CERTIFICATE

This is to certify that this project "**Design and Development of Institutional Repository"** submitted by **M. SARASWATI RAO** for the award of the **Master of Library and Information Science,(2006-2008)**, Andhra University, Visakhapatnam, is a bonafide work carried by him under my guidance and supervision.

Visakhapatnam, (Prof.K. SomasekharaRao)

Date: Project Guide

DECLARATION

I here declare that the project Report entitled **"Design and Development of Institutional Repository"** submitted by me for the partial fulfillment of course leading to the award of Master of Library and Information Science, (2006-2008), Andhra University, Visakhapatnam, is of my own work and it has not been submitted to this university or any other university or published at any time before.

Visakhapatnam, (**M. SARASWATI RAO**)

Date:

ACKNOWLEDGEMENT

I wish to express my deep sense of gratitude to Prof.K.SomasekharaRao, Project Guide, Dept. of Library and Information Science, who has conceived the idea and instilled in me to undertake **"Design and Development of Institutional Repository"** and for her valuable suggestions all through my project.

I feel extremely grateful to express my gratitude to Dr.J.VijayaKumari, Head of the Dept. of Library and Information Science and other faculty members Prof.C.Sasikala, Prof. R.S.R.VaraLakshmi, Dr. P. Bhaskara Rao and Dr.N. David Livingston for their kind advices and encouragement given to me whenever needed during my course.

Last but not least I am thankful to the Friends, Library Staff and Non Teaching Staff of the Dept. of LIS for providing me the timely and necessary information without which my course has not successfully completed.

(**M. SARASWATI RAO**)

CONTENTS

Chapter 1 Page No.

Concept of Digital Library 1-13

1.1 Introduction 1

1.2 What is Digital Library? 2

1.3 Definitions of Digital Library 3

1.4 Characteristics of Digital Library 6

1.5 Advantages of Digital Library 7

1.6 Limitations 9

1.7 Benefits of Digital Library 11

1.8 Types of Digital Libraries 12

1.9 Summary 13

Chapter 2

Institutional Repositories 14-32

2.1 Introduction 14

2.2 Digital library Initiatives 16

2.3 Institutional Repositories 17

2.4 Institutional Repositories in India 18

2.5 Planning for Institutional Repository 21

2.6 Summary 32

Chapter 3 Page No.

Digital Library Software 33-50

3.1 Introduction 33

3.2 Digital library software 33

3.3 Greenstone 38

3.4 DSpace Durable Digital Depository 44

3.5 Summary 50

Chapter 4 51-79

Model Institutional Repository(LIS)

4.1 Introduction 51

3.2 Installation and Customization of Dspace 52

3.3 The steps in developing LIS repository using dspace 59

3.4 The steps in developing LIS repository using GSDL 70

3.5 Summary 79

Chapter 5

Page No.

Methodology 80-81

5.1 Introduction 80

5.2 Problem specification 80

5.3 Objectives 80

5.4 Data gathering 80

5.5 Conversion of print articles into electronic form 81

5.6 Development of Digital Repository 81

4.7 Limitations 81

4.8 Conclusion 81

Bibliography 84-85

Concept of Digital Library

1.1 Introduction:

The Library is an important part of any Institute. Traditionally, the concept of the library was “a storehouse for books”. Where books were kept in the lock and key and users could not directly access them. Over a period of time, Libraries introduced Open Access Systems, where users could see, read and issue books easily. 20th century libraries started to use Computers and IT tools to automate the different library functions.

To be able to “Provide the right information to the right reader at the right time” is possible only if libraries and information centers use Information Technology as a tool Libraries are becoming knowledge centers or knowledge hubs. The Librarian has to work as a knowledge manager or information managed. With the changing needs of the user, Libraries have to use latest technologies to satisfy its user’s information needs. Effective use of technology enables us to retrieve information quickly and efficiently[1].

Rapid advances in information processing, storage and communication technologies have revolutionized the role of world ideal libraries in disseminating information services to their users. As a result, libraries are facing new challenges; new competitors, new demands, new expectations and a variety of information services from users tailored to their wants and needs. The new technologies have facilited the transformation of data into digital format. Over the last few years information has been produced digitally and libraries have become major providers of information in digital format[2].

"This is a fascinating period in the history of libraries and electronic publishing. For the first time, it is possible to build large-scale services where collections of information are stored in digital formats and retrieved over networks. The materials are stored on computers. A network connects the computers to personal computers on the users' desks. In a completely digital library, nothing need ever reach paper."

Enterprise libraries have a great deal to learn and adapt from innovations in Digital Library (DL) research. Apart from using DL technologies in providing improved services using external information sources, libraries can play a key role by participating in national and international distributed DL efforts and in using DL technologies for publishing and managing internal content.

It is very important at the outset to appreciate the fact DLs are much more than technology - they are also about people and organizations - as users, producers and managers of information. As we will see, DLs have several facets[3].

1.2 What is Digital Library?

A Digital Library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system.

The first use of the term digital library in print may have been in a 1988 report to the corporation for National research Initiatives The term digital libraries was first popularized by The National Science Foundation (NSF), United States Government agency/The Defence Advance

Research projects Agency (DARPA), United States Department o Defence/The National Aeronautics and space Administration (NASA), United States Government Digital Libraries Initiative in 1994. The older names electronic library or virtual library are also occasionally used, though electronic library nowadays more often refers to portals, often provided by government agencies, as in the case of the Florida electronic Library.

Digital libraries help sharing references without physical boundary. The author indexing system helps easily uploading electronic documents and facilitates sharing documents. Traditional libraries are limited by storage space; digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. As such, the cost of maintaining a digital library is much lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent and additional books. Digital libraries do away with these fees[4].

1.3 Definitions of Digital Library:

1. "Digital libraries are organized collections of digital information. They combine the structuring and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible." (Michael Lesk)

2. "An informal definition of a digital library is a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network. A crucial

part of this definition is that the information is managed. A stream of data sent to earth from a satellite is not a library. The same data, when organized systematically, becomes a digital library collection." (William Arms)

3. "A DL contains digital representations of the objects found in it - most understanding of the "DL" probably also assumes that it will be accessible via the Internet, though not necessarily to everyone. But the idea of digitization is perhaps the only characteristic of a digital library on which there is universal agreement." (Harter)
4. "In its most basic form a DL should encompass two functions: a) Provide digital content to virtual, geographically dispersed users, and b) Pull in digital information electronically from outside sources irrespective of location."
5. Digital library is "a focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization, and maintenance of the collection." (Ian Witten and David Bainbridge).
6. "Digital libraries are different [from traditional library automation] in that they are designed to support the creation, maintenance, management, access to, and preservation of digital content." (Bernie Hurley, the Director for Library Technologies at U.C.Berkeley. Quoted in Digital library technology trends. Sun Microsystems. August 2002)

7. A digital library is a collection of documents in organized electronic form, available on the Internet or on CD-ROM

(compact-disk read-only memory ...).
www.netaonline.org/PD-DigitalGlossary.rtf

8. A digitized collection usually available online. It is also a series of activities that brings together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge. ...
www.lib.fsu.edu/glossary
9. Collection of texts, images, etc., encoded so as to be stored, retrieved, and read by computer. Digital printing
www.sir.arizona.edu/resources/glossary.html
10. Digital libraries can include reference material or resources accessible through the World Wide Web. Digitized portions of a library's collection or original material produced for the web can also be included in a digital library.
www.collectionscanada.ca/vrc-rvc/040017-151-e.html
11. Digital libraries are full-text databases that replicate, in digital media, many of the functions of traditional libraries. They tend to contain a purposefully selected collection of texts plus various means of access to these texts. ...
www.scils.rutgers.edu/~carballo/glossary.html
12. An organized collection stored in digital/electronic format and accessible online.
edtech.tennessee.edu/~set33/extra/glossary.html
13. A digital library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The term digital libraries (plural) are also used for the academic sub discipline that studies digital libraries. ...
en.wikipedia.org/wiki/Digital library

1.4 Characteristics of Digital Library:

These characteristics have been gleaned from various discussions about digital libraries, both online and in print (See Arms, 1995; Graham, 1995a; Chepesuik, 1997; Lynch and Garcia-Molina, 1995):

- digital libraries are the digital face of traditional libraries that include both digital collections and traditional, fixed media collections. So they encompass both electronic and paper materials.
- digital libraries will also include digital materials that exist outside the physical and administrative bounds of any one digital library
- digital libraries will include all the processes and services that are the backbone and nervous system of libraries. However, such traditional processes, though forming the basis digital library work, will have to be revised and enhanced to accommodate the differences between new digital media and traditional fixed media.
- digital libraries ideally provide a coherent view of all of the information contained within a library, no matter its form or format
- digital libraries will serve particular communities or constituencies, as traditional libraries do now, though those communities may be widely dispersed throughout the network.

digital libraries will require both the skills of librarians and well as those of computer scientists to be viable.

1.5 Advantages of Digital Libraries:

The advantages of digital libraries as a means of easily and rapidly accessing books, archives and images of various types are now widely recognized by commercial interests and public bodies alike. You can also access any user account simply through a link like this.

Traditional libraries are limited by storage space; digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. As such, the cost of maintaining a digital library is much lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries do away with these fees.

Digital libraries can immediately adopt innovations in technology providing users with improvements in electronic and audio book technology as well as presenting new forms of communication such as wikis and blogs.

- **No physical boundary:** The user of a digital library need not to go to the library physically; people from all over the world can gain access to the same information, as long as an Internet connection is available.
- **Round the clock availability:** A major advantage of digital libraries is that people can gain access to the information at any time, night or day.
- **Multiple accesses:** The same resources can be used at the same time by a number of users.
- **Structured approach:** Digital libraries provide access to much richer content in a more structured manner, i.e. we can easily move

from the catalog to the particular book then to a particular chapter and so on.

- **Information retrieval:** The user is able to use any search term (word, phrase, title, name, subject) to search the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to its resources.
- **Preservation and conservation:** An exact copy of the original can be made any number of times without any degradation in quality.
- **Space:** Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them. When a library has no space for extension digitization is the only solution.
- **Networking:** A particular digital library can provide a link to any other resources of other digital libraries very easily; thus a seamlessly integrated resource sharing can be achieved.
- **Cost:** In theory, the cost of maintaining a digital library is lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Although digital libraries do away with these fees, it has since been found that digital libraries can be no less expensive in their own way to operate. Digital libraries can and do incur large costs for the conversion of print materials into digital format, for the technical skills of staff to maintain them, and for the costs of maintaining online access (i.e. servers, bandwidth costs, etc.). Also, the information in a digital library must often be "migrated" every few years to the latest digital media. This process can incur very large costs in hardware and skilled personnel.

- Improved ‘visibility’ to the intellectual out put
- Universally accessible
- Easy access to electronic resources.
- Optimizing use of IT environment.
- Knowledge content itself will be ubiquitous and inexpensive.
- Ability to deal with large datasets.
- Support wider range of materials.
- Access to latest information.
- No Storage Problems.
- Faster information retrieval.
- Increase in end users.

1.6 Limitations:

Digital Libraries are not without their limitations, at least for the time being. The very large volume of print material that already exists (and continues to get created) and the comfort attached to reading from print material, means that print and digital material are going to co-exist for a very long time to come. Following are some of the key issues facing DLs today:

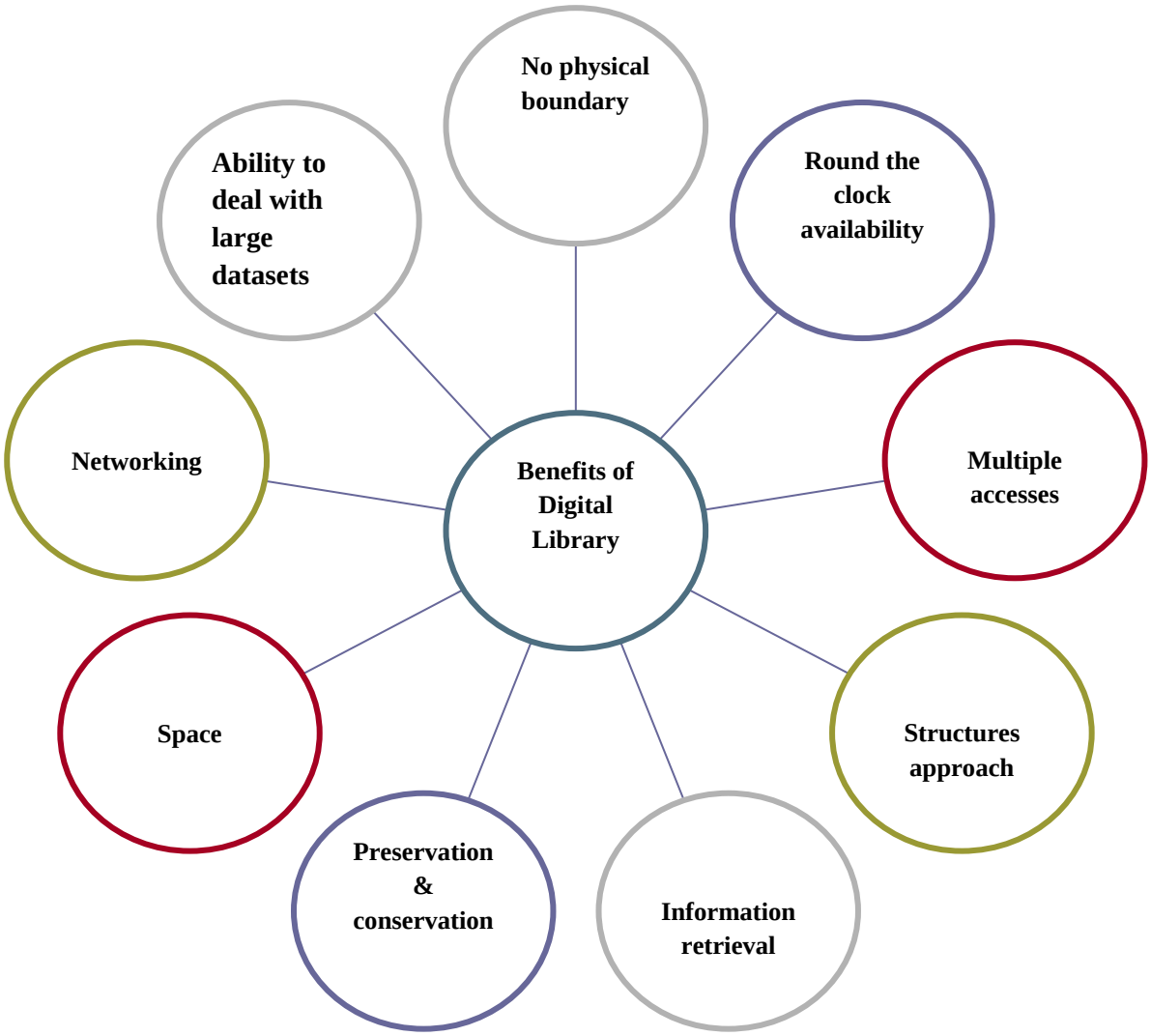
- Technological obsolescence:
  - Hardware: The major risk to digital objects is not physical deterioration, but technological obsolescence of the devices to read them. While the lifetime of optical and magneto-optical cartridges is expected to be in decades, those of reading devices are only about one decade. (Solution: Reasonable refreshing schedule and use of widely marketed devices than special purpose devices.)

- Software: A more serious problem is software obsolescence. It has been pointed out that the variety of software formats far exceeds the number of hardware devices manufactured, and that these programs come and go more quickly than the hardware does. (Solution: Libraries should rely on standards like MARC and SGML, which are expected to exist for the foreseeable future.)

- Cost of content refreshing:
  - Digital preservation will be an ongoing operation, requiring considerable recurring expense. This is a key issue for digital archiving. However, given the declining costs of technology, if a library can understand how it will fund the first refresh cycle in five to ten years, it can expect that the next refresh cycle will be so cheap as to be insignificant.
- Rights management:
  - It is very easy to copy, replicate, massage and distribute digital information. Enforcing copyright in digital environment is a major issue.
- Inter-operability:
  - Another key issue is that of interworking of different digital libraries. Given the distribution of library resources around the world, no one expects that there would be a single digital resource. This means that we need methods for finding either individual items or collections in different places, and assembling virtual collections that users can search or browse.
- Network bandwidth:
  - With the exploding number of users, multi-media content, web sites and applications (e.g. chat, games), bandwidth

available for accessing Digital Libraries is going to be a serious issue. Even a single speed CD-ROM drive (@ 150KBPS) provides faster access time than a 56 KBPS modem. As another example, over a 64 Kbps ISDN line, an ASCII text page arrives in less than half a second while an image page takes 4 seconds which is a definite lag.

1.7 Benefits of Digital Library



1.8 Types of Digital Libraries*:*

They are broadly categorized into three:

1. Academic Repositories.
2. Digital Archives.
3. Digital libraries as portals

1.8.1 Academic Repositories*:*

Many academic libraries are actively involved in building institutional repositories of the institution's books, papers, theses, and other works which can be digitized or were 'born digital'. Many of these repositories are made available to the general public with few restrictions, in accordance with the goals of open access. Institutional, truly free, and corporate repositories are often referred to as digital libraries.

A convenient definition is a “digital collection capturing and preserving the intellectual output of a single or multi-university community”. (The case for Institutional Repositories: a SPARC position prepared by Raym Crow)

1.8.2 Digital Archives*:*

Archives differ from libraries in several ways. Traditionally, archives were defined as:

1. Containing primary sources of information (typically letters and papers directly produced by an individual or organization) rather than the secondary sources found in a library (books, etc);
2. Having their contents organized in groups rather than individual items. Whereas books in a library are cataloged individually, items in an archive are typically grouped by provenance (the individual

or organization who created them) and original order (the order in which the materials were kept by the creator);

3. Having unique contents. Whereas a book may be found at many different libraries, depending on its rarity, the records in an archive are usually one-of-a-kind, and cannot be found or consulted at any other location except at the archive that holds them.

1.8.3 Digital Libraries as Portals:

With the growing acceptance of web portals on university campuses, many librarians are considering building library portals to make their collections accessible to users who expect all knowledge to be obtainable with just a few clicks from inside a web browser. After all, the World Wide Web has become the marketplace for documents, goods, services, and ideas.

For many people, especially students, if something is not on the web, it does not exist. This trend is especially challenging for libraries, which are the traditional keepers of knowledge but whose knowledge is largely kept in many millions of books and journals not on the web. The idea of a library portal is misdirected. It is vital that libraries have a presence on university enterprise portals, libraries should build portal pages, portal channels, and portal cameos rather than entire library portals.

1.9 Summary:

One thing digital libraries will not be is a single, completely digital system that provides instant access to all information, for all sectors of society, from anywhere in the world.

INSTITUTIONAL REPOSITORIES

2.1 Introduction:

In the information society, free flow of information is a fundamental principle for bridging the Knowledge gaps between privileged and under-privileged communities. Social inclusion and economic empowerment are also achieved in a society where citizens have universal access to information and knowledge, ranging from public information to specialized or customized information related to ones' profession, vocation or culture. Open access to information and knowledge is a key contributor in provisioning universal access to information and knowledge. Thus, the open access movement is gaining momentum in developed countries which already have necessary information infrastructure.

The emerging economies amongst the developing countries are not far behind in building up necessary information infrastructure, essential for sustainable economic development. These emerging countries, however, have limitations in terms of bridging the digital divide within their societies, due to the co-existence of marginalized and privileged communities. With the availability of advanced information and communication technologies (ICTs) and by building up necessary information infrastructure in South Asia, particularly in India, this sub-region becomes an active contributor to global open access literature, a contribution, almost proportionate to its flow of scholarly literature through subscription-based channels. For several thousand years, South Asian literature has been developing, explicit with knowledge of the time.

Traditional knowledge, presently available amongst indigenous people in folklore or other forms, is now being explored and documented. The digitization of century-old publications and rare documents is being carried out in different institutions across this sub-region to preserve them and make them accessible through networks of digital archives. E-content is are being produced by members of different communities, facilitated by certain development organizations, including the apex higher education agencies (viz., University Grants Commission, India; Higher Education Council, Pakistan); apex research councils, and grassroots-level civil society organizations. India has spearheaded the open access movement in developing countries since the last decade up scaling print-based scholarly journals into open access electronic journals, and establishing a number of open access repositories, both at national and institutional levels embracing free and open source software (FOSS). Establishment of open courseware and cross-archive search services are new fronts of open access initiatives.

In the information society, free flow of information is a fundamental principle for bridging the Knowledge gaps between privileged and under-privileged communities. Social inclusion and economic empowerment are also achieved in a society where citizens have universal access to information and knowledge, ranging from public information to specialized or customized information related to ones' profession, vocation or culture. Open access to information and knowledge is a key contributor in provisioning universal access to information and knowledge. Thus, the open access movement is gaining momentum in developed countries which already have necessary information infrastructure.

2.2 Digital Library initiatives*:*

The growth and the popularity of the Internet and WWW resulted in two major DL initiatives being taken in the mid 1990's in the USA.

First was the joint initiative of the National Science Foundation (NSF), Department of Defence Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA), in 1994, to fund six digital library development projects for a period of four years, among six academic institutions. The second was the signing of the National Digital Library Federation Agreement in May, 1995, led by the Library of Congress and 14 other research libraries. The purpose was to "bring together - from across the nation and beyond - digitized materials that will be made available to students, scholars and citizens everywhere".

