national societies devoted to such studies; ARROW—Australian Research Repositories Online to the World. (http://arrow.edu.au/) - the project aims to identify and test software or solutions to support best practice institutional digital repositories; ArXiv (http://arxiv.org/)- an e-print service in the fields of physics, mathematics, non-linear science, computer science, and quantitative biology; Author's

Addendum "(http://www.arl.org/sparc/author/addendum.html) - a form author uses to amend the document that the publisher asks author to sign." "By using this form author "retains the right to make article available in a non-commercial open digital archive on the Web.; BioMed Central (http://www.biomedcentral.com/)- an independent publishing house committed to providing immediate open access to peer-reviewed biomedical research; Canada's International Development Research Center (IDRC) (http://www.idrc.ca/en/ev-92447-201-1-DO_TOPIC.html) - establishes an OA depository; CLOCKSS Project (http://www.lockss.org/clockss)- serves as a failsafe repository for published scholarly content; Create Change (http://www.createchange.org/librarians/intro/aboutcc.html) - has as its core goal to make scholarly research as accessible as possible to scholars all over the world, to their students, and to others who might derive value from it; Creative Commons (http://creativecommons.org/) - a nonprofit organization that offers flexible copyright licenses for creative works; Daedalus (http://www.lib.gla.ac.uk/daedalus/index.html) —Data-providers for Academic E-content and the Disclosure of Assets for Learning, Understanding and Scholarship; ePrints UK Project (http://www.rdn.ac.uk/projects/eprints-uk/); eScholarship repository program (http://repositories.cdlib.org/escholarship), (http://www.cdlib.org/programs/escholarship.html) - facilitates innovation and supports experimentation in the production and dissemination of scholarship. And to improve all aspects of scholarly communication, HighWire Press (http://highwire.stanford.edu/) hosts the largest repository of free, full-text, peer-reviewed content; Information Access Alliance (http://www.informationaccess.org/) - believes that "access to a broad array of research information is critical to the health and wealth of society; International Consortium for the Advancement of Academic Publication (ICAAP) (http://www.icaap.org/)- "focused on creating technologies to facilitate sophisticated delivery of educational content; Internet Archive (http://www.archive.org/) - building a digital library of Internet sites and other cultural artifacts in digital form; OAI-PMH, the Open Archive Initiative's Protocol for Metadata Harvesting, (http://www.openarchives.org/OAI/openarchivesprotocol.html); Open Access Now campaigns "for freedom of research information." (http://www.biomedcentral.com/openaccess/links/); Open Access Working Group (http://www.arl.org/sparc/oa/oawg.html); Open Archives Initiative (OAI) http://www.openarchives.org/ "develops and promotes

interoperable standards that aim to facilitate the efficient dissemination of content; Open Content Alliance (http://www.opencontentalliance.org/; http://www.opencontentalliance.org/faq.html) - represents the collaborative efforts of a group of, multifaceted organizations from around the world that will help build a permanent archive of multilingual digitized text and multimedia content; Project DARE (Digital Academic Repositories)/ DAREnet (http://www.darenet.nl/en/page/language.view/home), (http://www.darenet.nl/nl/page/language.view/home) - gives digital access to academic research output in the Netherlands; "Project DARE launches NARCIS (National Academic Research and Collaborations Information System) (http://www.narcis.info/narcis/?language=en) - as a gateway to Dutch scientific research; and DARLIN (Dutch Archive for Library and Information Science), (http://www.nvb-darlin.nl/en/) - a new OA repository for Dutch publications in library and information science; Project RoMEO — Rights Metadata for Open Archiving —(http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html); Project SHERPA (http://www.sherpa.ac.uk/)- deals with investigating issues in the future of scholarly communication and publishing, with special reference to developing open-access institutional repositories in various research universities.; Public Knowledge (http://www.publicknowledge.org/issues/openaccess)- "works for open access to (1) taxpayer-funded research and (2) research that scientists and scholars consent to publish without payment; Public Knowledge Project (http://pkp.ubc.ca/) - a federally funded research initiative that seeks to improve the scholarly and public quality of academic research through innovative online environments; Public Library of Science (http://www.plos.org/about/index.html), (http://www.publiclibraryofscience.org/) - a nonprofit organization of scientists and physicians committed to making the world's scientific and medical literature a public resource; PubMed Central (http://www.pubmedcentral.gov/about/openaccess.html) - a digital archive of life sciences journal literature at the U.S. National Institutes of Health (NIH); SciELO—Scientific Electronic Library Online (http://www.scielo.org/index.php?lang=en); Scholarly Publishing and Academic Resources Coalition (SPARC) (http://www.arl.org/sparc/home/index.asp) - an alliance of universities, research libraries, and organizations and serves as a catalyst for action; TARDIS-(http://tardis.eprints.org/) -Targeting Academic Research for Deposit and Disclosure.

2.2. OA Corollary

While on one hand, archiving in open access repositories is information for researchers, as their research output gets widely disseminated, thereby leading to increase visibility, which ensures (generally) increased visibility, catching more citations that lead to high research impact? Besides preservation would also enable researchers controlling or monitoring of their own publications. On the other hand, it is equally important for institutions. For institutions pooling the organization's intellectual capital, having one stop source / point for the research output of an institution, providing scope for introspection / strategies / action plan, facilitating generation of reports and long term preservation are some of the advantages of digital archiving of the institution's intellectual output. However above all such institutional digital archives is a no doubt facilitating enhanced access to scholarly communications at low cost or no cost. Hence their generation and increased population is of utmost importance for making research output free from various clutches, more so in situations due to the prohibitive cost of scholarly publications.

Having said so, it may be appropriate at this point to focus on Open Access and institutional repositories before switching on to the building blocks of OA repositories - Open source software.

3.0. Open Access and Open Content Licensing

Open access (OA) means immediate, free and unrestricted online access to digital scholarly material[1], primarily peer-reviewed research articles in scholarly journals. OA was made possible by the advent of the Internet.

Alternatively, we can say that Open-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions. What makes it possible that the Internet and the consent of the author or copyright-holder.

According to the Bethesda statement –

"An open access publication is one that meets the following two conditions: The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display

the work publicly and to make an distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (2) as well as the right to make small numbers of printed copies for their personal use”

The first major international statement on open access [2] was the Budapest Open Access Initiative in February 2002[3]. This provided a definition of open access, and has a growing list of signatories [4]. Two further statements followed: the Bethesda Statement on Open Access Publishing [5] in June 2003 and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [6] in October 2003.

OA has since become the subject of much discussion amongst researchers, academics, librarians, university administrators, funding agencies, government officials, commercial publishers, and society publishers. Although there is substantial (though not universal) agreement on the concept of OA itself, there is considerable debate and discussion about the economics of funding open access publishing, and the reliability and economic effects of self-archiving.

The core principle of OA is to open up access to research and scholarship, especially that which is publicly funded. OA is entirely compatible with peer review, and all the major OA initiatives for scientific and scholarly literature insist on its importance. OA literature is not free to produce, even if it is less expensive to produce than conventionally published literature.

The question thus arises as to what is then an open access contribution

3.1 Open Access Contribution

Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material and the like

As stated in the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [7] Open access contributions must satisfy two conditions:

The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit

and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.

A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, inter operability, and long-term archiving.

3.2. Vehicles for Delivering OA Content

There are two primary vehicles for delivering OA to research articles: OA journals and OA archives or repositories.

(I) OA Self-Archiving (sometimes known as the “green” road[8] [9]), authors publish in a subscription journal, but in addition make their articles freely accessible online, usually by depositing them in an institutional or central repository such as PubMed Central[10]). OA archives or repositories do not perform peer review, but simply make their contents freely available to the world. They may contain unrefereed preprints, refereed post prints, or both. Archives may belong to institutions. Authors may archive their preprints without anyone else’s permission, and a majority of journals already permit authors to archive their post prints. When archives comply with the metadata harvesting protocol of the Open Archives Initiative (OAI-PMH) then they are interoperable and users can find their contents without knowing which archives exist, where they are located, or what they contain. High-energy physicists had been self-archiving centrally in arXiv (http://arxiv.org/) since 1991.

(II) OA Publishing (sometimes known as the “gold” road [11] authors publish in open access journals that make their articles freely accessible online immediately upon

publication. Examples of OA publishers [12] are BioMed Central and the Public Library of Science [13]. OA journals perform peer review and then make the approved contents freely available to the world. Their expenses consist of peer review, manuscript preparation, and server space. OA journals that charge processing fees usually waive them in cases of economic hardship. OA journals with institutional subsidies tend to charge no processing fees. Some OA publishers waive the fee for all researchers affiliated with institutions that have purchased an annual membership. There's a lot of room for creativity in finding ways to pay the costs of a peer-reviewed

About 92% of peer-reviewed journals have endorsed some form of self-archiving [14]. About 10% of peer-reviewed journals are now OA journals [15].

There is now open-source software for building and maintaining OAI-compliant archives, which is being discussed here and worldwide momentum for using it. The costs of an archive are negligible, needs only some server space and some technical support.

Since open access repositories are primarily institutional based, let us focus on some parameters of the institutional repositories.

3.3. Open Access Routes

The routes that lead to open access are either Gold or Green. Following the Gold route, the authors publish in OA journals that make their articles freely accessible online immediately upon publication. OA journals are peer-reviewed. Depending on the model, authors may have to pay publishers a fee to publish. While taking the Green route, authors publish in a subscription journal, but also make their articles freely accessible online, usually by depositing them in either an institutional repository or central repository (either peer-reviewed post prints or non-peer-reviewed preprints).

3.4. Publishers' Archiving Policies

The importance and the momentum that OA is gathering at all levels have made the commercial stakeholders of scholarly communication system to change their mindset and be part of the OA game. The publishing fraternity has realized the potential of this new movement and has changed colours. Various colours would depict different scenarios especially regarding the embargo timeline. For instance : **Green signals that you** can

archive pre-print and post-print; Blue signifies that you can archive post-print (i.e. final draft post-peer-review); Yellow means you can archive pre-print (i.e. pre-peer-review); while White warns that archiving not formally supported. (source: http://www.sherpa.ac.uk/romeoinfo.html#colours)

3.5. OA Resources

There are several OA resources available in the cyberspace. While these are getting populated regularly, new resources crop up for access by all. Some such resources are directories such as:

Directory of Open Access Journals http://www.doaj.org/; OpenDOAR—the Directory of Open Access Repositories http://www.opendoar.org/; ROAR—Registry of Open Access Repositories, formerly called Tim Brody's Institutional Archives Registry

http://archives.eprints.org/; HighWire Press, Stanford University: Free Online Full-text Articles http://www.highwire.org/lists/freeart.dtl; Free Full Text http://www.freefulltext.com/; Free Medical Journals http://www.freemedicaljournals.com/.

Similarly, several open access Forums, Blogs, and News are out there. Examples include :American Scientist Open Access Forum: http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/ - a complete Hyper-mail archive of the ongoing discussion of providing open access to the peer-reviewed research literature online (1998-2005); Budapest Open Access Initiative Forum

http://www.soros.org/openaccess/forum.shtml; OA Librarian http://oalibrarian.blogspot.com/; Open Access News http://www.earlham.edu/~peters/fos/fosblog.html; SPARC Open Access Forum http://www.arl.org/sparc/soa/index.html#forum and SPARC Open Access Newsletterhttp://www.arl.org/sparc/soa/index.html

4.0. Institutional Repositories (IRs)

4.1. Institutional Repository - Definition

In broader sense, Institutional repositories are digital collections that capture and preserve the intellectual output of a single or multi-university community [16].