NSF/DARPA/NASA DL initiative targeted three research areas:

- *Capturing, categorizing, organizing information*
  - Page, speech, video, graphics understanding
  - Indexing, hypermedia linking, knowledge representation
- *Searching, browsing, filtering, summarizing, visualization*
  - Theories, models, intelligent processing, learning
  - Simulation, navigation, metaphors, optimization
- *Networking protocols and standards, using networked information*
  - Security, know bots, compression, modeling, IPR

For details of these projects and other other initiatives (e.g. TULIP) taken in the USA, see the article by Padmini Srinivasan in DBIT special issue. The NSF/DARPA/NASA projects have been renewed for another four years. For current status of these projects, follow the links given in

the Berkeley digital library Sunsiste. The D-LIb magazine also provides leads to DL projects both in the USA and other countries.

2.3 Institutional Repositories:

Developments in information and communication technologies hold great potential for the advancement of knowledge and the good human kind through the open access of scholarly literature. Due to the fast growth of information value, proper description of distributed information is essential for effective information retrieval and filtering. Many proposals have emerged. Among these Institutional Repository, which promises to be extremely advantageous to scientists and scholars, particularly in the developing world? An institutional repository is a relatively new model for storing research output of a given institute or organization. The application of Institutional repositories provides ways of capturing, storing, indexing, preserving and disseminating digital objects. In its primary role as an institutional repository package this is the intellectual output of institution when such out put is in digital form. This can include, but it is not limited to, research papers, conference papers, book chapters, datasets, learning objects and E-theses. The idea of Intuitional Repositories has gained much ground at the international level. Free availability of literature on the public internet, permitting of users to read, download, copy, distribute, print, search or link to full texts of these articles.

2.4 INSTITUTIONAL REPOSITORIES IN INDIA :

In India, some institutions, like Indian Institute of Science; Indian Institute of Management, Kozhikode; Indian Statistical Institute, Bangalore; Indian Institute of Technology, Delhi; National Institute of Technology, Rourkela; National Aerospace Laboratories, Bangalore; National Chemical Laboratory, Pune; Information and Library Network (INFLIBNET), Ahmedabad; National Institute of Oceanography, Goa; Raman Research Institute, Bangalore; etc. have established open access institutional repositories (IRs) that disseminate research outputs of respective institution. Sometimes, these are self-archived. Otherwise, administrator of the repositories collects the research documents from different sources and submits the documents to the IR on behalf of the persons concerned .

Another band of digital repositories also exist in India that store and provide access to subject specific collections of documents. These repositories accept scholarly publications from any professional or researcher who belongs to the respective subject. *Librarian's Digital Library*(LDL) of Documentation Research and Training Centre (DRTC), Bangalore is an example of subject-specific repository for the library and information professionals. Another subject-specific repository established in India is *OpenMed@NIC*, maintained by National Informatics Centre, New Delhi. OpenMed@NIC stores and provides access to biomedical literature. Other kind of digital repositories existing in India stores and provides access to document type specific collections. Vidyanidhi of University of Mysore is an example of document type specific collection that stores and provides access to theses and dissertations. Vidyanidhi

accepts any thesis or dissertation from any researcher or student that is accepted in any of the Indian universities or institutions.

Now some open access archives in India also offer RSS (Really Simple Syndication or Rich Site Summary) feeds to the registered users. In this facility, the registered users get notification on addition of new items in regular intervals. The OpenMed@NIC has started providing RSS feeds. Table 1 shows an indicative list of established institutional repositories and subject repositories in India. These repositories follow self-archiving model and have initiated using two most popular open source software` – DSpace and GNU EPrints. Some institutions have customized certain interfaces at the time of installation or up gradation such as Opened of NIC and e-Prints of Indian Institute of Science. Among them, some repositories are growing very rapidly and reaching worldwide. Search engines and metadata harvesters index many of the operational repositories of India. Table 1 also indicates that maximum number of institutional repositories belong to scientific and technological (S&T) areas than humanities and social sciences areas. In *Appendix I*, screenshots of some well recognized institutional repositories of India are given.

INSTITUTIONAL REPOSITORIES IN INDIA

Name	Host Institution	URL	No. of Items	Types of Documents	Software Used
Librarian's Digital Library (LDL)	Documentation Research & Training Centre (DRTC)	https://drtc.isibang.ac.in/	236	Research Papers, Articles, Reports, etc.	DSpace
DSpace at GBPUAT University	G.B. Pant University of Agriculture & Technology	http://202.141.116.205/dspac e/	82	Research Papers, Articles, Reports, Thesis, etc.	DSpace
IIA Repository	Indian Institute of Astrophysics	http://prints.iiap.res.in/	725	Research Papers, Articles, Reports, Thesis, etc.	DSpace
EPrints@IIITA	Indian Institute of Information	http://eprints.iiita.ac.in/	22	Research Papers, Articles,	EPrints
DSpace@IIMK	Indian Institute of Management, Kozhikode (IIMK)	http://dspace.iimk.ac.in/	133	Research Papers, Articles, Reports, etc.	DSpace
EPrints@IIMK	Indian Institute of Management,	http://eprints.iimk.ac.in/	25	Research Papers, Articles,	EPrints
EPrints@IISC	Indian Institute of Science (IISC)	http://eprints.iisc.ernet.in/	3645	Research Papers, Articles,	EPrints
ETD@IISc	Indian Institute of Science (IISC)	http://etd.ncsi.iisc.ernet.in /	153	Theses & Dissertations	DSpace
EPrints@IITD	Indian Institute of Technology, Delhi (IITD)	http://eprint.iitd.ac.in/dspa ce/	1296	Research Papers, Articles, Reports, etc.	DSpace
DSpace at INSA	Indian National Science Academy (INSA)	http://61.16.154.195/dspa ce/	818	Conference Papers, Articles, Reports, etc.	DSpace
ISI Library, Bangalore	Indian Statistical Institute, Bangalore	http://library.isibang.ac.in :80 80/dspace/	10	Research Papers, Articles, Reports, etc.	DSpace
DSpace at INFLIBNET	INFLIBNET	http://dspace.inflibnet.ac.i n	428	Research Papers, Articles, Reports, etc.	DSpace
NAL Institutional Repository	National Aerospace Laboratories (NAL)	http://nal-ir.nal.res.in/	418	Research Papers, Articles, Reports, etc.	EPrints
DSpace at NCRA	National Centre for Radio Astrophysics	http://ncralib.ncra.tifr.res.i n/ dspace/	22	Research Papers, Articles, Reports, Thesis, etc.	DSpace
EPrints at NCL	National Chemical Laboratory (NCL)	http://dspace.ncl.res.in/	290	Theses, Research Papers, Articles, Reports, etc.	DSpace
OpenMED@NIC	National Informatics Centre (NIC)	http://openmed.nic.in/	1035	Research Papers, Articles, Reports, etc.	EPrints
Digital Repository Service of NIO	National Institute of Oceanography	http://drs.nio.org/drs/	55	Journal articles, conference proceeding	DSpace
Dspace@NITR	National Institute of Technology, Rourkela	http://dspace.nitrkl.ac.in/d sp ace/	223	Theses, Research Papers, Articles, Reports, etc.	DSpace
Digital Repository of RRI	Raman Research Institute	http://dspace.rri.res.in/	1064	Research Papers, Articles, Reports, Thesis, etc.	DSpace
Vidyanidhi	University of Mysore	http://www.vidyanidhi.org.in /	1835	Theses & Dissertations	DSpace

2.5 Planning for Institutional Repository:

The basic issues to be considered while planning to design and develop an I.R include:

- ✓ Users
- ✓ Content
- ✓ Intellectual Property Rights
- ✓ Technical / technologies
- ✓ Human Resources
- ✓ Workflow
- ✓ Access and use of digital repositories
- ✓ Standards
- ✓ Issues for Discussion

2.5.1 Users:

Who are we aiming at Institutional Repository (I.R)? For any endeavors to survive it must have an interested audience which is willing to use and evolve with it.

The audience is paramount for the growth of an IR.

2.5.2 Content:

Content-related issues will be crucial to the success of these repositories

Published material

Ex: journal papers (post-prints), book chapters, conference papers

Unpublished/gray material

Ex: Pre-Prints ,Working papers,minutes,theses and dissertations, technical reports, progress/ status reports, committee reports, course material, presentations, multimedia material,etc.

Supporting material

Ex: Data sets, models, simulations, all file types, including streaming media

2.5.3 Intellectual Property Rights:

Posting to I.R. a pre published article may involve copyright issues that have legal implications.

Three organizations – have developed self – sufficient addenda that address these issues

- MIt
- Science commons (through its Scholar's copyright project GNU General Public License
- SPARC
- The Ro MEO Project (Rights MEtadata for Open archiving)
- Project Partnering on Copyright (new project)
- Creative Commons
- GNU General Public License

2.5.4 Technical / technologies:

Creation, maintenance and provision of services using digital library collections require several resources. Two major resources required for setting up and operating digital library collections are: ***Technology Infrastructure*** and ***Personnel***. Technology infrastructure includes computing and other equipment, software and network connectivity. Personnel will be required for handling various tasks associated with the creation and maintenance of the collection.

Technology Infrastructure

1. Server computer
2. Desktop computers
3. Digitization equipment
4. Network connectivity
5. Other equipment

1. Server computer:

The server computer holds the digital library collection, runs the digital library software which provides the searching and processing applications. The server also handles the communication with the user over the network. The server needs to be reasonably powerful (typically measured in terms of processor speed, for example mega or giga hertz), has adequate main memory (RAM), has large amounts of hard disk storage and has good communication capability (network card and bandwidth). Options include: specialized servers or server-class PCs. Ensure that the server is expandable (RAM, storage, CPU) particularly if your digital library collection size and its usage is expected to grow rapidly. Options for Server operating system include Windows 2000, Unix and Linux. This is mainly determined by the digital library software you plan to use. License costs may be involved if it is proprietary operating system. Ensure maintenance support is available. In terms of capacity, it is useful to plan for 5 year life and size the resource requirements at 25-50% higher than the projected growth. Plan for online and offline storage backups for the digital library collection and procedures for taking regular backups. You should also have a contingency plan if the server computer fails - for example by moving

your collection to another server. This requires that you have the required digital library application programs supported on this server. Another option is to host the collection on one or more mirror servers located at different geographical locations.

2 Desktop computers:

Desktop computers will be required for staff handling various tasks related to the digital library collection. These include:

- Staff managing the digital library content (scanning, conversion, editing, cataloguing and quality checking);
- Programmers (development and technical support) and
- Staff who administer and maintain the digital library server and website.
- Also consider desktop computer requirements for users, if public access is provided locally.

You need to assess the actual requirements depending on the nature of your digital library project.

3 Capture devices:

These are digitization equipment needed for converting source material in physical/ analogue format to digital format. Digitization equipment include scanners for scanning print publications, video and audio capture cards for capture/conversion of digital video and audio files, and digital cameras for capture/production of video clips/still images. Requirement will be minimal if capture and conversion is largely outsourced.

4 Networking and location of the digital library server computer:

There are several options for providing online access to your digital library collection via the Internet. If your organization already has the requisite network infrastructure (intranet and Internet connectivity) then you can take help from your network administrator to connect and configure your digital library collection server to be accessible over the Internet. If your digital library collection is relatively small and does not require a dedicated server, then you host it on your library web server, if one exists, or on the institution web server, if this is permitted. If there is no network infrastructure in your organization and if you have the required resources, then you can obtain dedicated Internet connection to an ISP (Internet Service Provider) - but this could be expensive as this would require additional equipment like router, and DNS and Proxy servers. Maintaining this infrastructure will require more resources. Another option is to locate the digital library server computer in a data centre or ISP, on payment of rental fee. You can also hire storage space on a server computer owned by the ISP and host your digital library collection on this server. Several issues however need to be considered. Content updating may not be easy. Maintenance and troubleshooting the server or the collection may also be difficult. If you are hiring computer storage to host the collection, ISP may not provide the support for the digital library software you need.

5 Other equipment requirements:

Other equipment you may need include: CD Writer and label printing systems, laser printers, tape or CD-ROM backup systems, UPS support, local and offsite storage backup systems.

2.5.5 Personnel:

Personnel comprise the most important resource for the digital library - not only during the initial creation and set up, but also for its operation, maintenance and provision of services. Since access to the digital library is easy compared to a physical library, more users are likely to access it. If the digital library does not meet the expectations of the users in terms of currency and quality of content, they will lose confidence and are less likely to visit the digital library again. It is therefore important that you assign personnel with right skills and attitude to handle various tasks associated with the digital library project.

Actual personnel required for a digital library project depend on the type and volume of tasks carried out. Broadly, personnel will be required for following tasks:

- Project management
- Selection and preparation of source material
- Digitization and conversion
- Cataloguing and metadata assignment
- Quality assessment
- System administration and maintenance of digital library server and website
- System analyst/ programmer for digital library application/ interface development

- Promotion and provision of services

In a smaller institution, a part-time project manager working with jut one or two well-trained volunteers may be sufficient to successfully complete a small project. For many digital library projects (e.g. dissertations and theses), there may already be staff in the library handling the source material in paper-based form. In converting this material to a digital library, much of this staff may be trained to handle processes in the digital environment. For temporary and part-time tasks, you may also consider seeking support from enthusiastic volunteers from within the library or other departments. You can also take advantage of student's assistance/ intern programmes available in most educational institutions for employing students on hourly payment basis. It may also be possible to contract services, particularly for grant-based projects.

While hiring new personnel or re-deploying existing personnel, it is important that their education and skills match the requirements of the job. It is also important that all employed staff in the digital library project have their positions described in clear terms, including responsibilities and expected performance standards.

2.5.6 Digital library software:

This is probably the most important and central aspect of the technology infrastructure for the digital library. Digital library software works with the web server (and the operating system) in providing various digital library functionality including creation, organization and maintenance, indexing, search and retrieval.

Digital library software options:

Several software options are available today for creation and provision of access to digital library collections. A clear understanding of the requirements and features of the collection you plan to build is very important to assess and select the option that best suits your need. A good understanding of features expected of digital library software will also help you in making informed decision.

What features you may expect of digital library software? These include: support for different document types and formats, metadata support, online/batch content updating, indexing and storage, search and retrieval (metadata, full-text), multi-lingual support, interoperability support, access and usage management, collection administration, and support for standards compliance like Dublin Core, Unicode and XML.

In-house software development:

If you have in-house software development and maintenance support, another interesting possibility is to develop the software in-house. Simple, database-driven web applications can be developed to set up and provide access to your digital library collection. Open source database (e.g. MySQL and PostGress) and programming tools (e.g. Perl, PHP, Python and Java) available for Linux operating system platform can be used for developing the digital library application. Advantage of this approach is customization of the application to suit your requirement. Disadvantage is the cost of development as software projects are known to overrun schedules. Maintenance may also be difficult if you do not have dedicated software team.

Other software requirements:

Other software you may need include: Digitization software like Acrobat, OCR software (e.g. Fine Reader), image editors, web server and browsers, word processing software including HTML/XML editors, and DBMS package. Many of these are available freely, both for Windows and Linux operating systems.

2.5.7 Staff skills:

Skills include both hard (technical and IT skills) and soft (e.g. communication) skills. Apart from professional and technical skills, soft skills such as the following could be important: communication skills, ability to work in a team, flexibility, people skills, capacity to learn constantly and quickly, ability to work under pressure, attention to details, goal oriented, time-consciousness, capacity to work independently, public-service perspective, presentation skills, negotiation skills, and teaching/ training skills.

You may also like to consult professional staffing standards or guidelines developed by other similar libraries or projects.

Staff training:

It may not be possible always to find personnel with right background. Staff training becomes critical. Training should be appropriate to the skill requirements of the concerned operations. In-house training options include - special training programmes and on-the-job training. Staff may also be deputed to relevant training programmes. It will be useful to prepare a training manual which will be very useful for new staff. There also large number of online resources, including tutorials, available from authoritative sources related to digitization.

The rapid change in technology and practices in digital library technologies requires constant re-training and re-positioning of staff. This usually means that the project manager spends good bit of time keeping up-to-date on developments in the field, learning about emerging standards and best practices, and then take what is valuable, incorporate it in the project plan, and train other staff in its implementation.

2.5.8 Workflow and process:

- User identification and registration
- Document submission Content selection, moderation
- Organization (Physical and intellectual)
- Archiving
- Networking and interoperation
- Content access and delivery
- Usage monitoring
- Content maintenance and preservation

2.5.9 Access and Use of Digital repositories:

What type of delivery will be offered?

- Internal visibility of research output and access to the same can be made available through intranet or the existing campus network.
- External access is possible through Internet.

Organizations support open access

i. Open Archive Initiative(OAI)
ii. Open Directory of Open Access repositories (Open DOAR)

iii. Scientific Registry of Open Access Repositories(ROAR)

iv. SCIRUS

2.5.10 Standards:

Documents from OAI/others that provide guidance Harvesting – version 2.0(2002)

Open Archives Initiatives Protocol for Metadata Harvesting Implementation (http://www.openarchives.org/OAI/2.0/guidelines.htm)

Best Practices for OAI Data provider Implementations and shareable metadata; A joint initiative between the Digital Library Federation and the National Science Digital Library,2006.

Technical Standards (v.1.1)for JISC Information Environment

2.5.11 Issues for Discussion:

In India there are only 17 IRs,why?

Issue1: In spite of its immense value for the present and future information distribution, digital repositories are not getting the patronage of academic and scientific community the way it ought to be.

Issue 2: There is no evidence of making open access I.R.s mandatory in India by any of the government agencies or research institutes.

2.6 Summary:

Information technology infrastructure and personnel comprise two critical resources required for digital library projects. Actual requirements depend on nature of the digital library collection (one-off or live collection) and on factors such as volume of source material to be digitized, metadata requirements, document formats to be supported, search and retrieval expectations, and quality requirements. Hardware requirements include server computer for hosting the collection, desktop computers, digitization equipment, network connectivity, and other equipment. Digital library software is another critical technology component. Options include: Open source free digital library software, library automation software, commercial digital library software and in-house software development. Each of these has advantages and disadvantages. Personnel comprise the most important resource for the digital library apart from technology. Actual personnel required for a digital library project depend on the type and volume of tasks carried out.

Digital Library Software

3.1 Introduction:

Software is probably the most important and central aspect of the technology infrastructure and digital library is not an exception. Digital library software works with the web server (and the operating system) in providing various digital library functionality including creation, organization and maintenance, indexing, search and retrieval.

Several software options are available today for creation and provision of access to digital library collections. A clear understanding of the requirements and features of the collection we plan to build is very important to assess and select the option that best suits ours need. A good understanding of features expected of digital library software will also help us in making informed decision.

The features of digital library software include: support for different document types and formats, metadata support, online/batch content updating, indexing and storage, search and retrieval (metadata, full-text), multi-lingual support, interoperability support, access and usage management, collection administration, and support for standards compliance like Dublin Core, Unicode and XML.

3.2 Digital library software:

Few of the software options that are considered for **creation and provision of access** to digital library collection are discussed below.

3.2.1 Open source digital library software:

Several free digital library software packages are now available, enabling easy creation and sharing of information through digital library collections. Example packages include:

1. **Greenstone Digital Library Software**

   New Zealand Digital Library

   http://www.greenstone.org/

2. **ARNO**

   Academic Research in the Netherlands Online, Tilburg University, The Netherlands

   http://www.uba.uva.nl/arno

3. **CDSware**

   CERN Document Server Software (CDSware), CERN, Geneva, Switzerland

   http://cdsware.cern.ch/

4. **DSpace**

   MIT Libraries, Cambridge, MA USA

   http://www.dspace.org/

5. **Eprints**

   University of Southampton, U.K.

   http://software.eprints.org/

6. **Fedora digital object repository management system**

   University of Virginia, USA

   http://www.fedora.info/

7. **i-Tor**

   Tools and technologies for Open Repositories

   Netherlands Institute for Scientific Information Services

http://www.i-tor.org/en/toon

8. **MyCoRe**

Essen University Library, University of Duisburg-Essen,Germany

http://www.mycore.de/engl/index.html

All of these are available for Linux operating system platform, a freely available popular operating system. Most of these packages also support the OAI interoperability protocol, an attractive proposition to share digital library collection with other collections on the Internet.

3.2.2 Commercial digital library software:

Some examples of commercially available digital library software include:

1. ENCompass of Endeavor Information Systems (http://www.endinfosys.com).
2. DigiTool of Ex Libris (http://www.exlibris.co.il/, or (http://www.exlibris-usa.com).
3. Visual MIS (Multimedia and Imaging Solutions) of VTLS (http://www.vtls.com/).
4. TEAMS of Artesia Technologies (http://www.artesia.com/).
5. Insight of Luna Imaging (http://www.luna-imaging.com/)
6. MuseSearch of MuseGlobal, Inc. (http://www.museglobal.com/)

Onsite technical support is an advantage with commercial packages (may not be available in all places). On the downside these packages could be quite expensive. Further there is no possibility to extend their

functionality as these packages are proprietary and do not have access to source code.

3.2.3 Some Digital Library Software in Brief:

- ***Eprints Archive Software (EAS) :***

EPrints is generic archive software under development by the University of Southampton. It is intended to create a highly configurable web-based archive. EPrints primary goal is to be set up as an open archive for research papers, but it could be easily used for other things such as images, research data, audio archives - anything that can be stored digitally by making changes in configuration. It works on Linux O/s and it needs MySQL, Perl modules and Apache web server.

- ***CERN Document Server Software (CDSware) :***

The CERN Document Server Software (CDSware) is the software developed by, maintained by, and used at, the CERN Document Server. It allows us to run our own electronic preprint server, our own online library catalogue or a document system on the web. It complies with the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as its underlying bibliographic standard. The CDSware is free software, licensed under GNU General Public Licence (GPL).

- ***Open Archives Harvester :***

The PKP (Public Knowledge Project)Open Archives Harvester is a free metadata indexing system developed by the Public Knowledge Project through its federally funded efforts to expand and improve access to research. The PKP OAI Harvester allows us to create a

searchable index of the metadata from Open Archives Initiative (OAI)-compliant archives, such as sites using Open Journal Systems (OJS) or Open Conference Systems (OCS).

- ***Ibero-american and Caribbean Digital Library Project :***

Ibero-American and Caribbean Digital Library Project is an initiative between UNESCO, University of Colima on and other experts of the region. The methodology that the registry of resources of information (metadatos) of any type of digital object "books, journals, thesis, music, images, audio, video, etc. As well as the formats and techniques for the digitalization, Design and development of the Z39.50 standard for the exchange of information. The objective is to create a great digital library of the region that this free one of author rights. In order to obtain software http://bdigital.ucol.mx consults the page or the version in CD-ROM to ask for a copy to the person in charge.

- ***OCLC Site Search Open Source Project :***

The OCLC Site Search Software provides a comprehensive solution for managing distributed library information resources in a World Wide Web environment. It offers tools that integrate electronic resources under one web interface, provide flexible access to resources, and build text and image databases locally.

3.3 Greenstone:

3.3.1 Introduction:

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM in the form of a fully-searchable, metadata-driven digital library. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. Its developers received the 2004 IFIP Namur award for "contributions to the awareness of social implications of information technology, and the need for an holistic approach in the use of information technology that takes account of social implications." The company DL Consulting provides Greenstone commercial consulting, customization, support, maintenance and hosting services.

Greenstone runs on Windows, Unix and Mac OS X. The distribution includes ready-to-use binaries for all versions of Windows, and for Linux and Mac OS X. It also includes complete source code for the system, which can be compiled using Microsoft C++ or gcc. Greenstone works with associated software that is also freely available: the Apache Webserver and PERL. The user interface uses a Web browser: typically Netscape Navigator or Internet Explorer.

Greenstone is specifically designed to be highly extensible and customizable. New document and metadata formats are accommodated

by writing "plugins" (in Perl). Analogously, new metadata browsing structures can be implemented by writing "classifiers." The user interface look-and-feel can be altered using "macros" written in a simple macro language. A Corba protocol allows agents (e.g. in Java) to use all the facilities associated with document collections. Finally, the source code, in C++ and Perl, is available and accessible for modification.

Greenstone 3 is a complete redesign and reimplementation which retains all the advantages of Greenstone 2 (the current version)--for example, it is multilingual, multiplatform, and highly configurable. It incorporates all the features of the existing system, and is backwards compatible: that is, it can build and run existing collections without modification.

3.3.2 Overview of Greenstone:

Greenstone is a comprehensive system for constructing and presenting collections of thousands or millions of documents, including text, images, audio and video.

3.3.3 Collections:

A typical digital library built with Greenstone will contain many collections, individually organized—though they bear a strong family resemblance. Easily maintained, collections can be augmented and rebuilt automatically.

There are several ways to find information in most Greenstone collections. For example, we can *search for particular words* that appear in the text, or within a section of a document. We can *browse documents by title*: just click on a book to read it. We can *browse documents by subject*. Subjects are represented by bookshelves:

just click on a bookshelf to look at the books. Where appropriate, documents come complete with a table of contents: we can click on a chapter or subsection to open it, expand the full table of contents, or expand the full document into Ourr browser window (useful for printing).

3.3.4 Finding information:

Greenstone constructs full-text indexes from the document text—that is, indexes that enable searching on any words in the full text of the document. Indexes can be searched for particular words, combinations of words, or phrases, and results are ordered according to how relevant they are to the query.

In most collections, descriptive data such as author, title, date, keywords, and so on, is associated with each document. This information is called *metadata*. Many document collections also contain full-text indexes of certain kinds of metadata.

Metadata forms the raw material for browsing. It must be provided explicitly or be derivable automatically from the documents themselves. Different collections offer different searching and browsing facilities. Indexes for both searching and browsing are constructed during a "building" process, according to information in a collection configuration file.

3.3.5 Document formats:

Source documents come in a variety of formats, and are converted into a standard XML form for indexing by "plugins." Plugins distributed with Greenstone process plain text, HTML, WORD and PDF documents, and Usenet and E-mail messages. New ones can be

written for different document types. To build browsing structures from metadata, an analogous scheme of “classifiers” is used. These create browsing indexes of various kinds: scrollable lists, alphabetic selectors, dates, and arbitrary hierarchies. Again, Greenstone programmers can create new browsing structures.

3.3.6 Multimedia and multilingual documents:

Collections can contain text, pictures, audio and video. Non-textual material is either linked into the textual documents or accompanied by textual descriptions to allow full-text searching and browsing.

Unicode, which is a standard scheme for representing the character sets used in the world's languages, is used throughout Greenstone. This allows any language to be processed and displayed in a consistent manner. Collections have been built containing Arabic, Chinese, English, French, M 0Å 1ori and Spanish. Multilingual collections embody automatic language recognition, and the interface is available in all the above languages (and more).

3.3.7 Making Greenstone Collections:

The simplest way to build new collections is to use Greenstone's “librarian” interface (GLI). This allows us to collect sets of documents, import or assign metadata, and build them into a Greenstone collection. It supports five basic activities, which can be interleaved but are nominally undertaken in this order:

1. Copy documents from the computer's file space, including existing collections, into the new collection. Any existing metadata remains “attached” to these documents. Documents may also be gathered

from the web through a built-in mirroring facility.

2. Enrich the documents by adding further metadata to individual documents or groups of documents.
3. Design the collection by determining its appearance and the access facilities that it will support.
4. Build the collection using Greenstone.
5. Preview the newly created collection, which will have been installed on Our Greenstone home page as one of the regular collections.

Finally, an alternative way of building collections is provided by the Collector, which helps us to create new collections, modify or add to existing ones, or delete collections. It predates the librarian interface, and for most practical purposes the librarian interface should be used instead of the Collector.

3.3.8 The librarian's interface:

For expository purposes, the walkthrough takes the form of a single pass through the steps listed above. A more realistic pattern of use, however, is for users to switch back and forth through the various stages as the task proceeds. The librarian interface can be run in one of four modes: Librarian Assistant, Librarian, Library Systems Specialist, and Expert. Modes control the level of detail within the interface, and can be changed through 'Preferences' in the 'File' menu.

3.3.9 Working with existing collections:

When we enter the Collector, we have to specify whether we want to create an entirely new collection or work with an existing

one, adding data to it or deleting it. By creating all searching and browsing structures automatically from the documents themselves Greenstone makes it easy to add new information to existing collections. Because no links are inserted by hand, when new documents in the same format become available they can be merged into the collection automatically.

3.3.10 Add new data:

The files that we specify will be added to the collection. Make sure that we do not re-specify files that are already in the collection—otherwise two copies will be included. Files are identified by their full pathname, web pages by their absolute web address. We specify directories and files just as we do when building a new collection.

If we add data to a collection and for some reason the building process fails, the old version of the collection remains unchanged.

3.3.11 Edit configuration file:

Advanced users can edit the collection configuration file, just as they can when a new collection is built.

3.3.12 Delete the collection:

We will be asked to confirm whether we really want to delete the collection. Once deleted, Greenstone can not bring the collection back!

3.3.13 Export the collection:

We can export the collection in a form that allows it to be written to a self-contained, self-installing Greenstone CD-ROM for Windows. Because commercial software that creates self-installing CD-ROMs is expensive, this facility includes a homegrown installer module.

When we export the collection, the dialogue informs us of the directory

name in which the result has been placed. The entire contents of the directory should be written on to CD-ROM using a standard CD-writing utility.

3.4 DSpace Durable Digital Depository:

DSpace is a digital repository created to capture, distribute and preserve the intellectual output of MIT. As a joint project of MIT Libraries and the Hewlett-Packard Company, DSpace provides stable long-term storage needed to house the digital products of MIT faculty and researchers. For the user: DSpace enables easy remote access and the ability to read and search DSpace items from one location: the World Wide Web. For the contributor: DSpace offers the advantages of digital distribution and long-term preservation for a variety of formats including text, audio, video, images, datasets and more. Authors can store their digital works in collections that are maintained by MIT communities. For the institution: DSpace offers the opportunity to provide access to all the research of the institution through one interface. The repository is organized to accommodate the varying policy and workflow issues inherent in a multi-disciplinary environment. Submission workflow and access policies can be customized to adhere closely to each community's needs.

3.4.1 DSpace Definition, Features and Functionality:

It is an attempt to address a problem that MIT faculty have been expressing to the Libraries for the past few years. As faculty and other researchers develop research materials and scholarly publications in increasingly complex digital formats, there is a need to collect, preserve, index and distribute them: a time-consuming and expensive chore for

individual faculty and their departments, labs, and centers to manage themselves. The DSpace system provides a way to manage these research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time.

DSpace was built breadth-first: it supports every function that a research organization needs to run a production digital repository service, but as simply as possible. The project focus was on building a production quality system. It complements and was influenced by previous research in computer science and digital library architectures. With the help of developers at other institutions that adopt DSpace under its open source license, the Dspace team work to add features and improve the different functions of the system.

DSpace is designed to make participation by depositors easy. The system's information model is built around the idea of organizational "Communities"—natural sub-units of an institution that have distinctive information management needs. In the case of MIT (a large research university) "Communities" are defined to be the schools, departments, labs, and centers of the Institute. Each Community can adapt the system to meet its particular needs and manage the submission process itself.

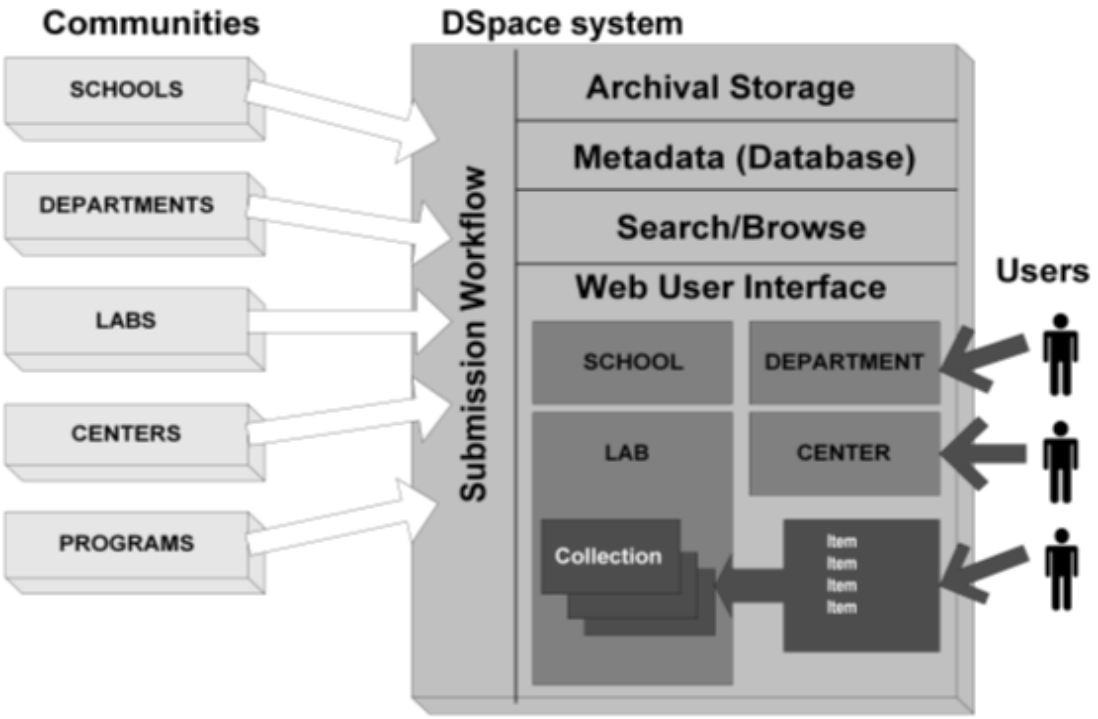


Figure 1: DSpace information model

3.4.2Metadata:

DSpace uses a qualified Dublin Core metadata standard for describing items intellectually (specifically, the Libraries Working Group Application Profile). Only three fields are required: title, language, and submission date, all other fields are optional. There are additional fields for document abstracts, keywords, technical metadata and rights metadata, among others. This metadata is displayed in the item record in DSpace, and is indexed for browsing and searching the system (within a collection, across collections, or across Communities). For the Dissemination Information Packages (DIPs) of the OAIS framework, the system currently exports metadata and digital material in a custom XML schema while we work with the METS community to develop the necessary extension schemas for the technical and rights metadata about arbitrary digital formats.

3.4.3 User Interface:

DSpace's current user interface is web-based. There are several interfaces: one for submitters and others involved in the submission process, one for end-users looking for information, and one for system administrators.

The end-user or public interface supports search and retrieval of items by browsing or searching the metadata. Once an item is located in the system, retrieval is accomplished by clicking a link that causes the archived material to be downloaded to the user's web browser. "Web-native" formats can be viewed immediately; others must be saved to the user's local computer and viewed with a separate program that can interpret the file (e.g., a Microsoft Excel spreadsheet, an SAS dataset, or a CAD/CAM file).

3.4.4 Workflow:

DSpace is the first open source digital repository system to tackle the complex problem of how to accommodate the differing submission workflows needed for a multidisciplinary system. In other words, different DSpace Communities, representing different schools, departments, research labs and centers, have very different ideas of how material should be submitted to DSpace, by whom, and with what restrictions. Who is allowed to deposit items? What type of items will they deposit? Who else needs to review, enhance, or approve the submission? To what collections can they deposit material? Who can see the items once deposited? All of these issues are addressed by the Community representatives, working together with the Libraries' DSpace user support staff, and are then modeled in a workflow for each collection to enforce their decisions. The system models "e-people" who have

"roles" in the workflow of a particular Community in the context of a given collection. Individuals from the Community are registered with DSpace, then assigned to appropriate roles.

3.4.5 Technology platform:

DSpace was developed to be open source, and in such a way that institutions and organizations with minimal resources could run it. The system is designed to run on the UNIX platform, and comprises other open source middleware and tools, and programs written by the DSpace team. All original code is in the Java programming language. Other pieces of the technology stack include a relational database management system (PostgreSQL), a Web server and Java servlet engine (Apache and Tomcat, both from the Apache Foundation), Jena (an RDF toolkit from HP Labs), OAICat from OCLC, and several other useful libraries. All leveraged components and libraries are also open source software. Libraries are bundled where possible. The system is available on SourceForge, linked from both the DSpace informational web site and the HP Labs site.

While DSpace is open source and freely available, neither MIT Libraries nor HP offer formal support for DSpace adopters. It is our assumption that institutions that use DSpace will have resources to use the system, including adequate hardware that runs the UNIX operating system, and a UNIX systems administrator to install and configure the system . Most institutions using DSpace will also want the services of a Java programmer who can localize and customize for them, or enhance it, although this is not absolutely necessary to run the system.

System Architecture:

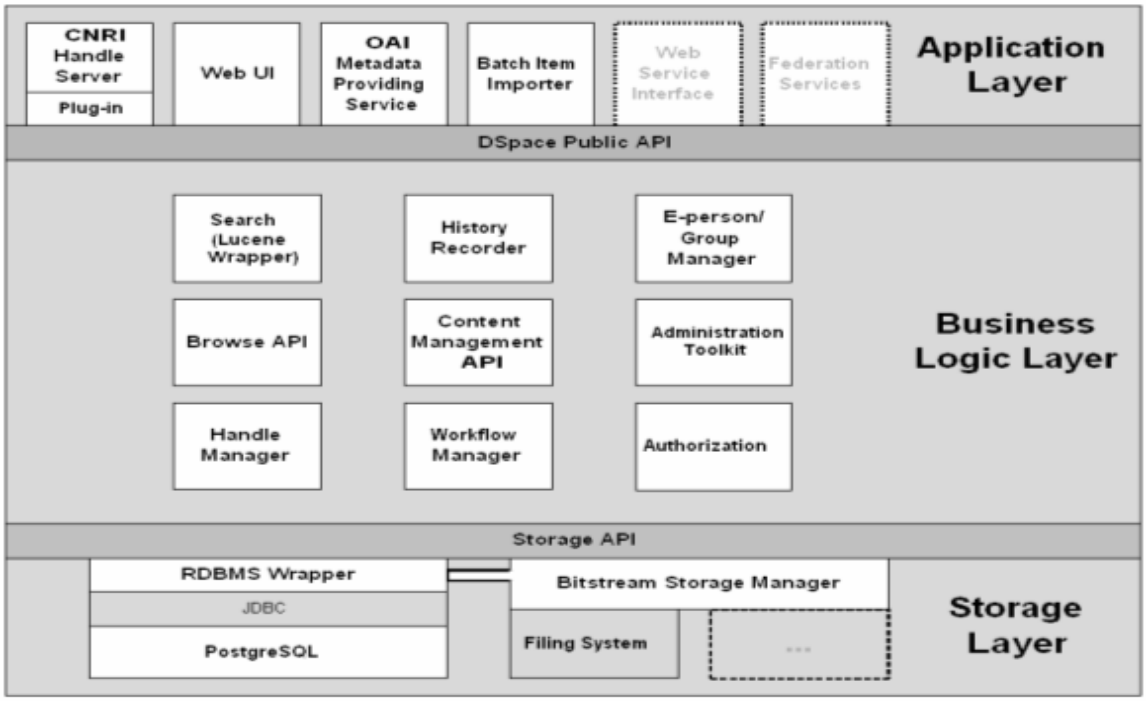


Figure 2: DSpace technical architecture

The DSpace architecture is a straightforward three-layer architecture, including storage, business, and application layers, each with a documented API to allow for future customization and enhancement. The storage layer is implemented using the file system, as managed by PostgreSQL database tables. The business layer is where the DSpace-specific functionality resides, including the workflow, content management, administration, and search and browse modules. Each module has an API to allow DSpace adopters to replace or enhance that function as desired. Finally, the application layer covers the interfaces to the system: the web UI and batch loader, in particular, but also the OAI support and Handle server for resolving persistent identifiers to DSpace items. This is the layer that will get much of the attention in future releases, as we add web services for new features and define Federation services across the range of institutions adopting DSpace.

3.5 Summary:

Large scale digitization projects are underway at Google, the Million Book Project, MSN, and Yahoo!. With continued improvements in book handling and presentation technologies such as optical character recognition, ebooks, and development of alternative depositories and business models, digital libraries are rapidly growing in popularity as demonstrated by Google, Yahoo!, and MSN's efforts. Just as libraries have ventured into audio and video collections, so have digital libraries such as the Internet Archive. So the digital library softwares are very playing an important role to develop repositories and there is need to improve existing free digital softwares and develop new ones which should be compatible to the present needs of the librarians and Institutions.

Model Institutional Repository
(Library and Information Science)

4.1 Introduction:

The Andhra University is a prestigious Institution having a lot of departments. The Dept. of Library and Information Science, A.U College of Arts & Commerce, is one of the oldest departments in A.U and also produced a very talented Human Resource. The Professors of this Dept. have been showing their presence in the Kingdom of Knwoledge through their articles in different Indian and Foreign Journals. I have collected all the articles of the Professors of Dept. of Library and Information Science and developed a Library and Information Science Repository on Dspace (Linux Debian Platform) and Greenstone (Windows Platform).

4.2 History of the Department:

The Department of Library and Information Science, Andhra University, had its origin in 1935. It grows from strength to strength under the leadership of Prof. K.Siva Reddy, who was the Head of the Department & Chairman, Board of Studies for long period. The services of eminent Professors, like Prof.G.V.S.L.N. Raju, Prof. K. Sarada, Prof. P. Soma Raju & Prof. B. Vijaya Lakshmi have helped the Department to attain the present status.

4.3 Installation and Customization

4.3.1 Installation of Dspace:

At first we should install Operating System. We have chosen **Debian (Linux)** Operating System for our project. After this the following steps should be done before installation of Dspace:

1. Create ***Dspace User***.
2. Install ***JDK***.
3. Create ***Postgresql*** user.
4. Configure ***pg_hba.conf*** and ***postgresql.conf*** files.
5. Create ***Dspace Database***.
6. Configure ***Dspace.cng*** file.
7. Copy ***jdbc drivers*** to ***lib*** folder.
8. Copy ***java_opts*** from ***Install.html*** file to ***bash rc***.

Then **Install Dspace**. The Following are the steps to install dspace.

- ✓ First restart the system and enter in ***dspace user name***.
- ✓ Open a terminal and type ***cd dspace 1.4.2 source.***
- ✓ Next type ***ant*** and press enter. It prompts ***Build Successful***.
- ✓ Type ***ant fresh_install*** and press enter. It prompts ***Build Successful***.
- ✓ Go to ***Dspace home*** -->***build*** and then copy two ***war*** files and paste it in ***Dspace home*** -> ***apache-tomcat-*** ->***webapps*** folder.