An institutional repository (IR) is a digital collection of a university's scholarly/creative output. Institutional repositories collect, preserve, and make accessible the knowledge generated by academic institutions. IRs also form part of a larger global system of repositories, which are indexed in a standardized way, and searchable using one interface, providing the foundation for a new model of scholarly publishing. Institutional repositories benefit scholars and the institution by bringing timely access, broader dissemination, increased use, and enhanced professional visibility of scholarly research and creative output while potentially raising the institutional profile. A growing number of universities around the world, such as the Massachusetts Institute of Technology (MIT) and the University of California have developed and are running institutional repositories, while many others are in the planning stages. [17].

Therefore, An insitutional repository stores and makes accessible to the educational, research and associated assets of an institution, where content are not limited to e-prints only. But the IRs can also include -research data, learning resources, image collections and many other different types of content. The primary goal of open access archive (OAA) is to maximize the accessibility of the research publications and their impact, as it is this that forms the basis for future scientific development.

4.2. Repositories - Rationale

Scientific Progress in developing countries is significantly hampered by the high cost of subscribing to the scientific and medical journals (in particular) that are essential for research to flourish. Open access archiving provides a means for fast track for building research capacity in developing countries and facilitate access. Open archive can transform the research scene from one of isolation and magnetization, to one of inclusion and international cooperation. It is an initiative of making sound archives available to wider academic community for fostering research, creativity and innovativeness.

4.3. Repositories: Need and Benefits

Repositories are important s an increasingly recognised means to capture, store and access the institutional knowledge base and intellectual assets which are growing in digital form. They support the open access goal of transforming scholarly communication and

becoming a major component in the evolving structure of scholarly communication. “IRs enhances the visibility of and improves access to research outputs; encourages data re-use and collaboration. Potentials of repositories are being recognised by funding bodies worldwide and there is an international trend of funding bodies requiring publication of research results through repositories (RCUK, Wellcome Trust, The US - National Institute of Health)

“. An institutional repository needs to be a service with continuity behind it. Institutions need to recognize that they are making commitments for the long term.” [18]. Digital information is more vulnerable to potential loss due to dependence on technology – preservation actions required within very short timeframe. Digital information is easily altered - measures required to ensure its continued integrity & authenticity. Guarantee of long-term preservation gives authors more incentives to deposit content and enhances a repository’s trustworthiness. Long-term preservation and access to scholarly and education material should be an important strategic area for all Academic/Research organizations.

4.4. Features & Functionality of IR

Some of the key common features of IR are:

- IR contains digital content (born-digital or digitized)
- It is community-driven (members - Institution or Consortium)
- Members of IR are also authors & copyright owners of the content
- IR provides persistent access to deposited documents; & open access to its content, (with some exceptions)
- IR is (generally) interoperable that facilitates in developing cross-archive aggregation and search services.

IR also supports several specific functions

- Registration of institutional users (authors)
- Document submission
- Approval/ moderation
- Archiving
- Dissemination
- Administration

4.5. Setting up an IR

There are several requirements and processes involved. Setting and managing an IR is a serious long-term undertaking.

Human resource is one of such components. It is useful to appoint an IR manager - responsible for the IR. Other tasks that will require (typically, part time in nature), personnel support that is: user support, advocacy, training and proxy/ mediated submissions.

The second component is the infrastructure, that will include; suitable IT and network infrastructure: Software, particularly open source, such as Greenstone Digital Library Software http://www.greenstone.org/cgi-bin/library, Dspace http://www.dspace.org/, Eprints http://www.eprints.org/, etc.

4.6. Economics of IR

Cost components involved in setting up and maintaining an IR involve start-up costs, ongoing costs and long term costs. Start-up costs may include, hardware (IR server, backup facility, network connectivity),software (free, if open source),installation and customization, drafting of policies and procedures. In addition to the start up costs, there are some ongoing costs, such as advocacy for getting content, support for IR hardware and software, user support, mediated submission and for upgrade/ migrations. While the long-term costs for digital preservation needs to be also considered.

4.7. IR: Planning, Implementation & Execution

The entire gamut of activities involved in planning, implementation and execution of IR can broadly be classified under four basic phases . These phases could be:

- Orientation Phase (be able to articulate the features, functionality and benefits of IR Study model IRs – their content, features and services.) Planning Phase (establish local aims for the IR and draw plans for setting up IR) Implementation Phase (Draw up implementation plan) Operational/Maintenance Phase (Draw up operational/ administrative/ maintenance procedures)

4.8. Open Source Tools—Applications, Packages, Platforms, Products, Systems, and Software

A number of tools, platforms, software and systems are out there to facilitate development and deployment of OA resources. Some such tools have been included here for reference. For instance Archimede – A Canadian software solution for institutional repositories. http://www.bibl.ulaval.ca/archimede/index.en.html; ARNO–Academic Research in the Netherlands Online. http://www.uba.uva.nl/arno ; CDSware —"CDSware, the integrated digital library system, is a suite of applications which provides the framework and tools for building and managing an autonomous digital library server." http://cdsware.cern.ch/cdsware/overview.html; Digital Commons — ProQuest —Launch OA Repositories http://www.proquest.com/products_pq/descriptions/digital_commons@.shtml,

http://il.proquest.com/products_umi/digitalcommons/; DiVA, the Digital Scientific Archive (Digitala Vetenskapliga Arkivet in Swedish) http://www.diva-portal.org/about.xsql; DPubS — Digital Publishing System — "DPubS (Digital Publishing System) is a powerful and flexible open-source system for publishing digital documents. http://dpubs.org/; DSpace— "The DSpace digital repository system captures, stores, indexes, preserves, and distributes digital research material." http://www.dspace.org/; EPrints — "The EPrints software creates OAI-Compliant Archives. "http://www.eprints.org/; ePublishing Toolkit — "The ePublishing Toolkit is a software package providing tools to help in publishing scientific content on the web." https://dev.livingreviews.org/projects/epubtk; Fedora — "Fedora open source software gives organizations a flexible service-oriented architecture for managing and delivering their digital content. At its core is a powerful digital object model that supports multiple views of each digital object and the relationships among digital objects." http://www.fedora.info/; GAPWorks — "GAPworks is the online publication system developed in the GAP project (funded by the German Research Foundation, DFG)." http://gapworks.berlios.de; Hyperjournal — "The HyperJournal is an Open Source software application which enables on-line as well as printed publishing in an innovative and significantly cost-cutting way." http://www.hjournal.org; iTor –Tools and Technologies for Open Repositories, Netherlands Institute for Scientific Information Services.http://www.i-Tor.org/en/