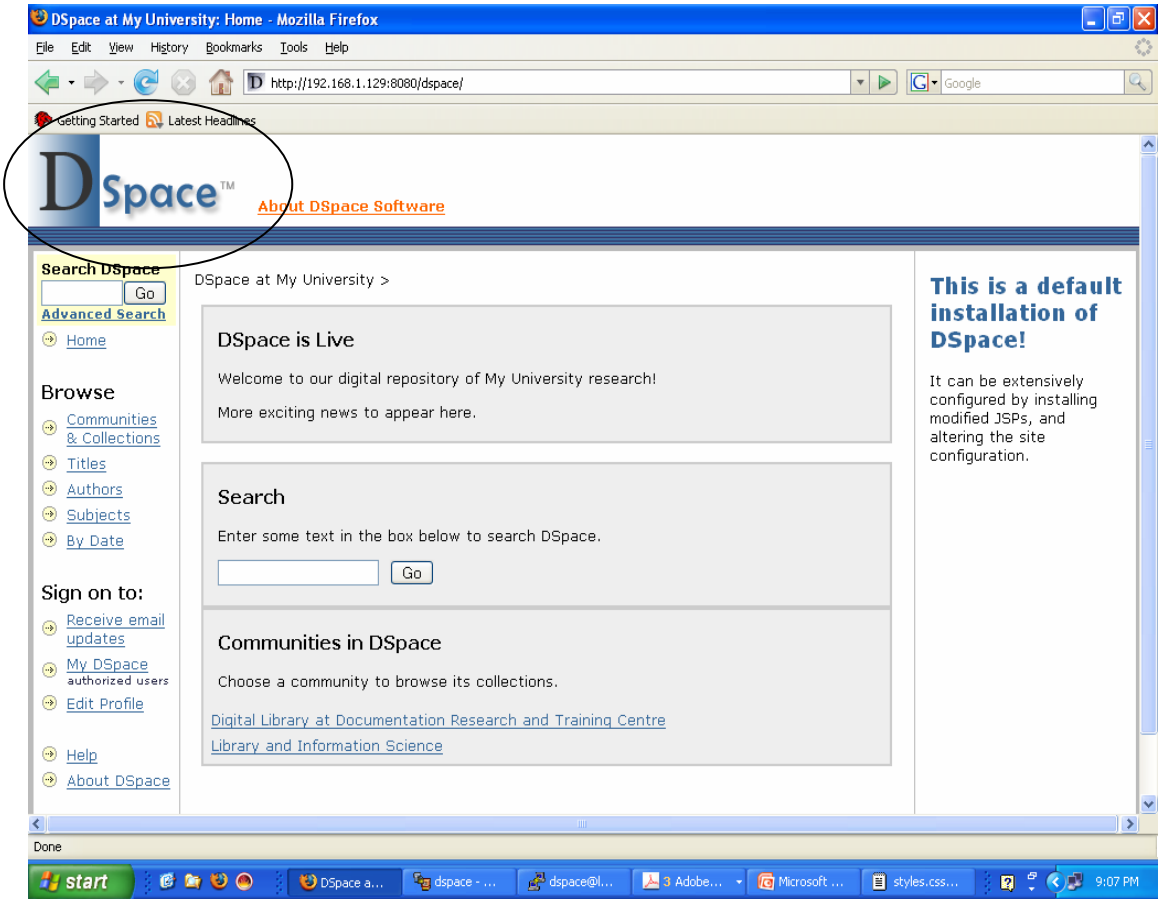
- ✓ Close the current terminal and open a new terminal.
- ✓ Type ***cd apache-tomcat 6.0.10/bin/*** and press enter.
- ✓ We should start the server. Type ***./startup.sh*** and press enter. This command asks the system to start ***server***.
- ✓ Open the browser and type the URL ***http://172.16.7.5:8080/dspace.***

4.3.2 Creation of Administrator:

- ✓ First shutdown the tomcat using the command ***./shutdown.sh.***
- ✓ Type ***cd bin/***
- ✓ Now we are in ***dspace@localhost :-/bin$***
- ✓ Type ***create administrator*** and press enter. Here it asks the details of Email address, First Name, Last Name and Password. After giving all these details it asks us the confirmation. We should type “Yes”.
- ✓ The administrator is created and we can start customization.

4.3.3 Customization of the site page:

The below figure is a default site page:



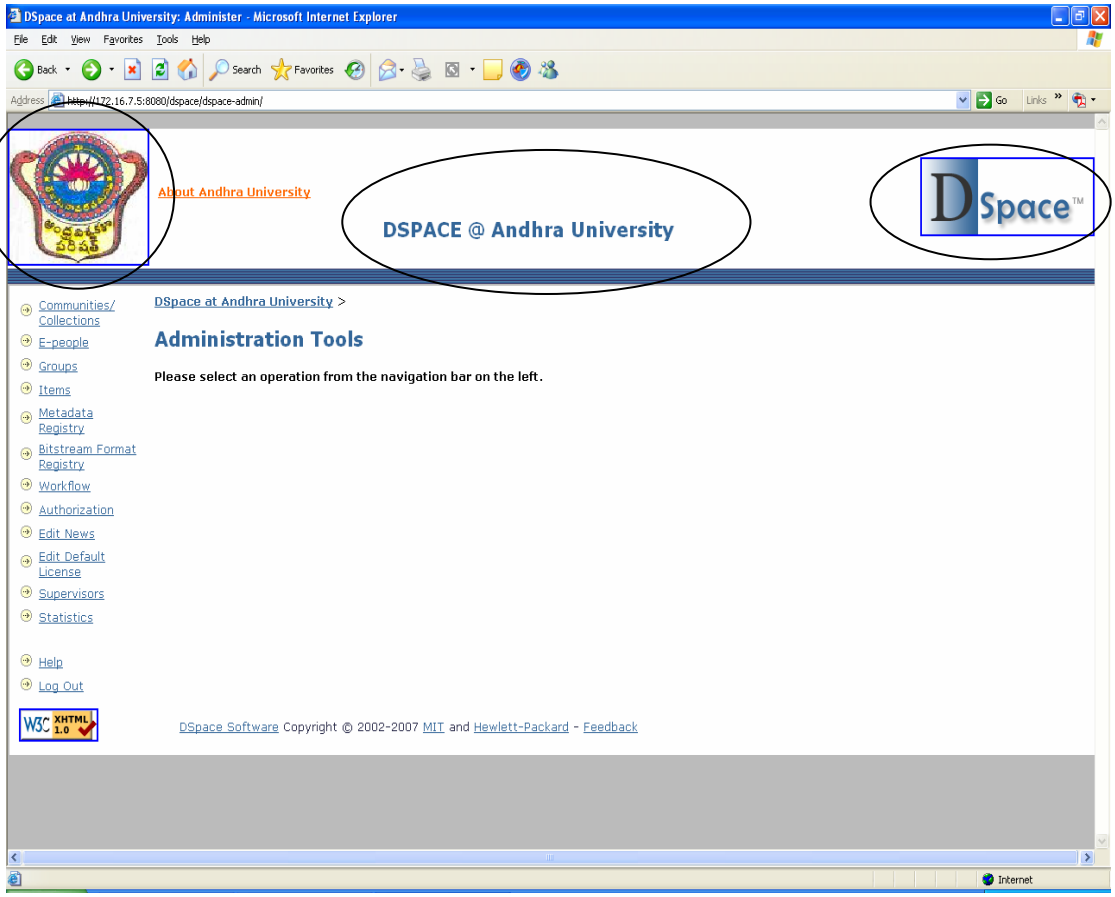
The following are the steps performed to change the banner of the site page:

- ✓ Open ***Dspace home-->apache-tomcat-6.0.10-->webapps-->dspace-->image***
- ✓ Here load the Andhra University Emblem image.
- ✓ Open ***Dspace home-->apache-tomcat-6.0.10-->webapps-->dspace-->layout/header-default.jsp*** change the settings and give Andhra University site address and image address. Here dspace-blue2.gif is the Andhra University emblem image.

These are the tags which are to be modified in the ***header-default.jsp*** file:

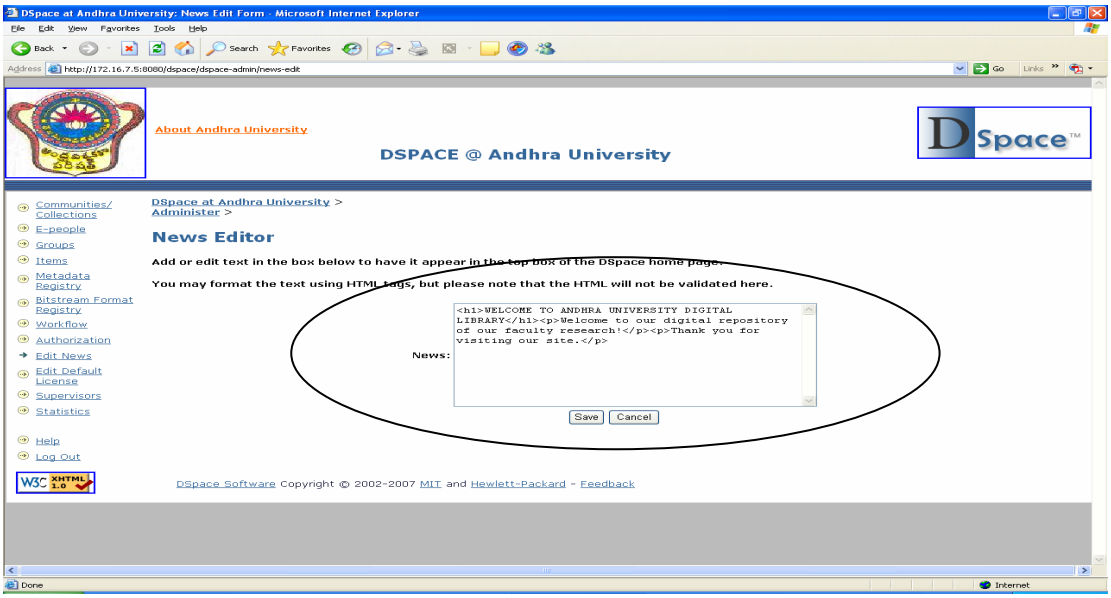
```
  <%-- DSpace top-of-page banner --%>
        <%-- HACK: width, border, cellspacing, cellpadding: for non-
CSS compliant Netscape, Mozilla browsers --%>

              <table class="pageBanner" width="100%" border="0"
cellpadding="0" cellspacing="0">

            <tr>

           <td>

        <a href="http://andhrauniversity.info"><img src="<%=
request.getContextPath() %>/image/dspace-blue3.jpg" alt="<fmt:message
key="jsp.layout.header-default.alt"/>" alt="Andhra University"
width="160" height="140" border="2"/></a>

</td>
                    <td class="tagLine" width="99%">
                    <a class="tagLineText" target="_blank"
href="http://andhrauniversity.info">About Andhra University</a>

<h1 align="center">DSPACE @ Andhra University</h1>

                </td>
    <td nowrap="nowrap" valign="middle">

                      <A
HREF="http://dspace.org/"><img src="<%= request.getContextPath()
%>/image/dspace-blue1.gif" alt="<fmt:message key="jsp.layout.header-
default.alt"/>" alt="DSpace" width="198" height="79" border="2"/></a>
<a class="tagLineText" target="_blank" href="http://dspace.org"></a>
</td>
                   </A>

                      
                </td>

            </tr>

            <tr class="stripe">

                <td colspan="3"> </td>
            </tr>

        </table>

        <table class="centralPane" width="100%" border="0"
cellpadding="3" cellspacing="1">

            <tr valign="top">

            <td class="navigationBar">

        <%-- Localization --%>
```

The changed settings are shown in Bold letters.The below figure is the page after changing the settings.



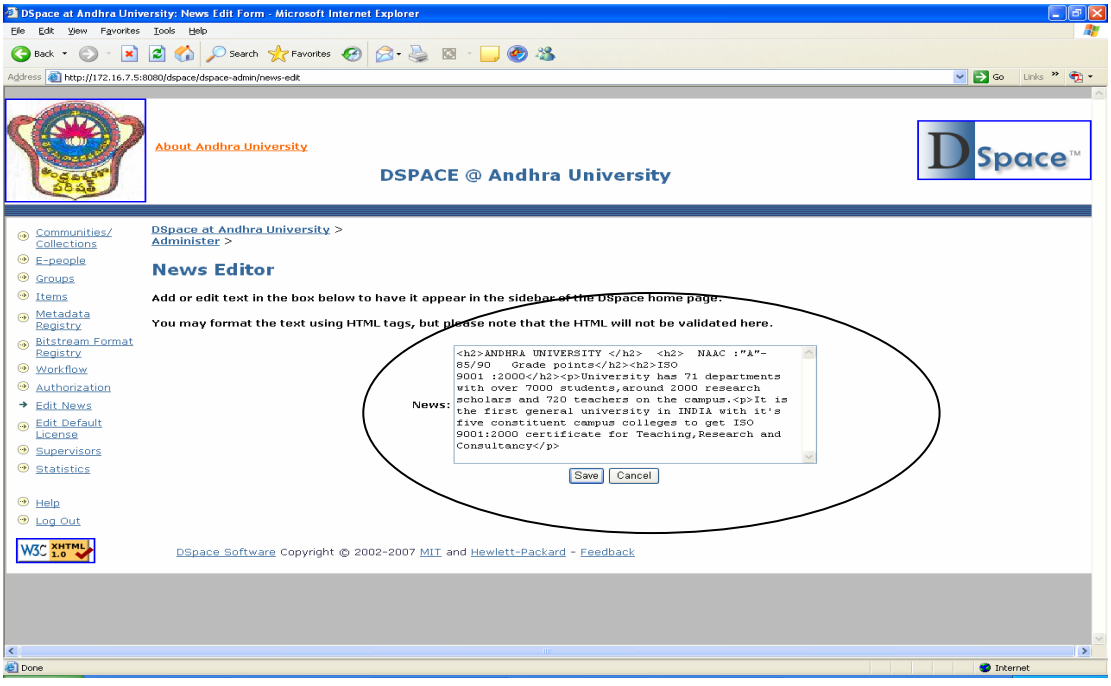
TOP NEWS Settings:

*<h1>WELCOME TO ANDHRA UNIVERSITY DIGITAL LIBRARY</h1> Welcome to our digital repository of our faculty research! Thank you for visiting our site. *

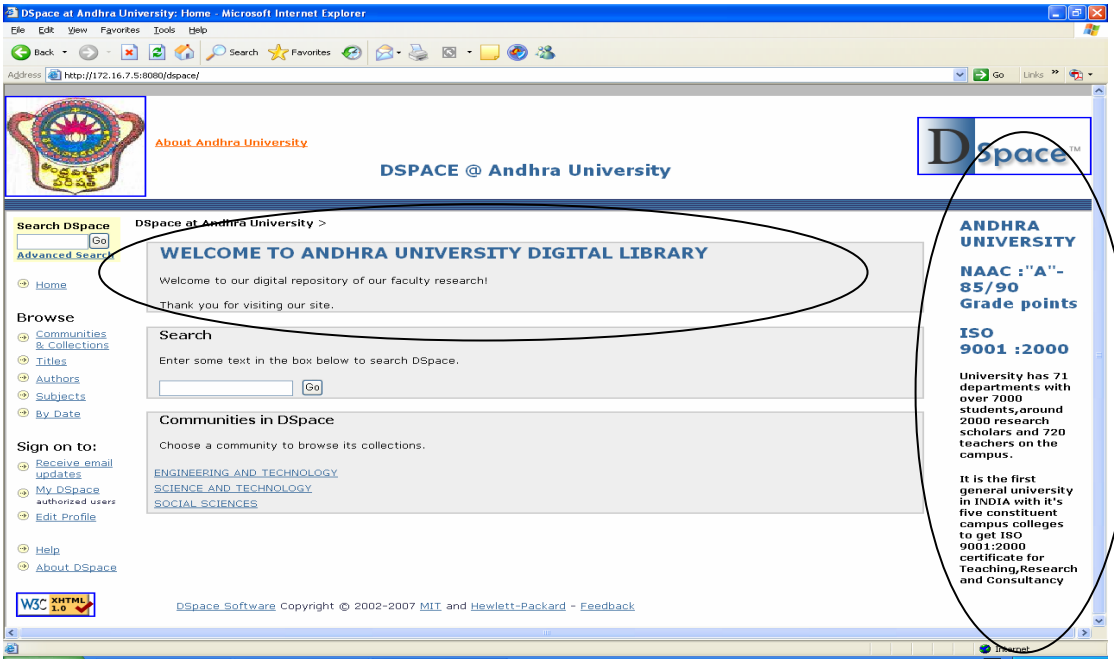


SIDE BAR settings:

*<h2>ANDHRA UNIVERSITY </h2> <h2> NAAC :"A"-85/90 Grade points</h2><h2>ISO 9001 :2000</h2> University has 71 departments with over 7000 students,around 2000 research scholars and 720 teachers on the campus. It is the first general university in INDIA with it's five constituent campus colleges to get ISO 9001:2000 certificate for Teaching,Research and Consultancy *



After changing the settings in the Sidebar News and Top News, the site page is:



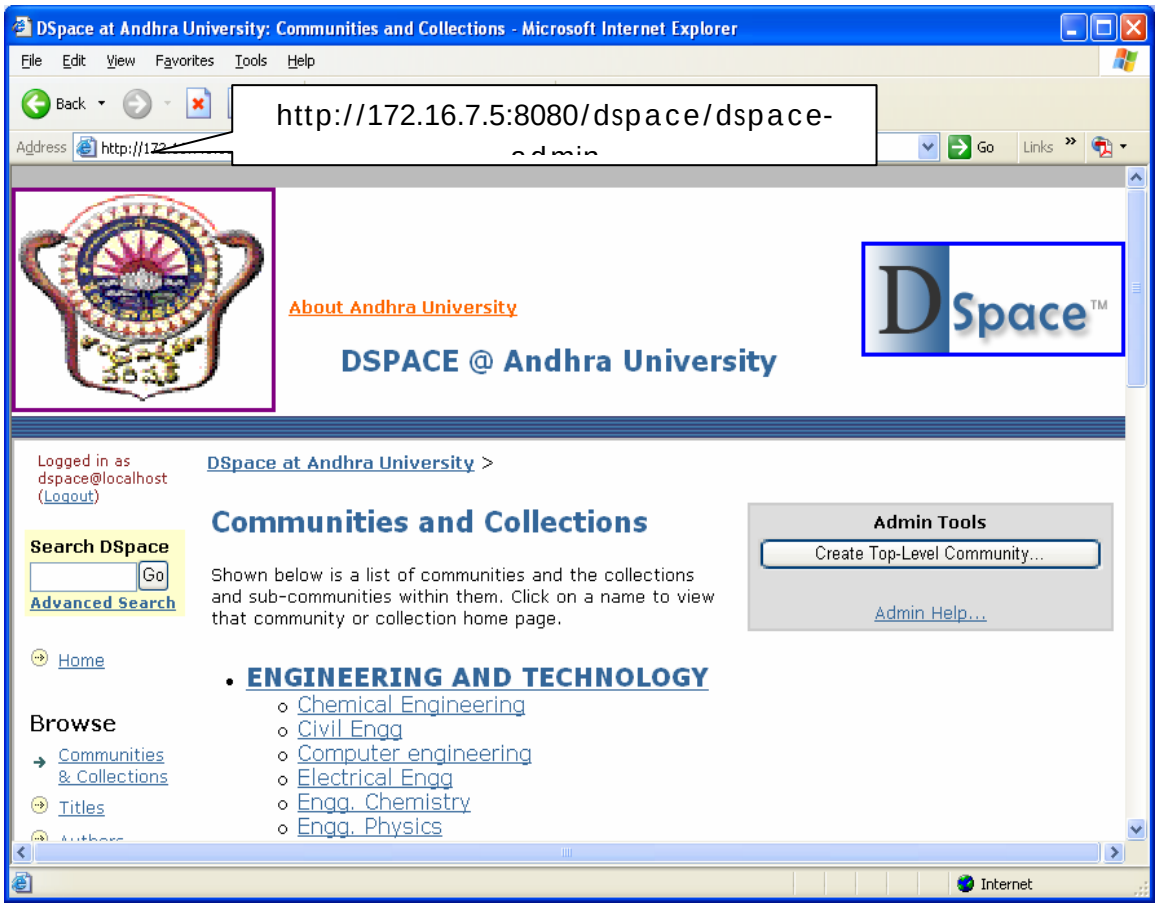
4.4 INSTITUTIONAL REPOSITORY

(LIBRARY &INFORMATION SCIENCE) USIG Dspace

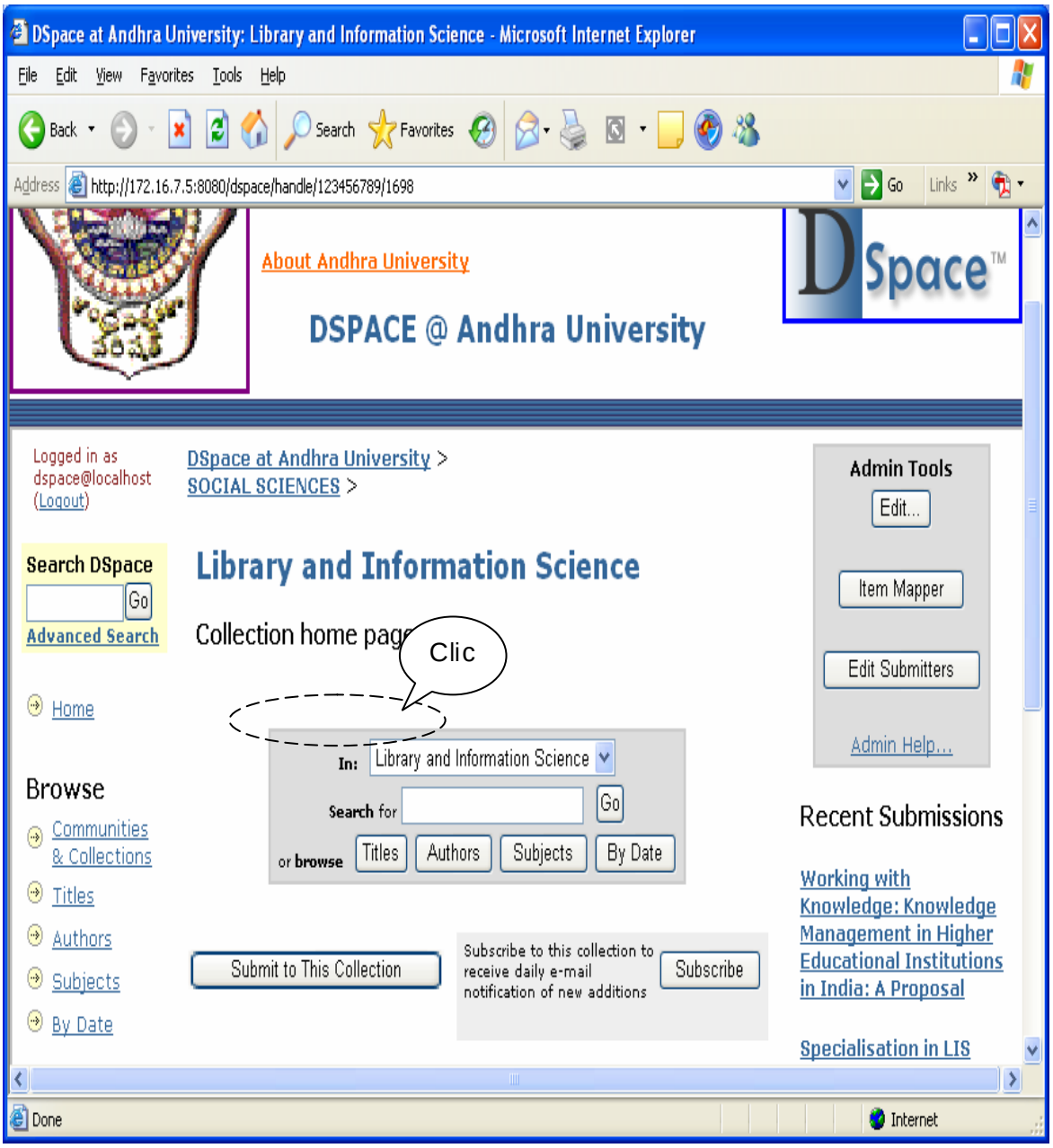
The Steps Involved in Developing LIS Repository in Dspace

Step1:Open the A.U Digital Library Website i.e.,

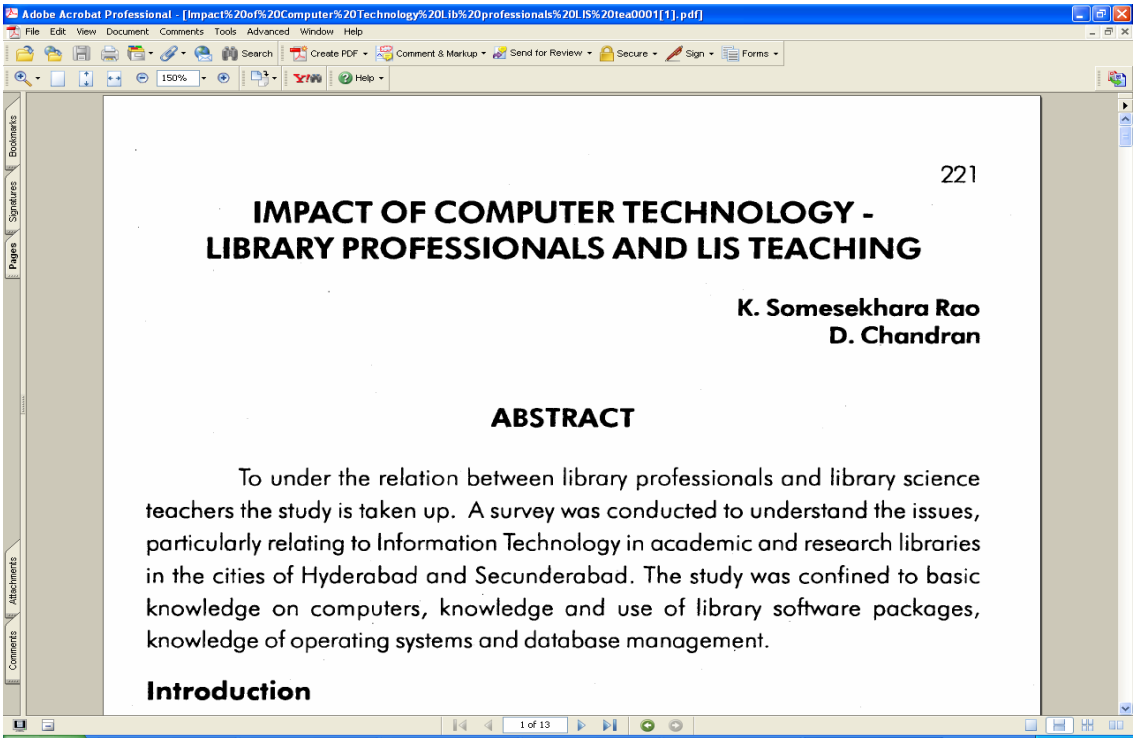
http://172.16.7.5:8080/dspace/dspace-admin in Administrators Mode.

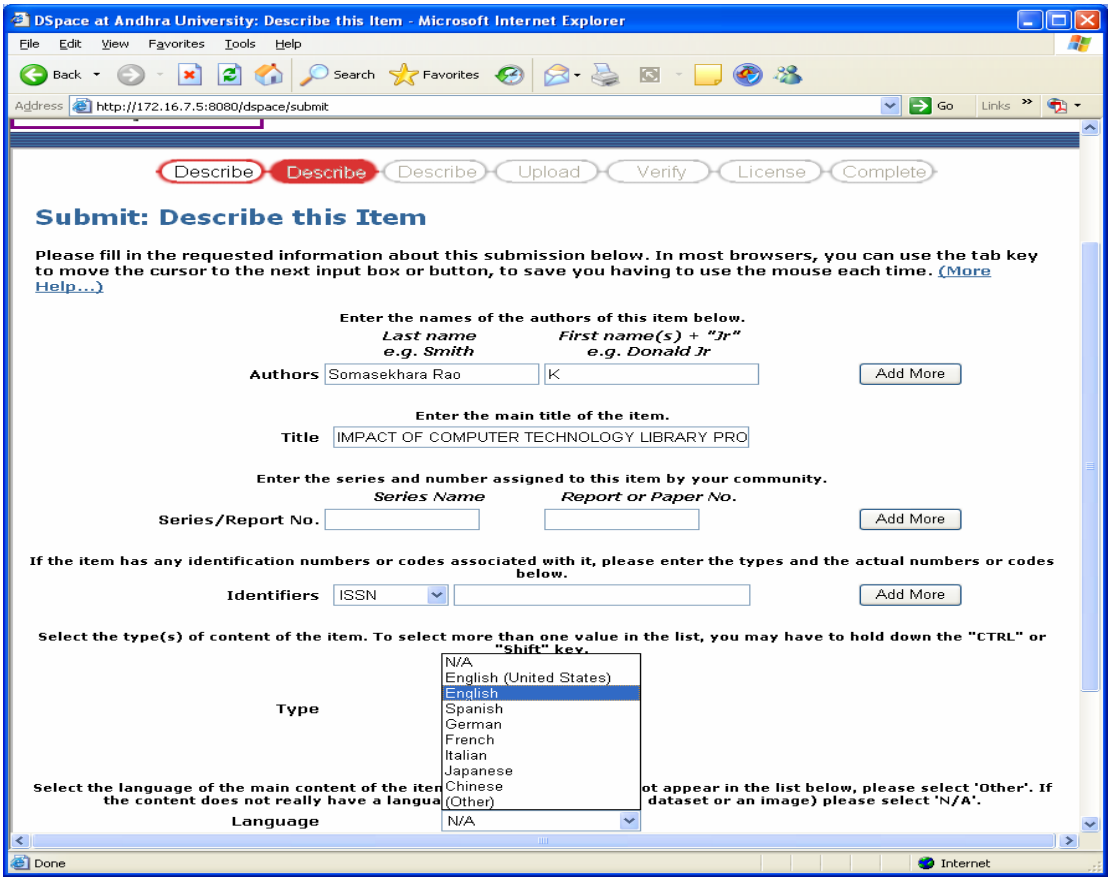


Step 2: Select the Physics ***Collection*** on Science and Technology ***Community*** and click the ***Submit to this collection***.

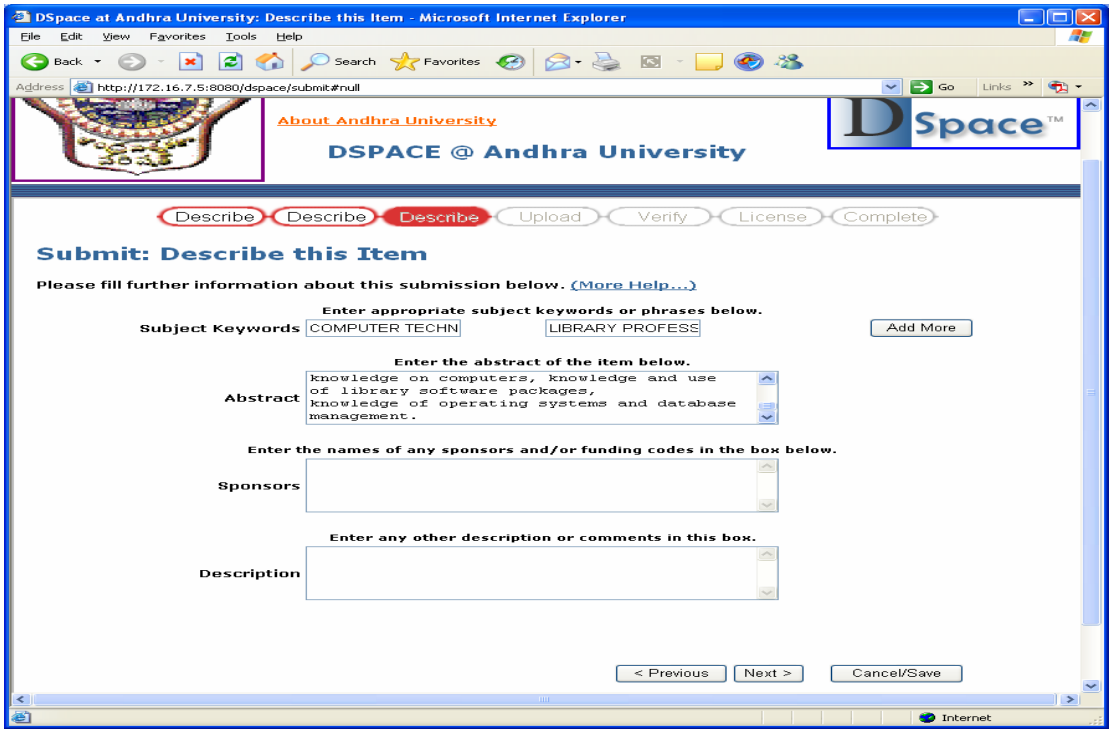


Step 3: Select the document and fill all the required items in ***Describe the Item*** menu.



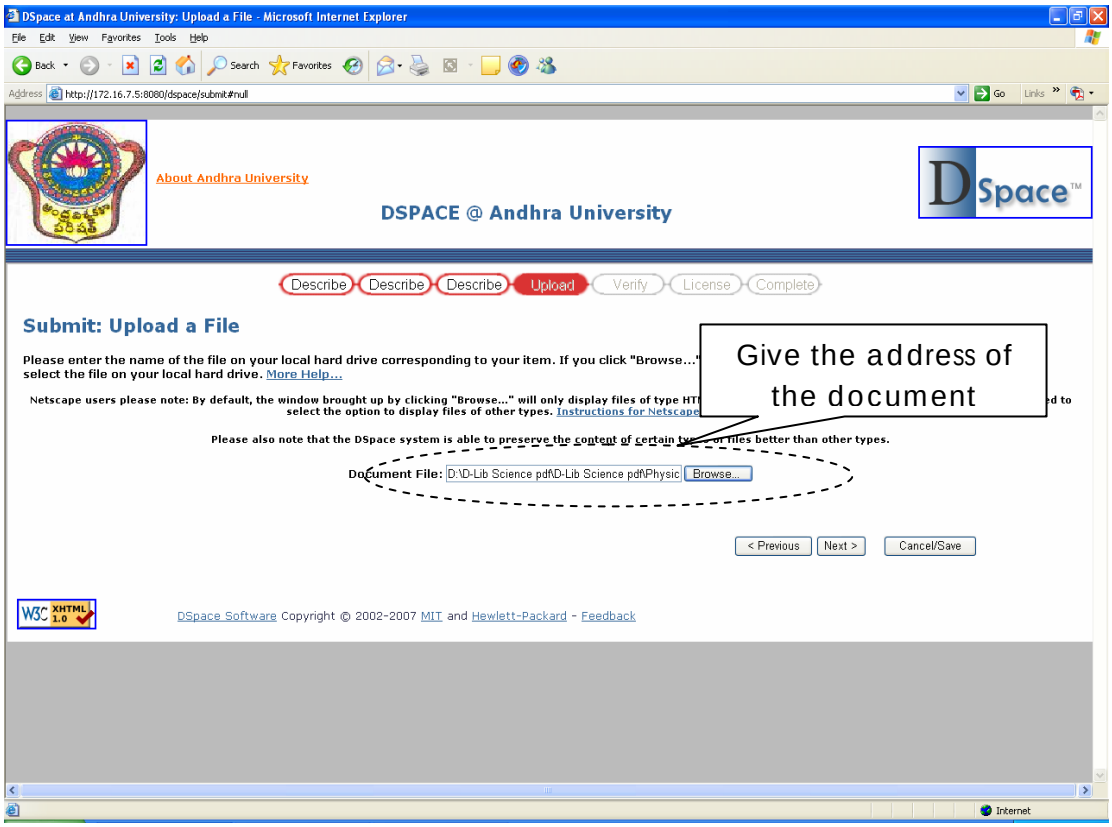


Click Next

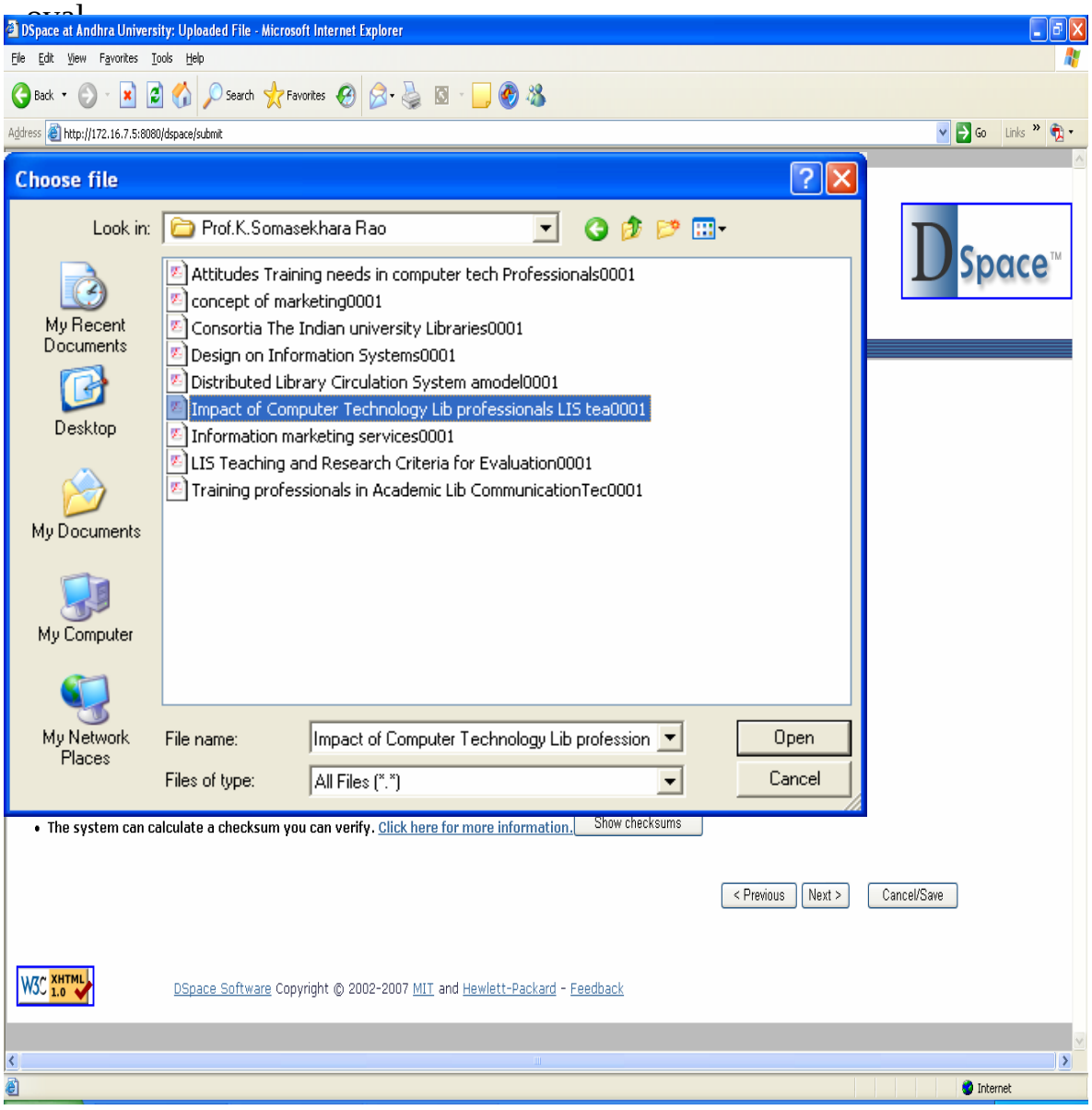


Click Submit

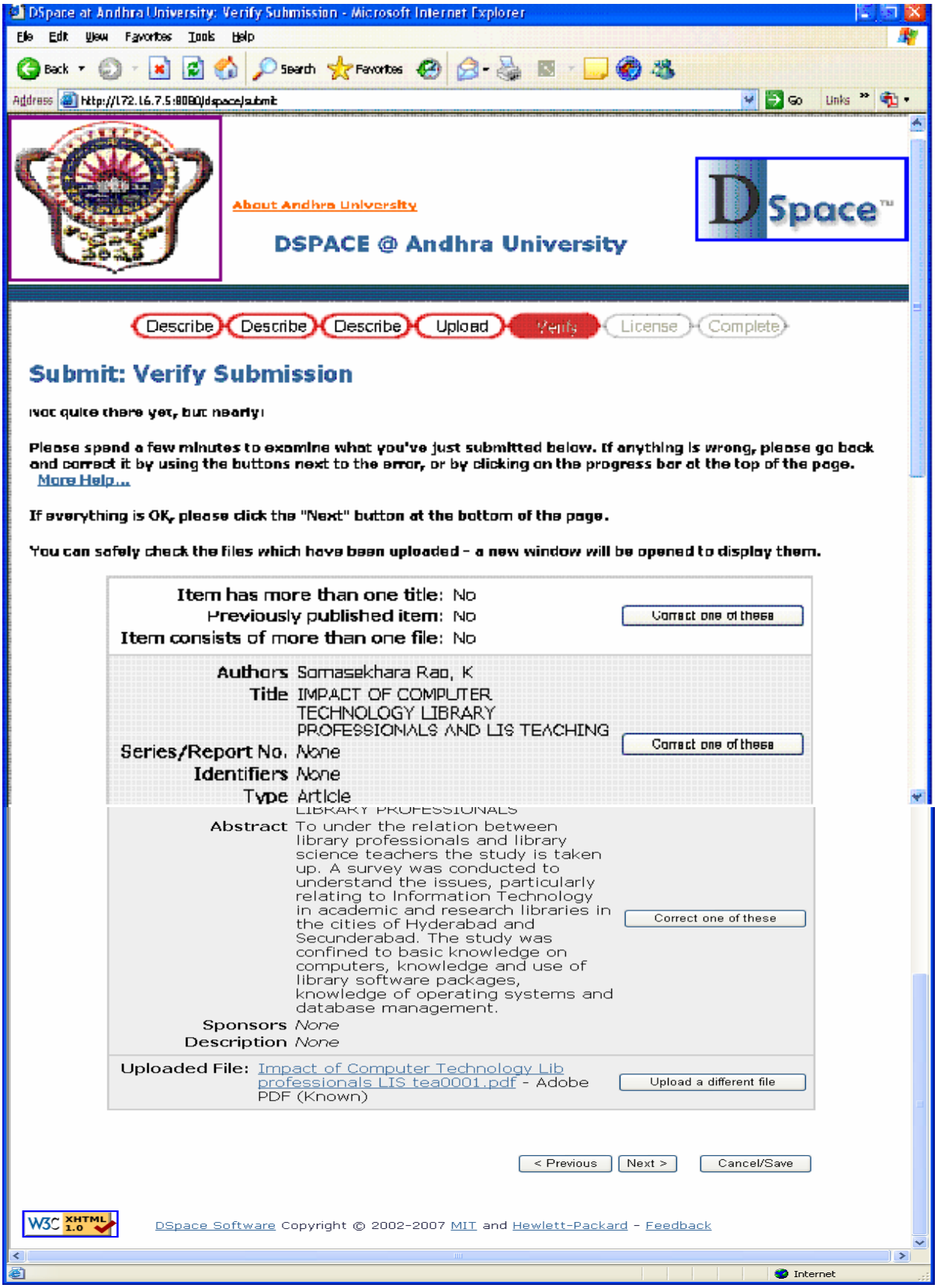
Step 4: Give the address of the document in ***Upload a File*** menu.