4.9. Building Blocks of IRs

The key component of an IR is the repository management software. There are several software's now available under open source license, that comply with OAI metadata harvesting protocol and are released and made publicly available. Some such IR softwares are:

ARNO - Tilburg University, The Netherlands (http://www.uba.uva.nl/arno);

CDSware - CERN, Geneva, Switzerland (http://cdsware.cer.ch/);

DSpace - MIT Libraries and the HP Labs, USA (http://www.dspace.org/);

EPrints - University of Southampton, U.K. (http://www.software.eprints.org/);

FEDORA - University of Virginia, USA (http://www.fedora.info/);

i-Tor - Netherlands Institute for Scientific Information Services (http://www.itor.org/en/toon);

MyCoRe - University of Duisburg- Essen, Germany (hhtp://www.mycore.de/engl/index.html).

However a list of such softwares are also available at http://www.soros.org/openaccess/software/

5.0. Defining OSS

Open-source software refers to any computer software whose source code is available under a license or arrangement such as the public domain that permits users to study, change, and improve the software, and to redistribute it in modified or unmodified form. It is often developed in a public, collaborative manner.

Open source describes the principles and methodologies to promote open access to the production and design process for various goods, products, resources and technical conclusions or advice. The term is most commonly applied to the source code of software that is made available to the general public with either relaxed or non-existent intellectual property restrictions. This allows users to create user-generated software content through either incremental individual effort, or collaboration. Some consider

open source as one of various possible design approaches, while others consider it a critical strategic element of their operations.

5.1. OSS: Compliance with the Criteria

Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria: [19].

(i) Free Redistribution : The license shall not restrict any party from selling or giving away the software and shall not require a royalty or other fee for such sale.

(ii) Source Code : The program must include source code, and must allow distribution in source code as well as compiled form. The source code must be the preferred form in which a programmer would modify the program.

(iii) Derived Works : The license must allow modifications and derived works, and must allow them to be distributed under the same terms as the license of the original software.

(iv) Integrity of The Author's Source Code : The license may restrict source-code from being distributed in modified form only if the license allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time.

(v) No Discrimination against Persons or Groups : The license must not discriminate against any person or group of persons.

(vi) No Discrimination against Fields of Endeavor : The license must not restrict anyone from making use of the program in a specific field of endeavor.

(vii) Distribution of License : The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.

(viii) License Must Not Be Specific to a Product : The rights attached to the program must not depend on the program's being part of a particular software

distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's license, all parties to whom the program is redistributed should have the same rights as those that are granted in conjunction with the original software distribution.

(ix) License Must Not Restrict Other Software : The license must not place restrictions on other software that is distributed along with the licensed software.

(x) License Must Be Technology-Neutral : No provision of the license may be predicated on any individual technology or style of interface.

5.2. Key Features of OSS

The OSS generally facilitates to capture and describe digital material using a workflow. It

Provides an interface for online submission of research material (Intranet) and also provides access to this material over the web (metadata and/or full pub). While it makes provision for preserving digital material over long period of time, it also exposes metadata through OAI-PMH protocol for developing cross-archive aggregation and search services. By default: unqualified Dublin Core is supported, besides other metadata standards. OSS supports several content streams or categories that includes published material for example: journal papers (post-prints), book chapters, conference papers; unpublished / gray material for example: pre-prints, working papers, minutes, theses and dissertations, technical reports, progress/ status reports, committee reports, course material, presentations, multimedia material, etc. and supporting material for example: data sets, models, simulations.

5.3. Content Producers and Content Functionality

Generally content here is referred to one produced by the respective institutions creating the repository of their intellectual output. This content could be varying forms as depicted schematically below.