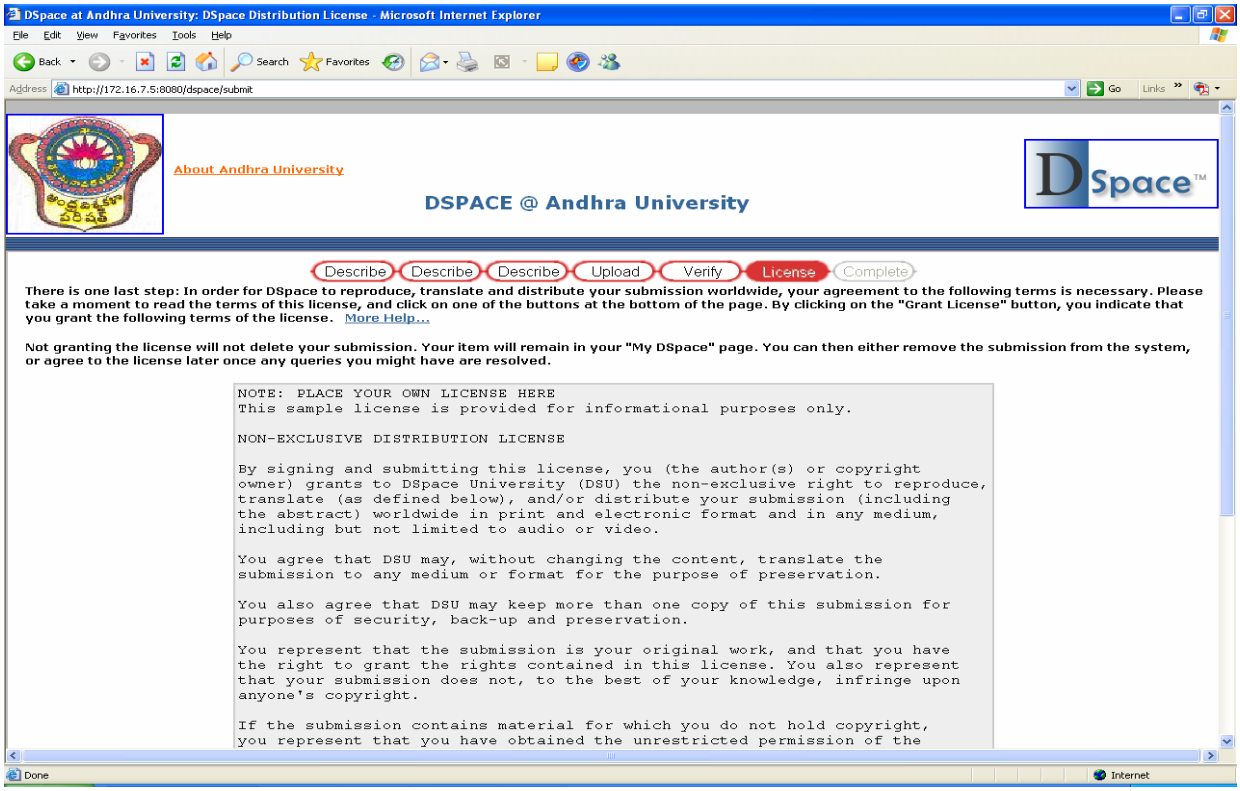
If the Address is correct then it gives the successful statement and shows the details like file name, size and file format which is shown in dotted oval

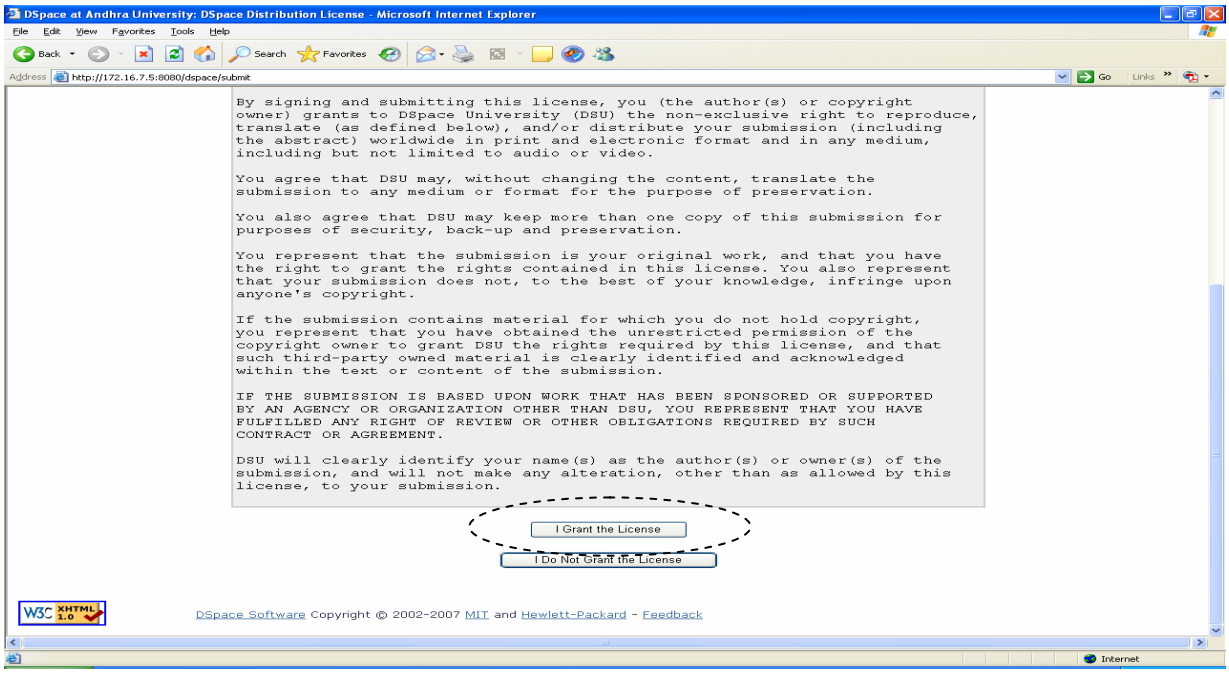


Step 5: Verify the Submission.



Step 6: Grant the License if you want to give the document for access.

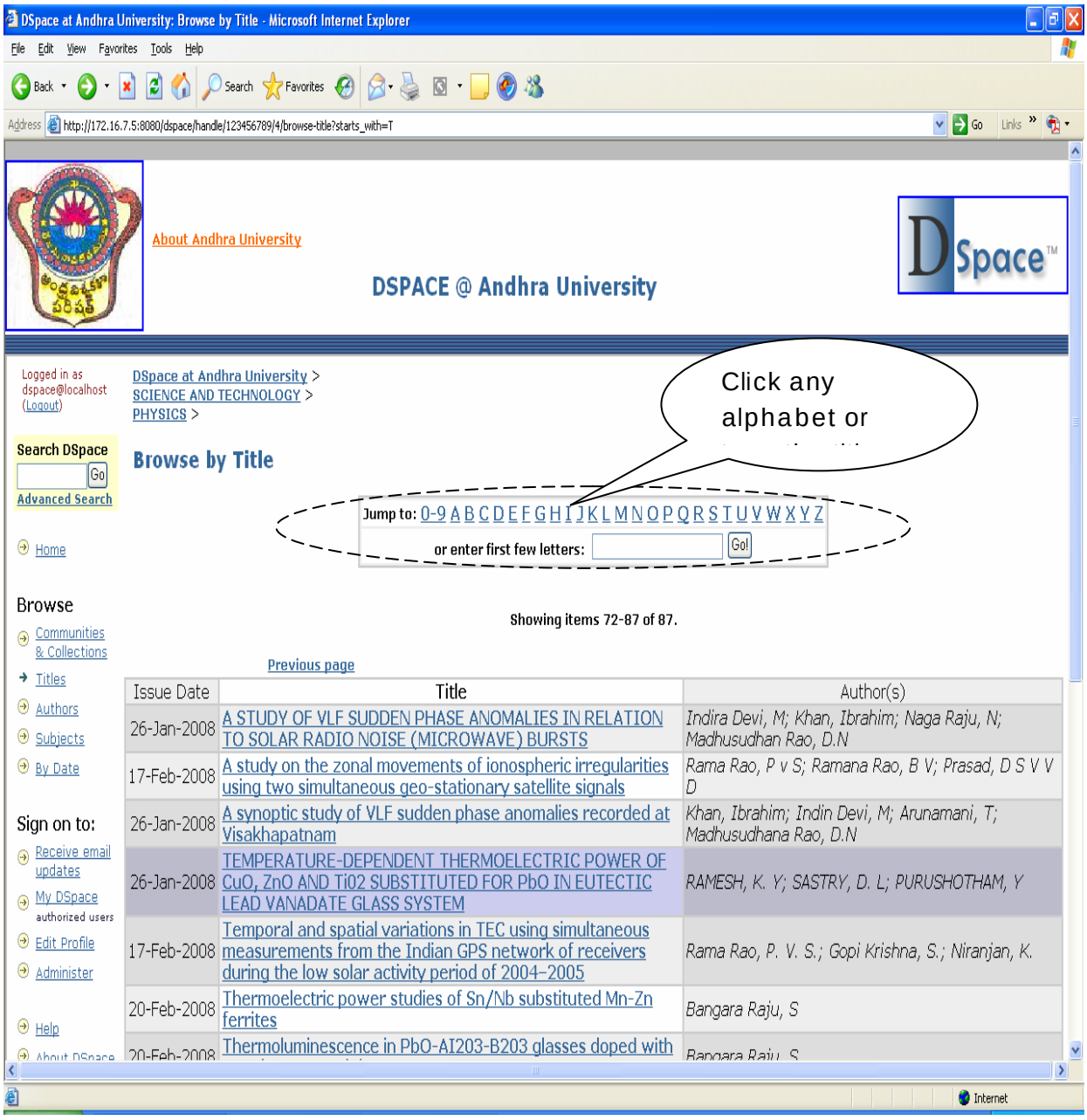




After granting, it shows the ***Submission Complete*** Statement and now the document is ready to access.

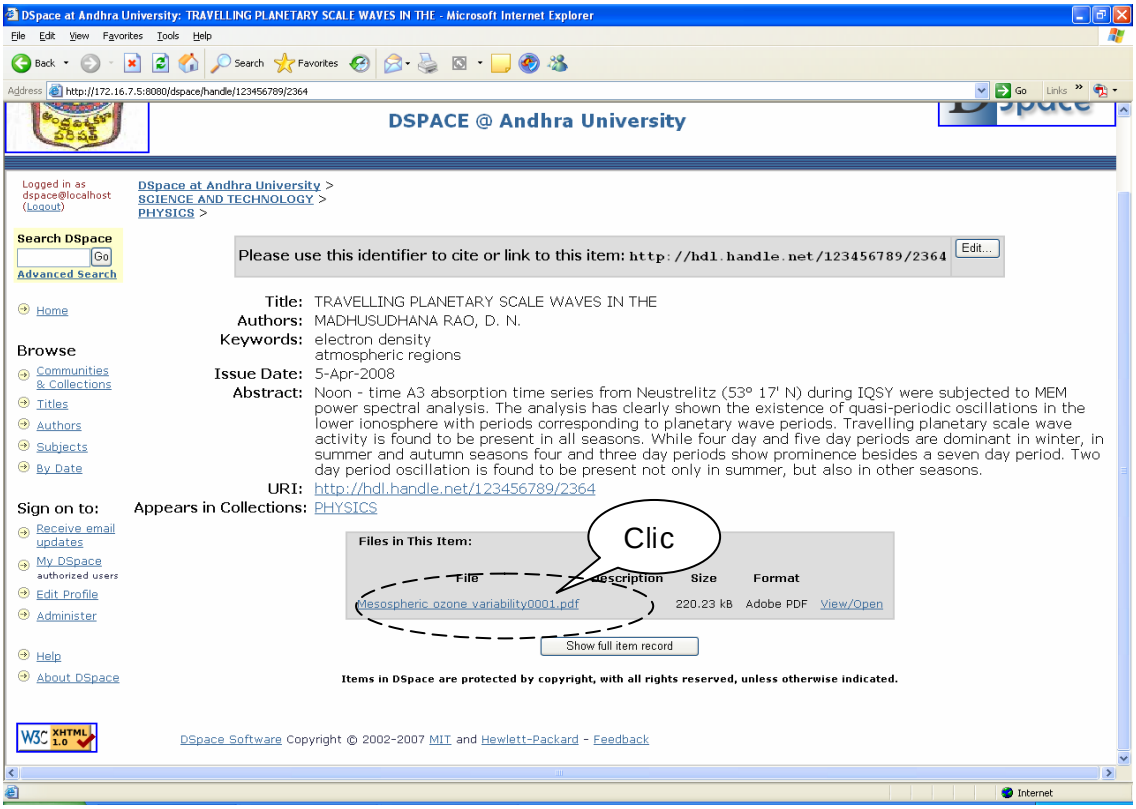
You can **Browse the document by Author, Key Words, Subject, Title and Date of submission.** It is shown in step by step.

Step 1: Click Title and choose the alphabet i.e., first letter of the title or type the title in the box.

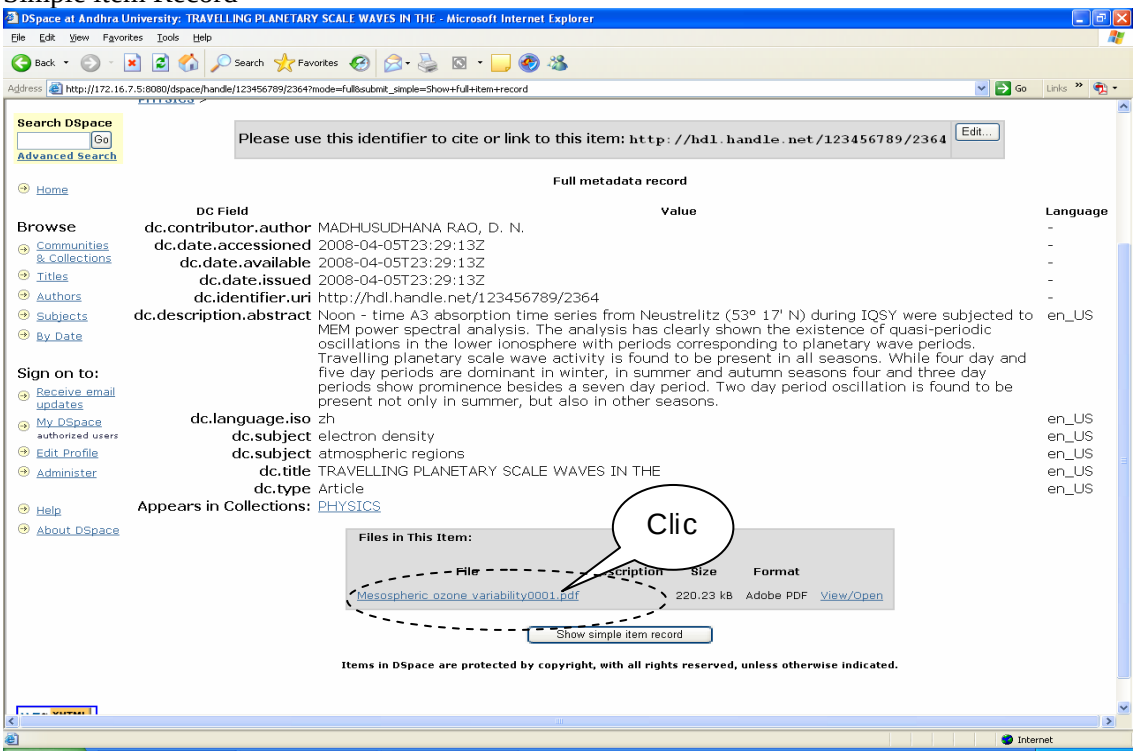


Step 2: Click the required item.

Step 3: You can see the simple or full Metadata. Click on the file for reading the document.

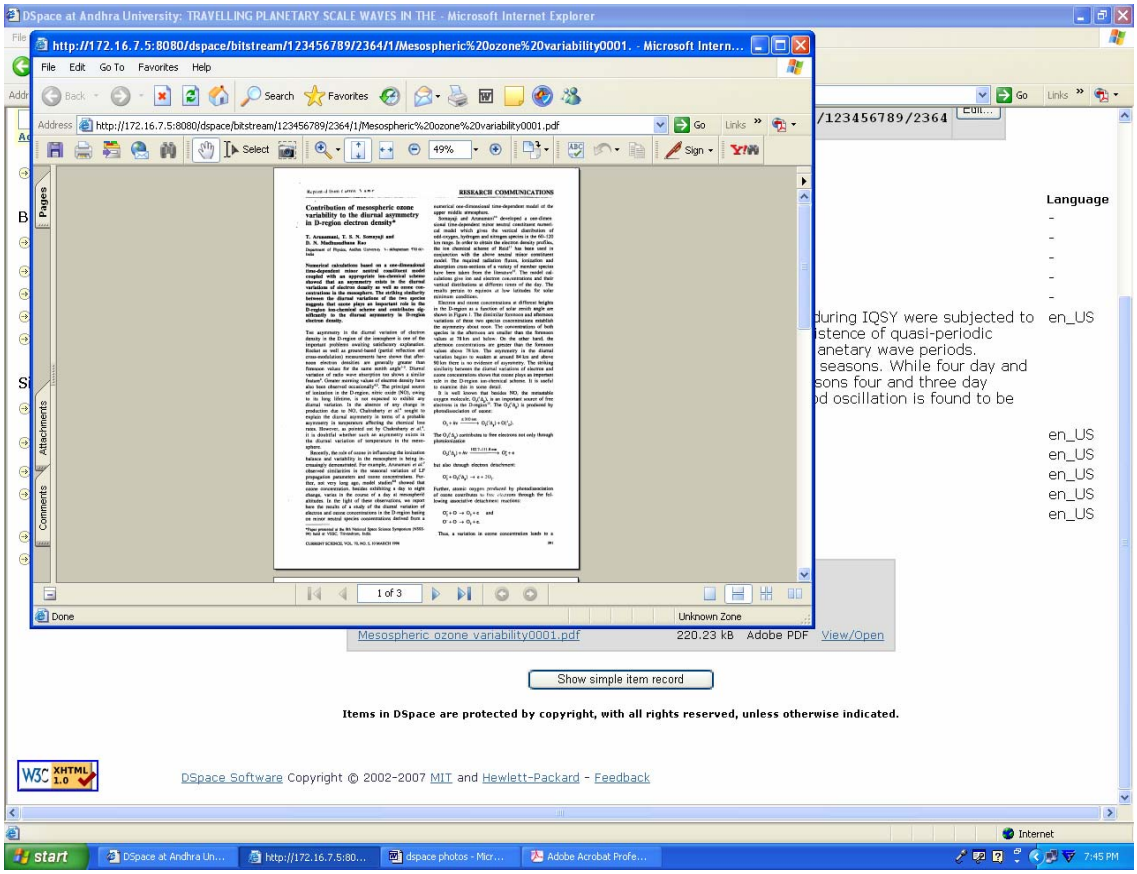


Simple item Record



Full item Record.

Step 4: The document is displayed.

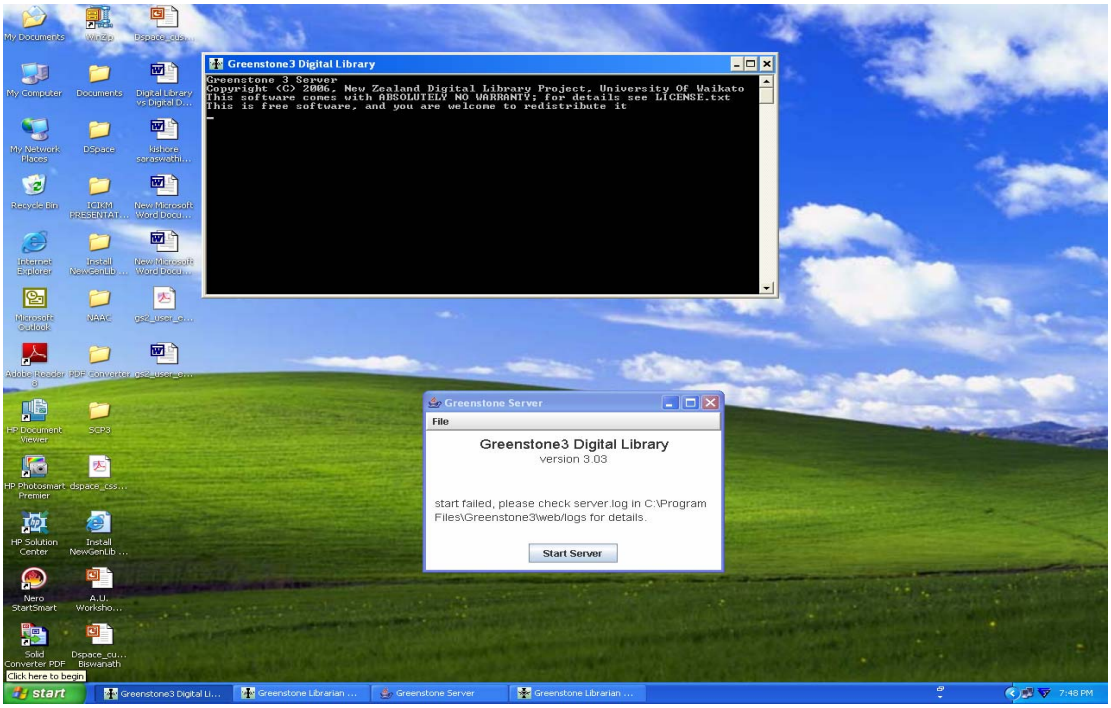


4.5 INSTITUTIONAL REPOSITORY

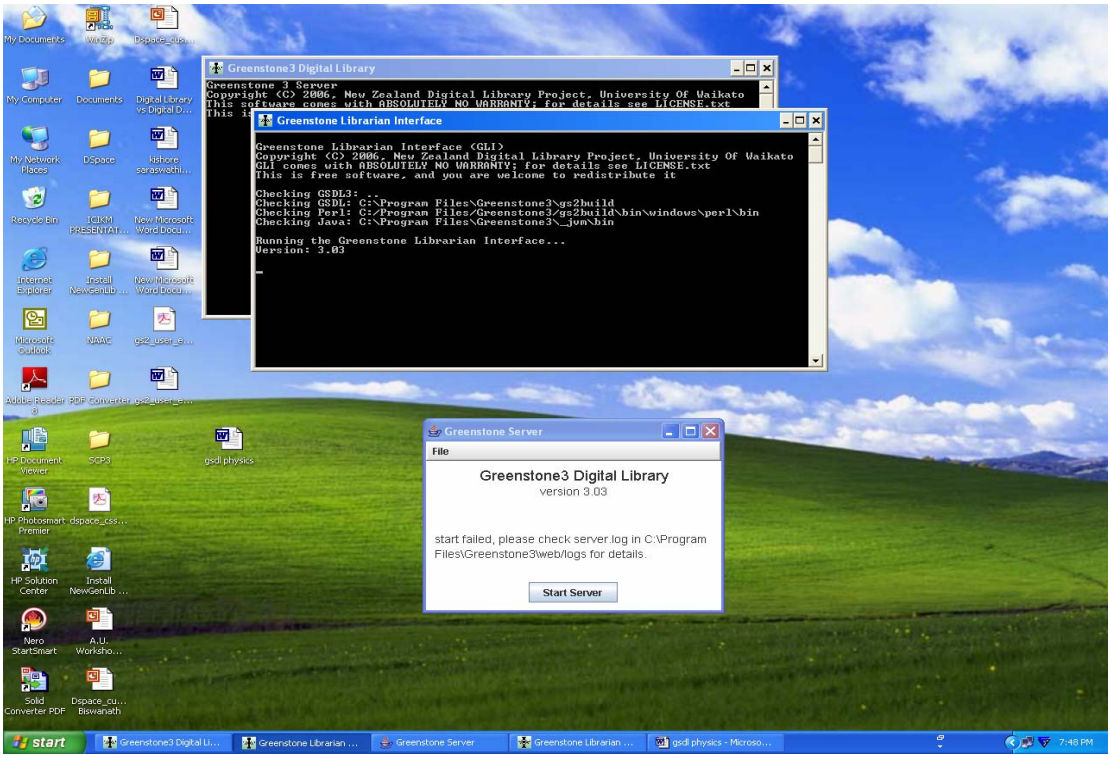
(LIBRARY &INFORMATION SCIENCE) USIG GSDL

The steps involved in developing LIS repository in GSDL

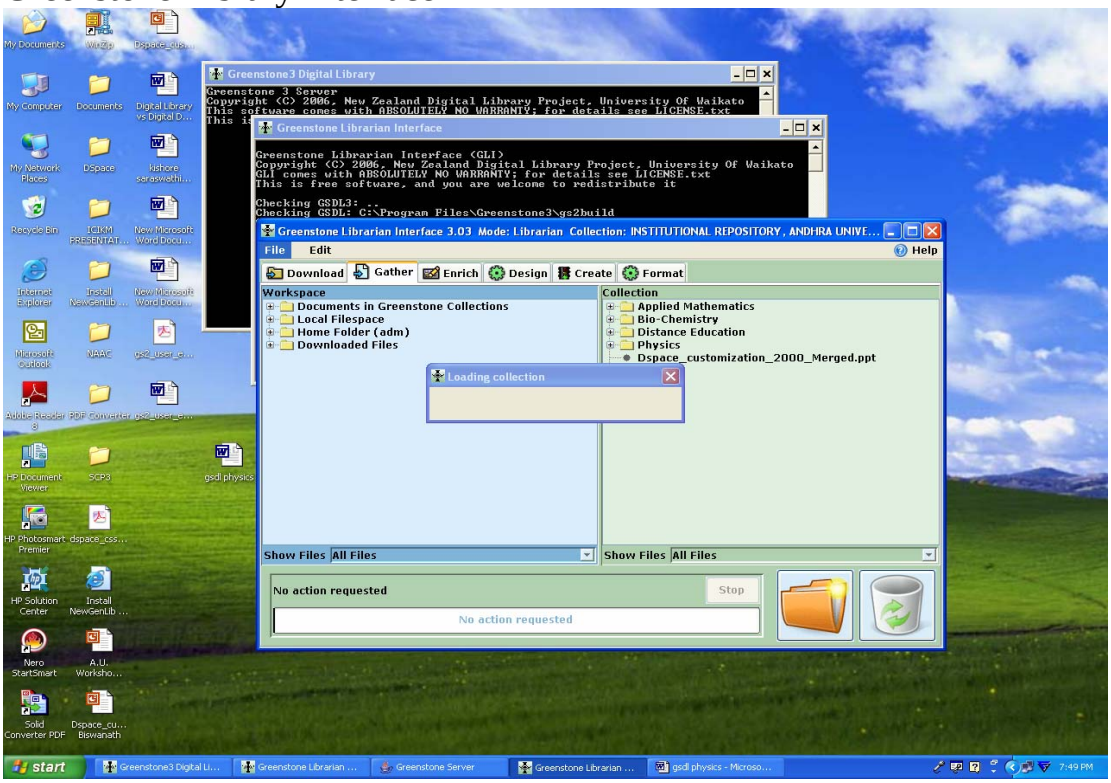
Step1:Open the Greenstone Digital Library, Greenstone Library Interface and start the server.



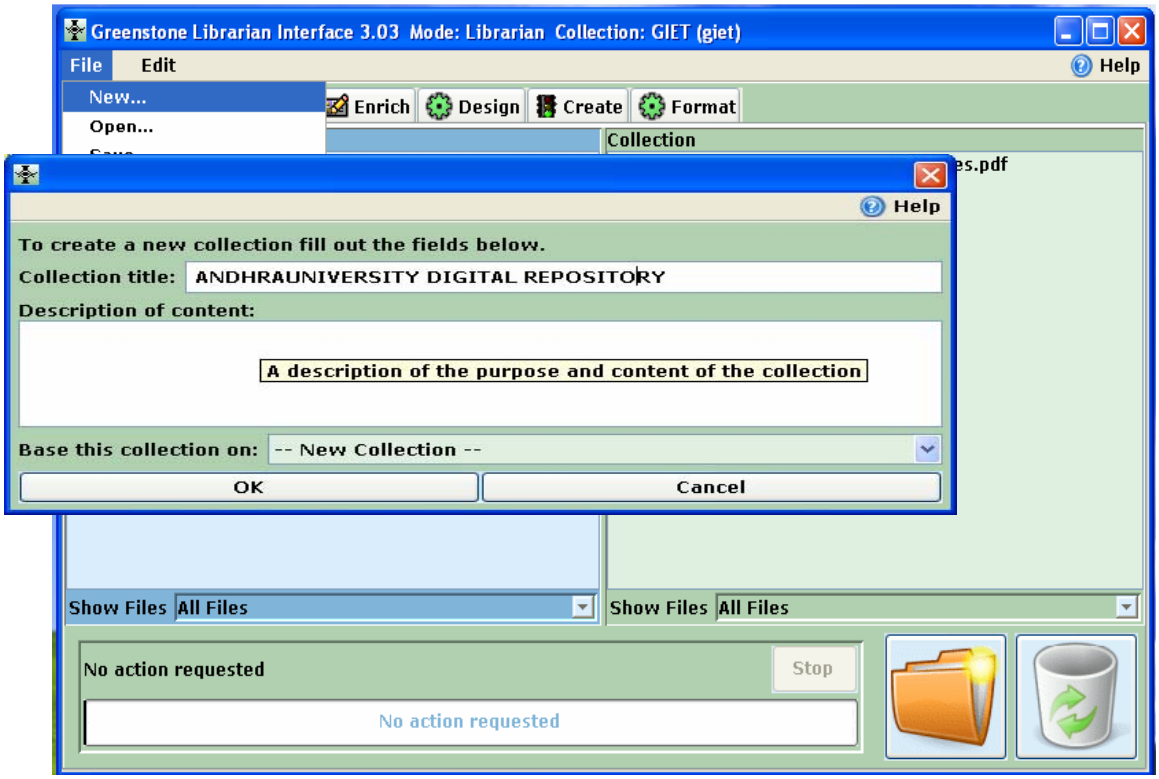
Greenstone Digital Library



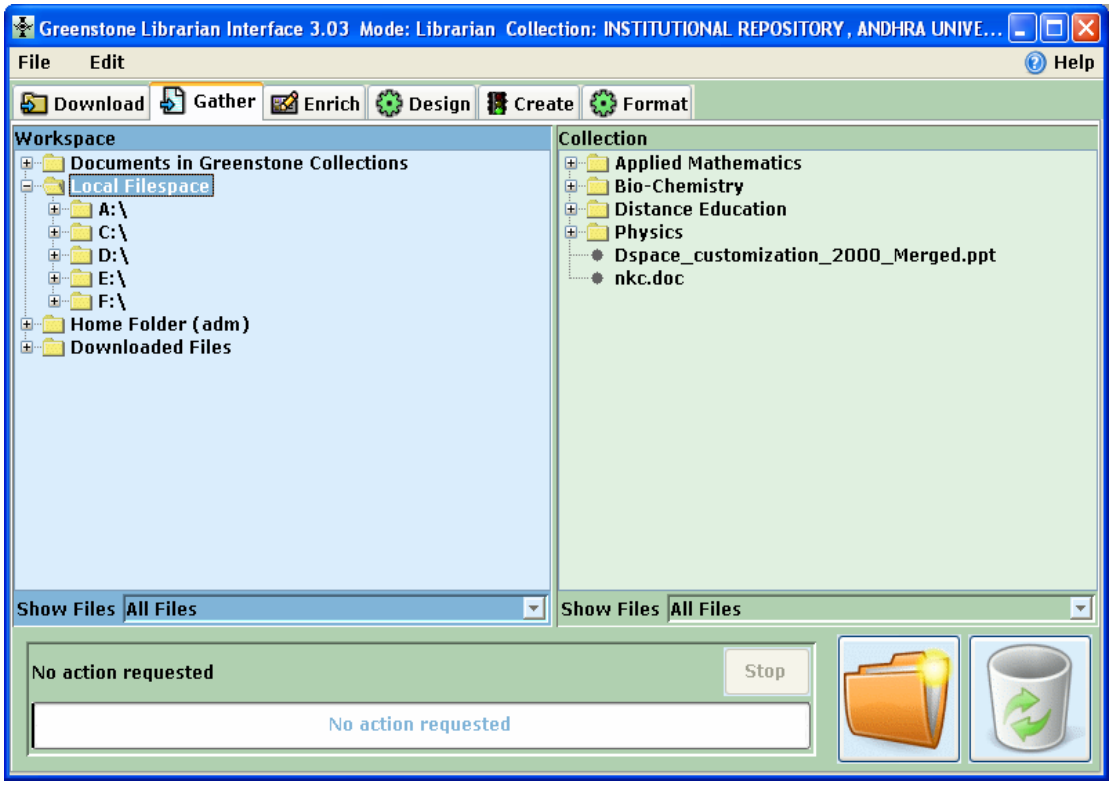
Greenstone Library Interface



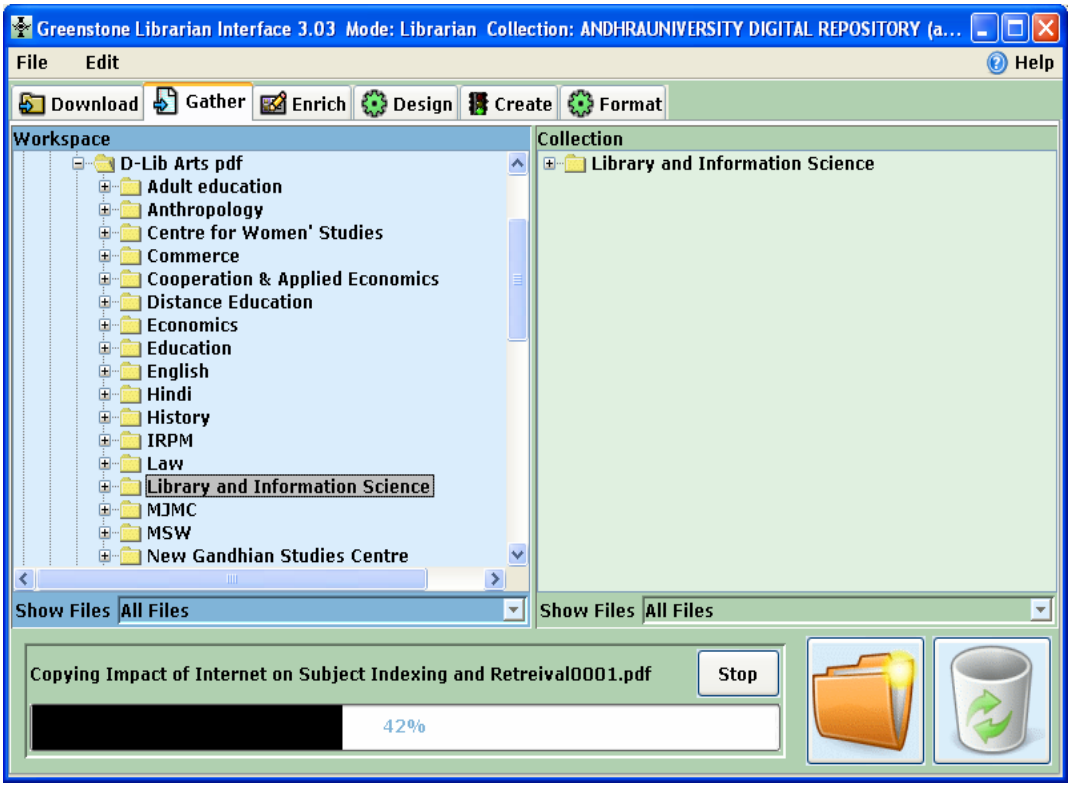
Step 2:Open New file and Gather the Required document.