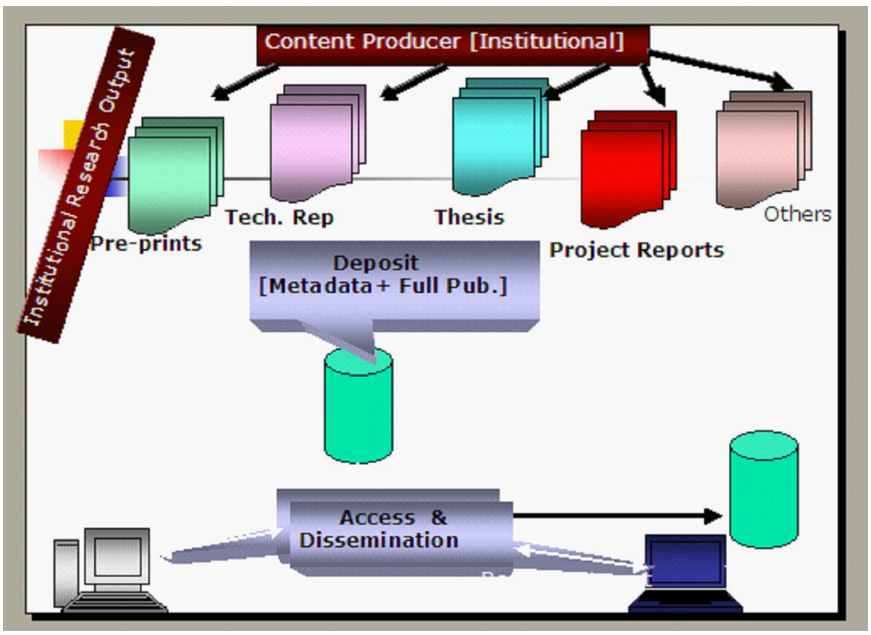

Figure 1. Content Producers

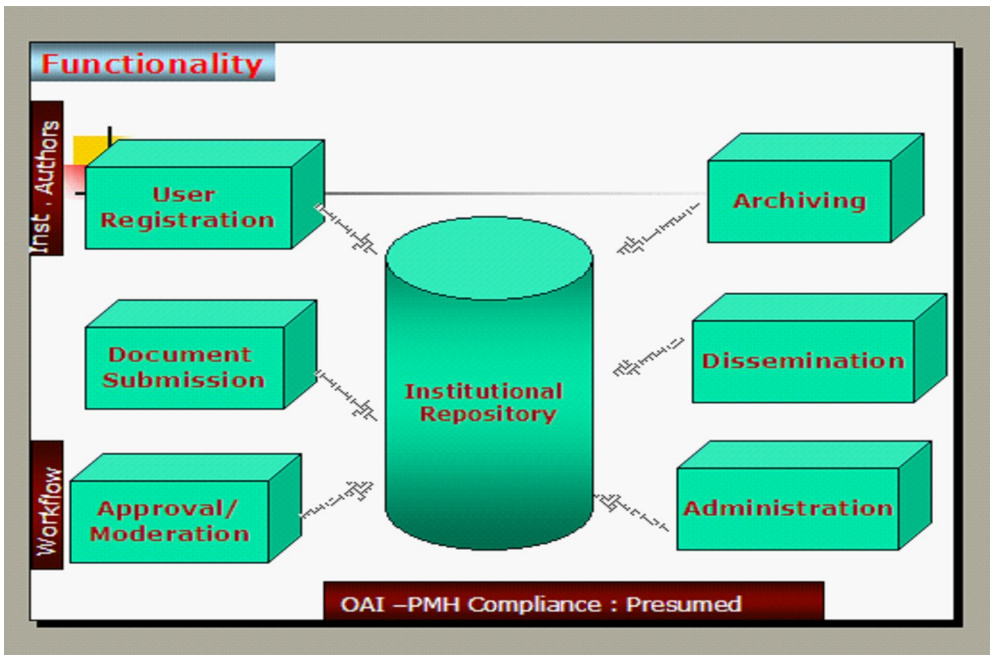

Figure 2. Content Functionality

The OSS has number of functions as is being schematically represented in Fig2.

In addition to the above indicated functions, additionally these support:

- Unicode based Multilingual content/ interface support
- Batch import/export
- Other interoperability protocols
- Z39.50, OpenURL
- Other metadata standards and crosswalks (e.g. METS, MARC)
- Persistent URL (Identifier)

5.4. OSS for Creating IRs: A look into the features of Dspace Suite

Before we begin to look into the suite of features of the Dspace, let us find out what Dspace is.

A digital library (DL) system generally supports – capture, store, index, preserve & redistribute functions for intellectual capital of an institution's research output in digital formats.

5.5. What is Dspace

Dspace is an Open Source Technology Platform (since 2002) that can be customized and its capabilities extended. Primarily a digital object asset management system. It is considered as a service model for open access and/or digital archiving for perpetual access. An institutional repository platform which holds collections those are searchable & retrievable on the web; which is open and interoperable. It is treated as a production service of the library to its local research community A federation of digital repositories run by multiple academic research institutions and is a joint project of MIT Libraries and HP Labs

5.6. Dspace Suite for Digital Repositories: Features

5.6.1. Dspace for Creating Digital Archives

The use of Dspace for creating institutional digital archives may be attributed to its following strengths.

- Communities / collections based model
- Backed up by MIT and HP
- Simple yet powerful documentation
- Strong workflow support
- Handle-based identifier
- Better articulation of preservation strategy
- Default support for qualified DC
- User (E-) Groups, Lists, User Meets ...

Features

The goodies that make the software to be acceptable to many and have led to substantial installations for creation of the institutional repositories are generally attributed to the following:

- Modular architecture, well-defined APIs
- 100% open source
- Programmed in java
- RDBMS and SQL for metadata
- CNRI "handles" for persistent identifiers
- OpenURL linking
- Metadata Standards - Supports
    - Dublin Core;
    - OAI-PMH (Open Archive's Initiative Protocol for metadata harvesting.) for exposing metadata

Other Goodies

The other capabilities of the software includes –

- Export**:** Exports in XML format. The latest version has METS (metadata encoding and transmission standard) for export
- Harvesting in DSpace: Using OAI Harvester -Open Archives Initiative – Protocol for Metadata Harvesting, the repositories can be located and harvested for scholarship purposes. It's necessary to exchange knowledge between repositories, avoiding

heterogeneity issues and giving an interpretation of metadata harvested. A Harvester: is basically a client application that issues OAI-PMH requests. It is operated by a service provider as a means of collecting metadata from repositories. DSpace has data provider functionalities. Thereby augmenting DSpace with the harvester support and then, importing DSpace in the broader context of Institutional repositories is feasible and more ingenious.

- UNICODE: - For creation of digital libraries in vernacular languages using UNICODE, the Unicode needs support from various quarters, such as *Operating systems, Programming languages, Application software, Word processors.* In the case of compatibility of UNICODE and Dspace, the requisite parameters are supported. For instance Linux supports Unicode; PostgresSQL supports Unicode; Java supports Unicode. You only require fonts to display in the browser and you need to modify Tomcat files •$TOMCAT_HOME/conf/server.xml. Make sure the line having port number look like this •<Connector port="8080" URIEncoding="UTF-8" ....

Information Model

Broadly speaking, the information model is based on the concept of Communities; Collections; Items Files.