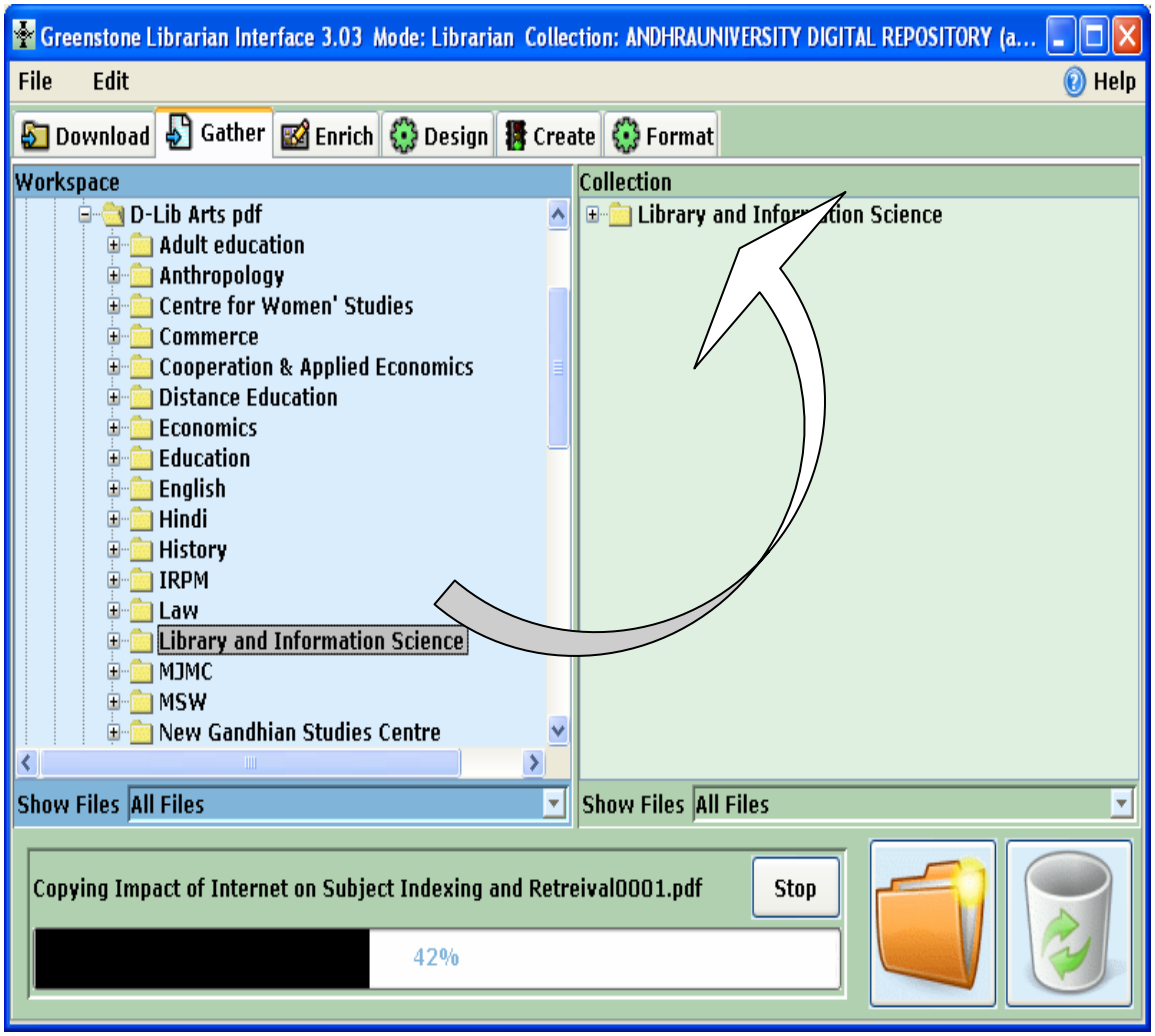
Open the Local File Space



Browse the drive
Browse the folders and click the document

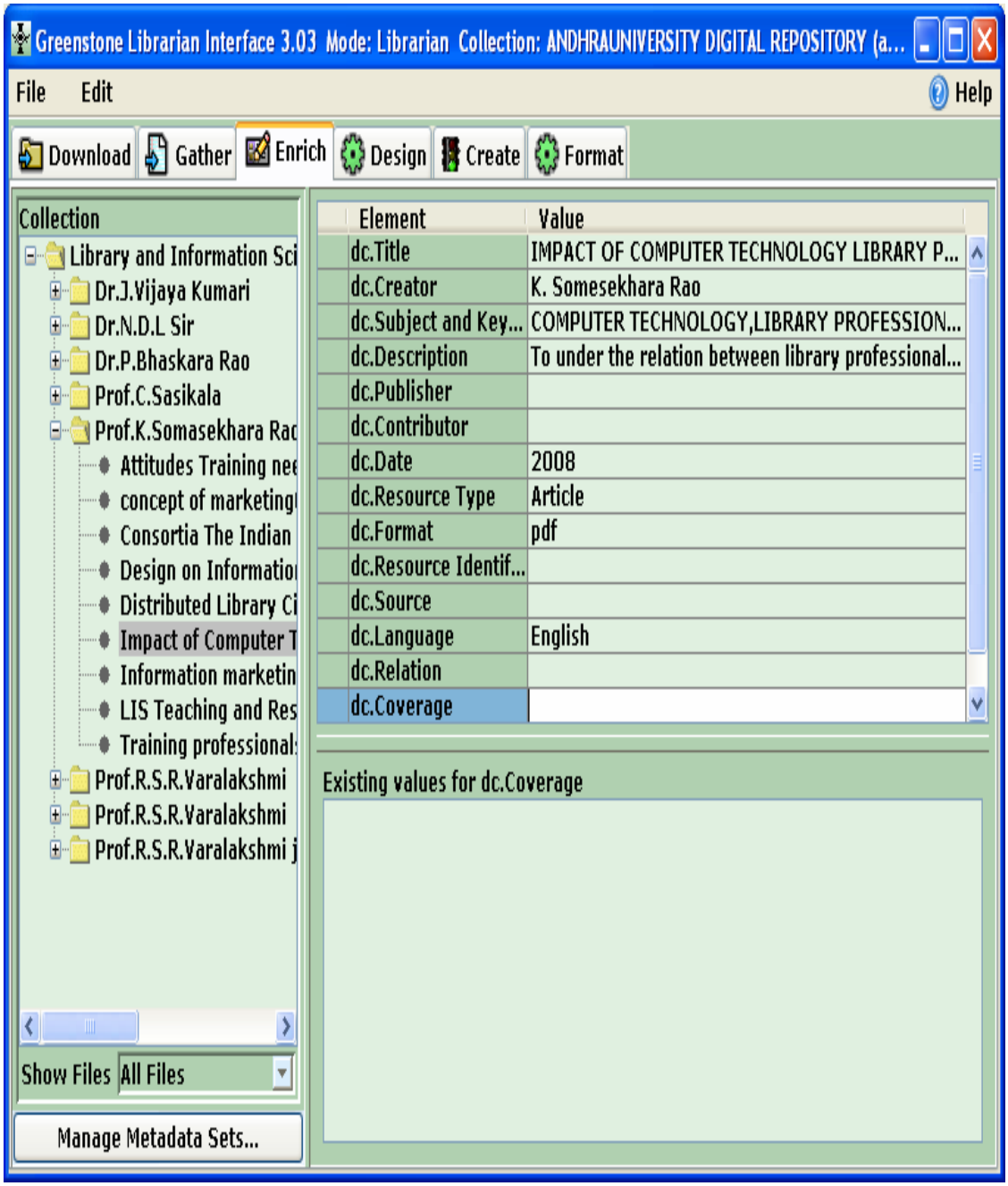


Click Library and Information Science and drag the Workspace to collection

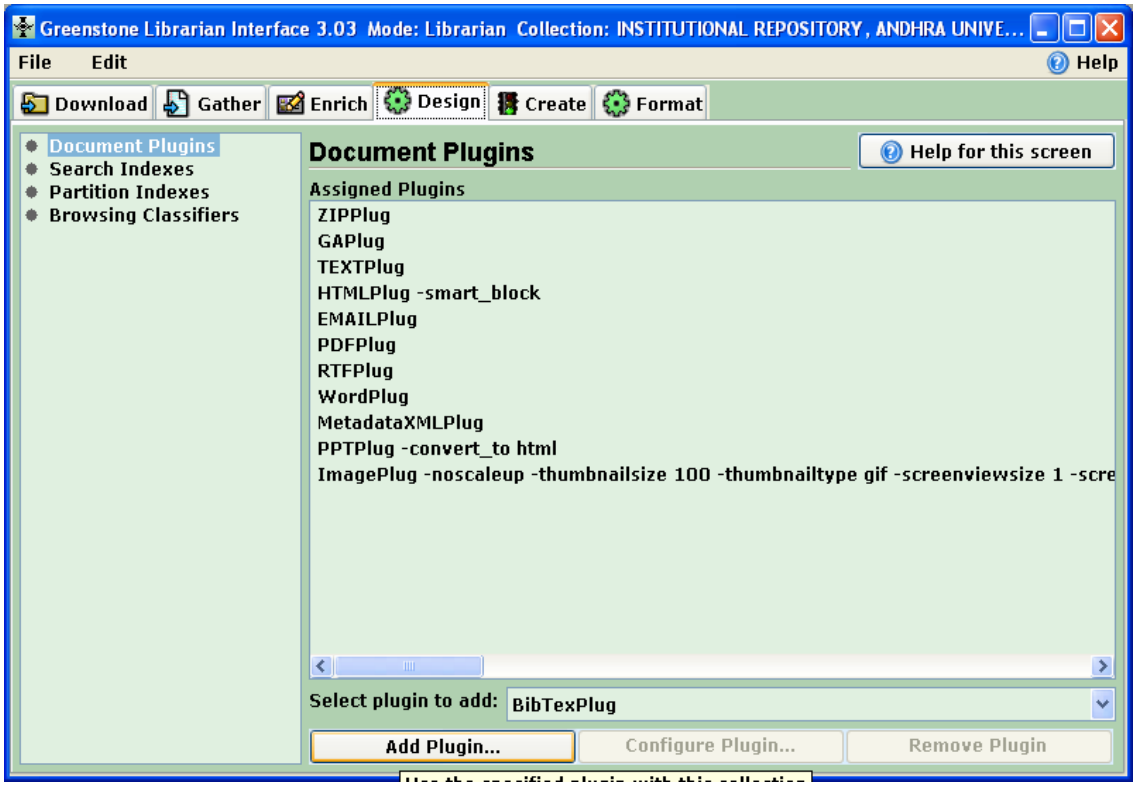


Drag the document to right side of the window and drop it to left side window

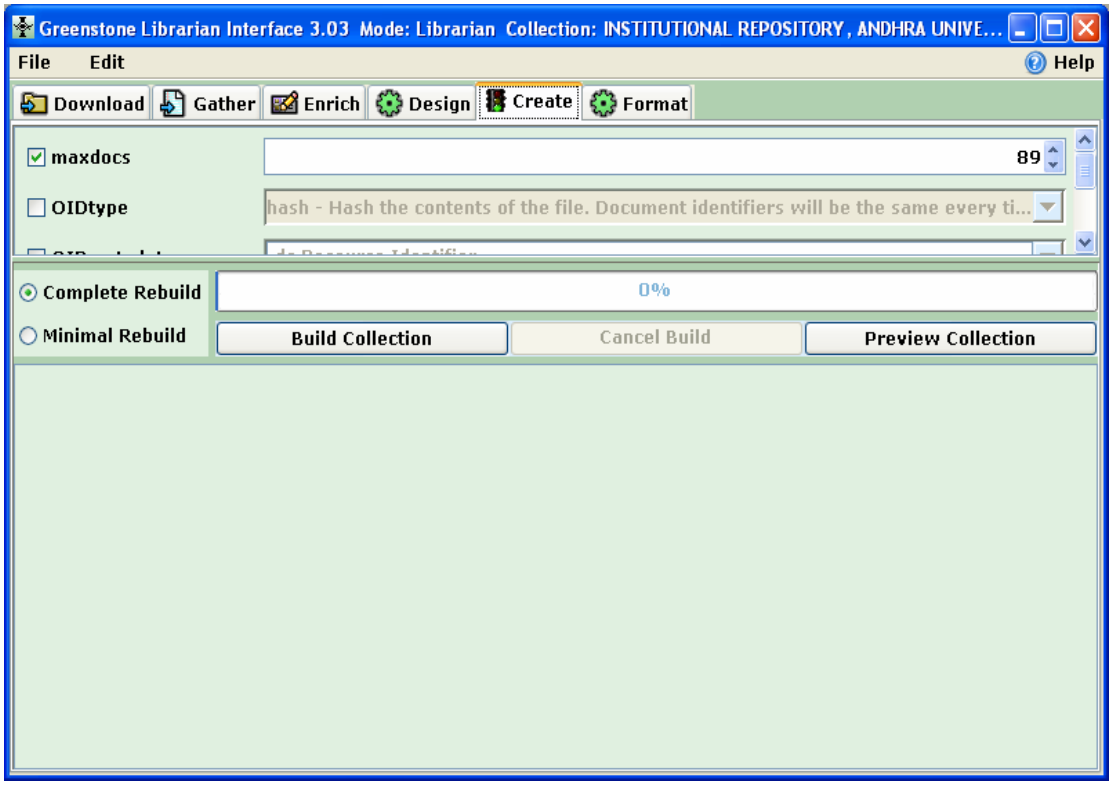
Step 2: Give the details of the document in ***Enrich Menu.***

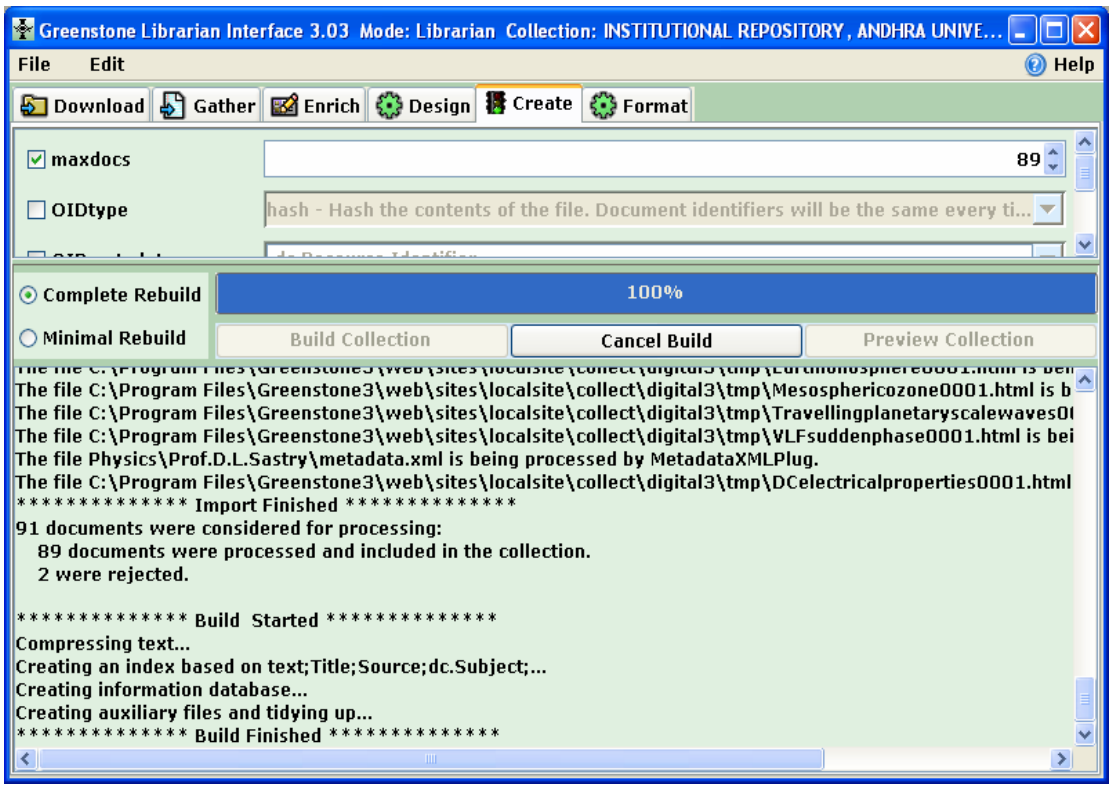


Step 3: Assign the Plug-in's in ***Design Menu.***

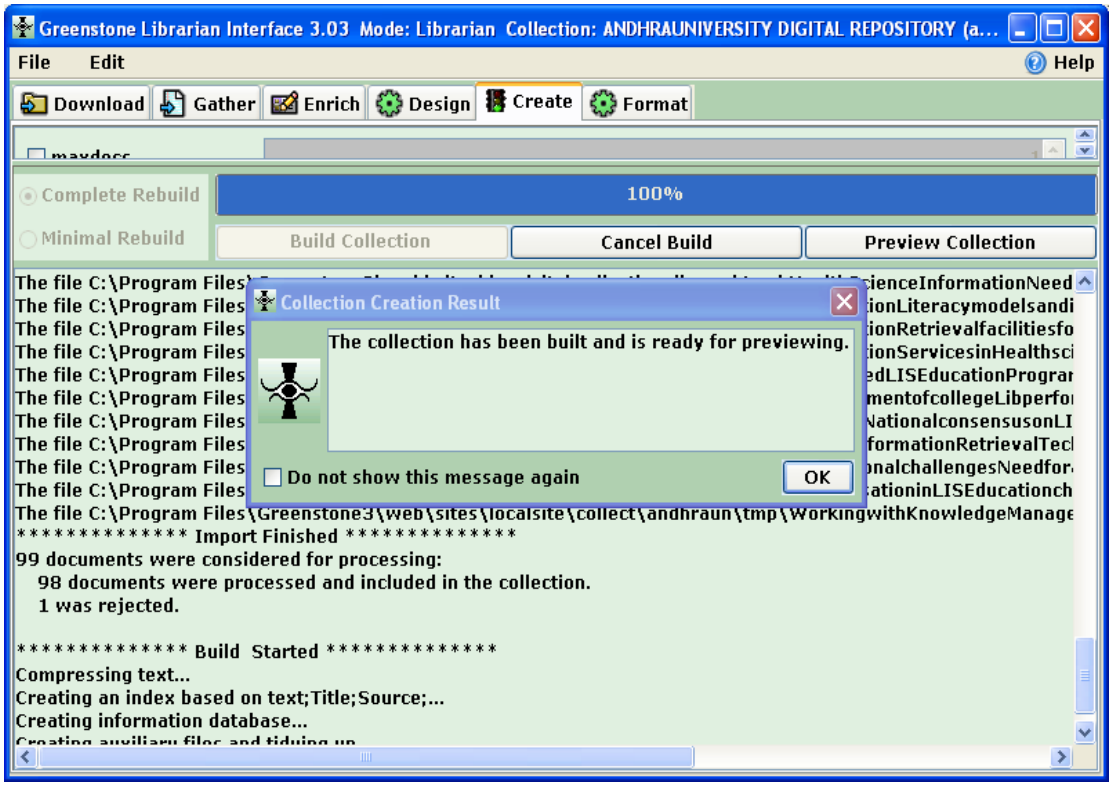


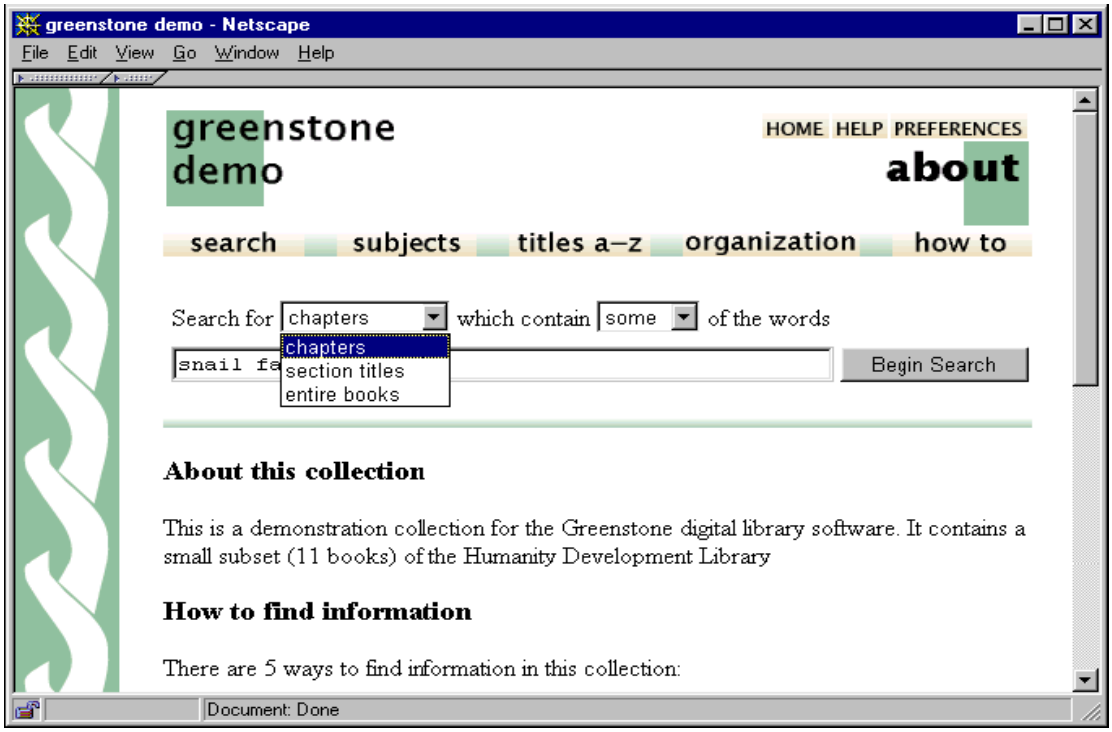
Step 4: Click ***Build Collection*** in ***Create Menu***.





After building the collection, it prompts ***Build Finished*** and you can see the preview of the document.





We can browse the documents by Title, Search, Subject etc.,.We can get the document in PDF format, Web layout (HTML) format etc.,.

Summary:

Institutions or Universities produce substantial quantity of research output. A portion of this output gets publisher in peer reviewed journals (international and national).However, due to various reasons access for them may not be possible means. How do we improve global access to our research output? Open Access Publishing has emerged as a viable alternate to provide global access to research out put. In simple it is a digital library of publications of an institution that cab be accessed by others through Internet. A major development has been the inter-operability framework provided by the Open Archives Initiative (OAI – www.openarchives.org) by organizing institutionally generated information accessible online for on campus and global users, universities and Institutions can get several advantages.

- ✓ Improved visibility to the intellectual output of institutions and the results of investment.
- ✓ The digital repositories contribute for a sound protection and preservation of institutions intellectual property.
- ✓ Interoperable repository removes access barriers and supports interdisciplinary research and discovery.
- ✓ The repository is accessible anywhere through Internet.
- ✓ Improved citation of research publications as the repository will be interoperable and accessible globally.

Methodology

5.1 Introduction:

The availability and use of Information Resources in electronic form and the technological developments for converting Manual and print forms into digital form is also growing in the 21st century. This has paved the way for designing and development of digital libraries, Institutional Digital Repositories etc. The LIS professionals are engaged in designing and developing digital libraries at National, Regional and Institutional levels. The present study is an attempt to study the designing of a digital library. There are a number of digital library softwares available in the market and on the web. A few are open source softwares. For the present study Dspace and GSDL are chosen for developing digital repository.

5.2 Problem specification:

Development of a Digital Repository of LIS collections of Andhra University using Dspace and GSDL.

5.3 Objectives:

Development of a digital repository by

1. Conversion of print articles into electronic form by scanning the documents.
2. Applications of Dspace and GSDL to create, enrich, design and build digital collection.
3. Providing Open Access.

5.4 Data gathering:

The printed articles which were published in different Books and Journals were gathered from the faculty of Dept. of LIS, A.U. School of Arts, Andhra University. I had taken articles from six faculty members. They had given permission for developing this Institutional Repository.

5.5 Conversion of print articles into electronic form:

The gathered articles were converted into electronic form by scanning each page of the articles and stored them in PDF form. The electronic collections were stored in different folders.

5.6 Development of Digital Repository:

- ✓ Dspace and GSDL softwares were installed.
- ✓ Communities and Sub Communities were created.
- ✓ The articles were scanned.
- ✓ The documents were enriched with metadata like Author, Title, Abstract, Language, Publisher, Place, Year, Subject, Keywords etc.
- ✓ Customization was done.
- ✓ The collections were tested for their retrieval capabilities.

5.7 Limitations:

Presently the collection is available on University's Network only. But the idea is to put them open to all the people of the world. It requires certain Infrastructure like separate SERVER and need to put it in University Website.

5.8 Conclusion:

This is a fascinating period in the history of libraries and electronic publishing. Now, it is possible to build large-scale services where collections of information are stored in digital formats and retrieved over networks. The materials are stored on computers. A network connects the computers to personal computers on the users' desks.

Enterprise libraries have a great deal to learn and adapt from innovations in Digital Library research. Apart from using Digital Library technologies in providing improved services using external information sources, libraries should play a key role by participating in national and

international distributed Digital Library efforts and in using Digital Library technologies for publishing and managing internal content.

Regarding this project, the two softwares Dspace and GSDL are very helpful and also good for developing an Institutional Repository for Open Access. More over collection development and maintaining Dspace is easy and flexible and it also coming with open Operating System software linux. We have faced certain problems while doing this project. A few are given below :

- Gathering of collection is difficult because the facult is busy with their work so they slowly respond to our request. Because of this lot of time was taken to gather all the collections.
- Installation of Dspace is difficult and we need support from technical persons but Installation of GSDL is easy compared to Dspace.
- The scanning should be done in PDF Searchable mode otherwise we face difficulties while enriching the documents with metadata.
- The hard copy should be in good form otherwise the scanning is not done properly.
- While scanning and creating metadata or uploading, the software is not recognizing certain characters like special symbols of Science subjects and also superscripts and subscripts.
- If an article is an contributory work then multiple authors will be there. In Dspace, we upload the document for one time only because it takes every author as main author and can retrieve on any contributors of the article hence the

redundancy is reduced but in GSDL it asks for author and contributors separately.

- For Customization, there needed some Knowledge on jsp, html.
- It requires constant updates of digital softwares for working with latest formats.

Ultimately I can say that a person with Software Knowledge i.e., Technological knowledge and Library Profession knowledge can develop and maintain Digital Libraries in a better way.

Bibliography

1. IS 214 - Digital Library and Information Services in Enterprises, Dr. T.B. Rajashekar, National Centre for Science Information (NCSI),Indian Institute of Science.

2. Digital Libray of Nepal Health Research council: Initiatives and Problems Chandra Bhushan yadav; Librarian, Nepal Health Research Council, Kathmandu, Nepal. vbhushan@nhrc.org.np

3. Open access and institutional repositories – a developing country perspective: a case study of India S. B. GHOSH (IGNOU) New Delhi – 110068, India. Email: sbghosh@hotmail.com, ANUP KUMAR DAS, Department of Library and Information Science, Jadavpur University, Kolkata – 700032, India. Email: anupdas2072@hotmail.com.

4. Digital Libraries: Definitions, Issues and Challenges, Gary Cleveland, UDT Core Programme, From INFLANET.

5. Digital library From Wikipedia, the free encyclopedia www.en.wikipedia.org/wiki/Digital library

6. UNESCO official website.

7. What Are Digital Libraries? By Donald J. Waters; From CLIR Issues, Number 4, July/August 1998.

8. Digital Library and Information Services in Enterprises by Dr. T.B. Rajashekar, National Centre for Science Information (NCSI), Indian Institute of Science, Bangalore-560 012. http://www.ncsi.iisc.ernet.in/raja/is214/is214-2006-01-04/topic-1.htm.

9. DSpace: An Open Source Dynamic Digital Repository from D-Lib Magazine, January 2003, Volume 9 Number 1, ISSN 1082-9873.

10. www.lib.fsu.edu/glossary
11. www.collectionscanada.ca/vrc-rvc/040017-151-e.html
12. www.scils.rutgers.edu/~carballo/glossary.html
13. www.edtech.tennessee.edu/~set33/extra/glossary.html
14. http://dspace.org/
15. http://sourceforge.net/projects/dspace/
16. www.greenstone.org