Data Files, also called bitstreams, are organized together into related sets. An item is an "Archival Atom" consisting of grouped, related content & associated description [Metadata]. An item's metadata is indexed for browsing and searching. Items are organized into collections of logically-related material. A community is the highest level of the DSpace content hierarchy. They correspond to parts of the organization such as Departments, Labs, Research Centres, Units, and Divisions. Dspace's modular architecture allows for creation of large, multi-disciplinary repositories that ultimately can be expanded across institutional boundaries. Dspace is committed to going beyond reliable file preservation to offer functional preservation where files are kept accessible as technology formats, media, and paradigms evolve over time for as many types of files as possible.

The end user interface supports browsing and searching the archives. Once an item is located, web-native files can be displayed in a web browser while other formats can be downloaded and opened with a suitable application program. .

Search Features

The search features of the Open source software generally supported are:

- Fielded –sUse the *fielded search* to reduce or limit the range of a query in order to increase the relevance of the search results. A fielded search is an advanced query feature that enables users to select and associate the different document fields to which he wishes to limit the query, to then use the required keywords within these fields
- Boolean - *Boolean searching* facilitates to narrow down your search by using special terms before your keywords. Boolean logic refers to the logical relationship among search terms, Boolean logic consists of three logical operators: OR; AND; NOT. Hence with Boolean searching the following are being made use of:

Boolean term	**Used for:**
AND	To make sure a keyword is included
AND NOT (ANDNOT, NOT)	To make sure a keyword is not included
OR	To give alternative keywords

- Exact term – The search strategy facilitates retrieval of documents having the exact term or phrase as specified in the search parameter.
- Proximity - In text processing, a *proximity search* looks for documents where two or more separately matching term occurrences are within a specified distance, where distance is the number of intermediate words or characters. For example, a search could be used to find “red brick house”, and match phrases such as “red house of brick” or “house made of red brick”.
- Wild Cards - *Wild cards* in a search are characters that will match any character in the field. You can use them where you’re trying to find something like ‘a code that starts with ‘S’ and includes an ‘x’’. EX: a?e Matches ‘are’, ‘database’, ‘A/e’, a*e Matches ‘Alvechurch’, ‘database’, and ‘Mary had a little lamb’ because * matches any number of characters - including one.
- Fuzzy - *Fuzzy searching* will find a word even if it is misspelled. For example, a fuzzy search for *apple* will find *appple*. Fuzzy searching can be useful when you are searching

text that may contain typographical errors, or for text that has been scanned using optical character recognition (OCR).

- Range – *Range search* problems arise in database and geographic information system (GIS) applications. Any data object with *d* numerical fields, such as person with height, weight, and income, can be modelled as a point in *d*-dimensional space. E.g.asking for all people with income between $1000 and $10,000, with height between 6'0'' and 7'0'', and weight between 50 and 140 lbs. defines a box containing people whose body and wallets are both thin.
- Boosting Terms – *Boosting terms* provide the relevance level of matching documents based on the terms found. The higher the boost factor, the more relevant the term will be. Boosting allows you to control the relevance of a document by boosting its term. For example, if you are searching for: jakarta apache and you want the term "jakarta" to be more relevant boost it using the ^ symbol along with the boost factor next to the term. You would type: jakarta^4 apache

5.6.2. What Can You Do with DSpace

Dspace can help you in the four broad activities: by facilitating you the following

(i) Capture : Captures digital content in any format directly from creators e.g. researchers, authors-Via simple web forms Web based interface makes it easy for a submitter to create an archival item by depositing files.

(ii) Distribute : It facilitates the distribution of the content via secure web server with access control mechanism using persistent identifiers.

(iii) Describe : Provides option for descriptive rights, metadata, etc.

(iv) Preservation : Facilitates large scale, stable, long term and managed storage for content archives.

5.6.3. Content Streams

Dspace was designed to handle any format from simple text documents to datasets and digital video.

The content Streams that it can handle has wide variety of range. For instance repository could contain archives of Preprints, Articles, Postprints, working papers, Technical Reports, Conference Papers, Theses/Dissertations, Datasets, e.g. statistical, geospatial, scientific, Non-Institutional Research and the like.

5.6.3.1 Format - Content Streams

The format of these content steams could be images, visual, scientific, etc., Audio files, Video files, reformatted Digitized library collections, etc. However it may be mentioned here that Dspace handles various formats differently. There are three categories or types of managing various formats, such as:

(i) Supported: Fully supports the format

(ii) Known: recognizes the format, but cannot guarantee full support

(iii) Unsupported: Cannot recognize a format; these will be listed as “application/octet-stream”, — Unknown

For instance, document types or document support, Dspace supports or otherwise: let us take a few examples so as to make it more explicit.

The supported – non-proprietary formats like GIF, XML, JPEG, MARC, PostScript, TIFF.

The Known – proprietary formats: Excel, Mathematica, Photoshop, PowerPoint, Word WordPerfect

The Unknown: MIME Type: application/octet-stream

5.6.4. Architecture and System Requirements

The DSpace system is organized into 3 layers (i) Storage layer, (ii) Business layer, and (iii) Application layer, each of which have number of components.

Storage Layer: This layer is responsible for physical storage of metadata and content.

Business Layer: The layer deals with managing the content of the archive, users of the archive (e-people), authorization and workflow

Application Layer: The layer contains components that communicate with the networked world outside of the individual space installation. For example: the Web user interface and the modules for metadata harvesting service.

The source code is organized to cohere very strictly to this three layer architecture. Also, only methods in a components public API are given the public access level. This signifies that the Java compiler helps ensure that the source code conforms to the architecture.

Each layer invokes the layer below it. For instance, the application layer may not use the storage layer directly. Each component in the storage and business logic layer has a defined public API. The union of APIs of those components are referred to as the Storage API, in case of the storage layer, and in case of the Business logic layer, it is called as Dspace Public API. These APIs are in process Java classes, objects and methods.

It may be worthwhile to note that each layer is trusted, although the logic for authorizing actions is in the business layer, the system relies on individual applications in the application layer to correctly authenticate e-people. If an antagonistic or insecure application were allowed to invoke the public API directly, it could very easily perform actions as any e-person in the system. The reason for this design choice is that authentication methods will vary widely between different applications, so it makes sense to leave the logic and responsibility for that in these applications.

5.6.5. Backend Technology

Broadly speaking the backend technology for keeping dspace functionality up and running comprises: Apache, Tomcat, OpenSSL/mod_ssl; Java 1.3, JSP 1.2, Servlet 2.3

PostgreSQL (RDBMS); Java Database Connectivity (JDBC) to insulate their code from any particular database; CNRI Handle System (for Persistent ids); Lucene 1.2 (Index/ search). However a broader perspective of the requirements is reflected below.

5.6.6. Hardware and Software Requirements

Software Requirements

The software requirements for letting Dspace function would depend on several aspects, for instance the operating system that it would need to fall back upon for its execution, database management system, various other softwares and then Dspace itself. The

following are the software components that are required to be installed for successful implementation and running of the Dspace.

Operating System: Recommended Operating Environment is Linux or Windows XP/ 2000. However it may be mentioned here that many installations as of now are running on LINUX.

Database Management System: Uses Postgre SQL relational database.

Pre-Requisite Software: The following are suite of requisite software. However it may be pointed out here that the versions may change with the passage of time as the developments take place and new features and/or some changes are incorporated in the existing ones.

- Java SDK 1.4.2
- Apache 2.0.54
- Tomcat 5.0.28
- Apache Ant 1.6.5
- PostgreSQL 8.0.2

DSpace Software

- Dspace 1.4 .1

Utilities. It may be nice to have the following utilities.

- Acrobat Reader
- Acropad
- PDF995 Editor
- PDF995 Maker
- Winzip 9.0

Hardware Requirements

The hardware to host the requisite contents, database, etc. For instance the server(s) for storage, backup, local, remote on-site and off-site storage, or for migration of files as the technology changes.

5.7. How it Does, What it can ? ,

Before we discuss the processes, it may be indicated here that users are assumed to be institutional, with all the necessary resources to use the system (or the means to

outsource this), including hardware (with running OS) and a systems administrator to install and configure the system – and, in most cases, a Java programmer who can localize and customize the system.

Presuming your infrastructure as indicated above is in place. The next thing you need to ensure is the requisite softwares are installed and your system is set up for takeoff Broadly, two processes are to be followed for software, (i) Installation, (ii) DSpace Configuration & Customization.

(i) Installation: You need to install the software in sequence, such as indicated below for reference purposes:

Software Requirements

- Java SDK 1.4.2 (http://java.sun.com/j2se/1.4.2/download.html)
- Apache 2.0.54 (http://httpd.apache.org/download.cgi)
- Tomcat 5.0.28 (http://jakarta.apache.org/site/binindex.cgi) & mod_jk 2.0.4
- Apache Ant 1.6.5 (http://ant.apache.org/bindownload.cgi)
- PostgreSQL 8.0.2 (http://mirror.tomato.it/ftp/pub/PostgreSQL/binary/v8.0.2/win32/)
- DSpace 1.4 .1 (http://prdownloads.sourceforge.net/dspace/dspace-1.4-alpha-1-source.tar.gz)

Recommended DSpace file Configuration

1. J2SE SDK Path: C:\J2sdk1.4.2_09
2. Apache Path: C:\WWW\Apache2
3. Tomcat Path: c:\WWW\Tomcat5
4. Apache Ant Path: C:\ Apache-ant- 1.6.5
5. PostgreSQL Path: C:\Program Files\PostgreSQL
6. DSpace Path: C:\Dspace

To Start/Stop Services

Three Services are required for Dspace, Service names are given below:

- Apache Tomcat
- Apache2
- Postgre SQL Database Server 8.0

To start services go to

Start» Control panel » Performance and Maintenance »Administrative Tools » Component Services»Services [Local] »

[As a short cut, you may also get this at -Start»Run»services.msc]

Select the “Service” to be started from the list and Right click on it and Start

the service.

(ii) Dspace Workflow and Customization:

- Ensure that all the services [Apache, Tomcat, and PostgreSQL]are running.
- Visit the DSpace Home Page http://localhost:8080/dspace
- Login to My DSpace (this is mandatory for you to enable Dspace administration**).**
- Create a Community
- Click on the Create Top Level Community button at the top of the page.
- Create a Collection
- Note: You need to create a community for building collection/s.
- Start a New Submission
- Describe your item - [Fill out the form as needed] – 3 Stages
- Enter the Details of article
- Upload a File
- Verify your Submission
- Grant DSpace Distribution License by clicking “I grant the License” button.
    - Click-through license during submission
    - Grants DSpace non-exclusive right to acquire, manage, preserve, distribute the item
    - However License does not grant DSpace copyright
    - Copy of license is stored with the item Submission Complete

5.8. Publishing the Archive

For Publishing the Archive on the internet:

- Identify a suitable domain name ['dspace.mydomain']
- Register the Domain with public IP.
- Configuring email system
- Configuring CNRI Handle
- Mapping Items from Other Collections

6.0. Policy Issues

Policy issues needs to be formulated for local policy decisions. For instance policies regarding Submission criteria (who? what? when?); Definition & responsibilities of user communities; Acceptable file formats (supported, unsupported); Levels of access (public, restricted); Metadata standards for quality assurance; IPR issues (management of rights & obligations) and the like.

6.1. Copyright Issues

'A bunch of exclusive rights which the law gives to authors and creators to control certain activities relating to the use, dissemination and public performance of their original works. These issues are many times becoming the stumbling blocks for many stakeholders to promote OA implementation in the respective institutions.

There has been debate on various issues concerning copyright and various players involved. Amongst these players the main character in the whole story the author – let us see as to what should be the author's rights which can make lot of difference. Many leading OA supporters opine that the author should have right to reproduce the work in any manner or form; publish the work if it has not been published before; perform the work in public; broadcast the work; cause the work to be transmitted in a diffusion service and make an adaptation of the work. Besides, authors should not sign away all their rights. A description of Author's Rights and Addendum (SPARC) - are available at http://www.arl.org/sparc/author/. For reference

The detailed information on various Copyright and Licenses are available at Creative Commons Licenses http://creativecommons.org/licenses/; An Education in Copyright Law: A Primer for Cyberspace" by Robert N. Diotalevi http://libres.curtin.edu.au/libres13n1/; "Stealing the Goose: Copyright and Learning" by Rory McGreal http://www.irrodl.org/content/v5.3/mcgreal.html; Piet Zwart Institute, Open Content Licenses http://pzwart.wdka.hro.nl/mdr/research/lliang/open_content_guide; Publisher Copyright Policies and Self-Archiving http://www.sherpa.ac.uk/romeo.php.

6.2. OA Mandate

At the international level several agencies have come forward to mandate the OA either in their institutions or the research funded by them. The most significant step in this direction is that of National Institutes of Health (NIH), perhaps the first agency to take such a step. At the national level also several forums took place to address issues concerning copyright, open access and FLOSS. The special mention may be made of the meeting organized by Indian National Science Academy on April 26th 2008 and the forum has come out with set of recommendations. These recommendations are expected to sensitize funding bodies to mandate the deposition of scholarly content emerging out of public funding in the OA pools.

7.0. Epilogue

There is extensive debate and ongoing discussion in academic communities with special reference to scientific communities, literature, regarding open access and scholarly communication. The open source and open access movement is gathering momentum. The open access movement is supported and advanced by a spectrum of interest groups and activities such as national and international organizations, publishers, individuals, and many special events. While open access is gaining strength and popularity as the new model for dissemination of information, there are still many issues not completely resolved. The open access movement has had a tremendous worldwide impact and involves not only the academic and publishing communities but also many other players.

Open source software like Dspace and others have several goodies associated with them as indicated above. A much detailed description of Dspace can also be obtained at http://www.dspace.org. The open source software beyond doubt can help in generating digital, interoperable repositories. An end to end documentation site http://dspace.org/ implement/ for *building an Institutional Repository with Dspace* can be refereed to for creating one. Many organizations have used this software to create repositories of intellectual output of their R&D personnel and facilitated one stop access to such collections and also in the process trying to harvest the OAI-Compliant repositories in their respective areas of interest. For instance Indian Statistical Institute, Kolkata has such repository at

http://ir.isical.ac.in/ Librarians Digital Library http://ir.isical.ac.in/ at DRTC, Indian Statistical Institute, Bangalore; e-sholarship@ Indian National Science Academy [INSA], New Delhi http://www.insaindia.org/informatics_centre/service.htm and others. However large number of examples of public DSpace collections can be found at: http://wiki.dspace.org/index.php//DspaceInstances. Besides there are several Dspace projects and the details of the same are available at http://wiki.dspace.org/index.php/DspaceProjects. Also, Open Access Repositories listing can be located in ROAR registry at http://archives.eprints.org/. Moreover, the resources such as JISC Digital Repositories Programme: at http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories; JISC Briefing Paper on Digital Repositories, at http://www.jisc.ac.uk/index.cfm?name=pub_repositories and Supporting Digital Preservation and Asset Management in Institutions available at http://www.jisc.ac.uk/index.cfm?name=programme_404 are other valuable references for consultation, while designing and developing institutional repositories using open source softwares for digital libraries.

8.0. References

1. **Suber, Peter.** Open Access Overview : Focusing on open access to peer-reviewed research articles and their preprints. (http://www.earlham.edu/~peters/fos/overview.htm) visited on march 17, 2007
2. Budapest Open Access Initiative (http://www.soros.org/openaccess/read.shtml) visited on March 17, 2007.
3. Budapest Open Access Initiative: Frequently Asked Questions. (http://www.earlham.edu/~peters/fos/boaifaq.htm) visited on March 17, 2007.
4. Budapest Open Access Initiative: (http://www.soros.org/openaccess/view.cfm)
5. Bethesda Statement on Open Access Publishing. (http://www.earlham.edu/~peters/fos/bethesda.htm)
6. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (http://oa.mpg.de/openaccess-berlin/berlindeclaration.html)

7. **Ibid**

8. **Harnad, Stevan,** et.al. The Access/Impact Problem and the Green and Gold Roads to Open Access. Serials Review Volume 30, Issue 4, 2004, Pages 310-314 doi:10.1016/j.serrev.2004.09.013

9. Budapest Open Access Initiative (http://www.soros.org/openaccess/read.shtml)

10. **PubMed** Central (PMC) -A free archive of life sciences journals (http://www.pubmedcentral.nih.gov/)

11. **Op.cit Harnad, Stevan** , et.al.

12. **DOAJ**-Directory of Open Access Journals (http://www.doaj.org/)

13. **PloS** – Public Library of Science (http://www.plos.org/journals/index.html) visited on May 11, 2007.

14. Journal Policies - Summary Statistics So Far (http://romeo.eprints.org/stats.php) visited on May 11, 2007

15. Op.Cit DOAJ

16. **Crow, R.** (2002) The Case for Institutional Repositories: A SPARC Position Paper, SPARC, July http://www.arl.org/sparc/IR/ir.html visited on May 11, 2007.

17. **Queen's** Institutional Repository Portal (http://library.queensu.ca/webir/) visited on May 11, 2007

18. **Lynch, Clifford A.** RLG DigiNews, 2004 http://www.rlg.org/en/page.php?Page_ID=19481 visited on May 11, 2007.

About Author

Dr. Usha Mujoo Munshi, Librarian, IIPA, New Delhi