

Developing An Institutional Repository Using Greenstone Digital Library (GSDL) Software

O Seshaiyah

Veeraanjaneyulu

Abstract

This is an age of information explosion. Its impact on library and information centres is profound. With the growing number of e-resources, it has become imperative for information professional to redefine their role in disseminating information to the users. Institutional Repository is a new concept for collecting, managing, disseminating and preserving scholarly works created in digital form by faculty and students in individual universities and colleges. Greenstone is an open source software for building digital repositories. This paper explains how to build an institutional repository using Greenstone Digital Library Software.

Keywords: Institutional Repository, Digital Library, Open Source Software, Greenstone Digital Library

1. Introduction

This is an age of information explosion. Its impact on library and information centres is profound. With the growing number of e-resources, it has become imperative for information professional to redefine their role in disseminating information to the users. Institutions use modern information and communication technologies for information management and dissemination. The institutions converting their assets into digital format for creating digital collection for the last few years.

Institutional Repository is a new concept for collecting, managing, disseminating and preserving scholarly works created in digital form by faculty and students in individual universities and colleges¹. It may also include many of the digital assets generated by an institution such as working papers, lectures, conference proceedings, learning objects, administrative documents, course notes, etc. The learning objects may include study materials,

assignments, question papers, audio-video materials and multimedia presentations.

2. Process Of Digitization

2.1 Scanners and Scanning

The first step in converting paper documents into a digital library collection is to obtain images of all pages of all publications into digital format. The digitization process requires a scanner capable of working at a resolution of 300 dpi (dots per inch). Scanners are available in all price ranges, and all shapes and sizes.

Low-cost flat-bed units are the cheapest and most widely available types of scanners. There are many brands: HP, Agfa, Acer, etc. Consequently these scanners are useful only for small jobs with limited numbers of pages—no more than 200 to 400 pages a month on a regular basis, or one-time jobs of up to 1000 or 2000 pages.

Professional duplex scanners are reliable, heavy-duty machines capable of processing a large volume of pages—typically from 2000 pages to 10,000



pages per day. They have an automatic sheet-feeder tray system that processes batches of about 50 to 200 pages. The best and fastest are duplex machines that scan both sides of the page at once.

Using software provided with the scanner, a digital image of each paper page is scanned and transformed into a Bitmap or TIFF image. Compressed TIFF IV is the best format to use. An average page scanned and converted to this format occupies only 50 Kb, compared to perhaps 2 Mb for the equivalent page in uncompressed Bitmap form. Typically a scanning resolution of 300 dpi is needed, although sometimes 200 dpi is acceptable. The final goal of scanning is either to OCR the pages to obtain perfect word processor or HTML versions of the publications, or to produce enhanced image files such as PDF image files.

2.2 Optical Character Recognition

An optical character recognition or OCR system transforms a scanned image into text. The input is a digitized image in TIFF or Bitmap format—preferably a clean, high-quality image. The output is a word processor or web file, typically in RTF, Word, or TML format. There is several OCR software available in the international market. FineReader (Abbyy Software House), Omnipage (ScanSoft), TextBridge(Omnipage), Readiris are some of the leading OCR software. A good OCR software normally has features to solve the problems such as multiple columns spanned by large headlines, a combination of type styles and faces, text wrapped around illustrations, indented lists, numbered lines, headers and footers, tables, and a number of other characteristics that would become a jumbled mess if the characters on the page were simply recognized and placed into lines of text².

2.23 Alternatives to OCR

There are two alternatives to OCR. They are:

- ◆ Manual retyping
- ◆ Image files

A very low cost alternative to OCR is simply to use a PDF image version of the document pages. The downside is that these files are not searchable. Also, they are quite large—usually 50 Kb per page.

3. Creating An Electronic Collection

Three important aspects should be kept in mind when deciding to create digital collections. First, the collection must be organized. The more content there is, the greater the need for indexes and powerful search systems. For collections of 3000 to 5000 pages or more, indexes and search systems are essential. Second, the needs of end-users must be prevailed. The target groups that will use the collection should be identified, and a process of regular consultation should be set up. Third, the available budget will determine how much can be done.

Institutional Repositories, in India, use Open Source software like Greenstone Digital Library Software (GSDLS), DSpace, GNU Eprints and so on. Here GSDLS is being used to develop institutional repository.

The Greenstone Digital Library software creates a structured digital library including a very powerful search and retrieval engine.

3.1 Greenstone Digital Library Software

Greenstone is a suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. Greenstone is produced by the New Zealand Digital Library

Project at the University of Waikato, and developed and distributed in collaboration with UNESCO and the Human Info NGO. It is open-source software under the terms of the GNU General Public License.

3.11 Special Features:

- ◆ It suits both Windows and Unix (Linux Sun OS) any of these systems can be used as a web server.
- ◆ The administration function build in it enables the items to authorize new users to build collection, protect documents so that they can only be accessed by registered users on presentation of password.
- ◆ It builds collection with effective full-text searching and metadata-based browsing facilities. Collection containing millions of documents upto several gigabytes can be built. Full-text searching is fast because compression is used to reduce the size of the indexes and text users can browse the list of authors, titles, date, class no., etc.
- ◆ Plug Ins can be written to accommodate new document types. The collection can contain pictures, music, audio, video clips, etc. It also supports multilingual documents.
- ◆ Collection can be updated and new one brought online any time with out bringing down the system³.

3.12 Software Requirement (for Windows)

- ◆ Windows with IIS
- ◆ Java 2 Runtime Environment
- ◆ Web browser
- ◆ ImageMagick
- ◆ GSDL 2.70

Here are two separate Windows binary programs on the CD-ROM: the Local Library and the Web

Library. The default installation described above selects the Local Library version. It is strongly recommended to use this version. The Web Library, which is much harder to set up, is only necessary if you already run a web server and want to use it for Greenstone. Despite its modest name, the Local Library offers a complete, self- contained, web-serving capability.

3.13 Steps for Installation of Software

The following steps are required for installation:

1. Install the Java 2 Runtime Environment (latest version).
2. After installing J2RE, go for GSDL folder choose setup gsdl 2.70.
3. The Install Shield Wizard will begin the installation of GSDL software. Click <next>.
4. Accept all the terms of license agreement by clicking on <yes> button.
5. Choose setup Language. English (US) is the default.
6. Choose the type of installation (Local Library).
7. Set the admin password.

For the simplest installation, just accept the default at each point by clicking the Next button. That's all you are needed to do! Greenstone is installed.

Once installation is completed, to start your Greenstone system click on the Start button, open the Program menu, and select Greenstone Digital Library. This brings you a dialogue box: just click Enter Library. This automatically starts your Internet browser and loads the Greenstone Digital Library home page. This can be seen in the example in Figure 1. You enter the Greenstone Demo collection by clicking on its icon.

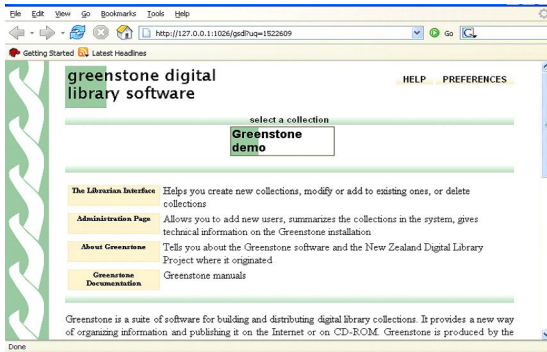


Figure 1 GSDL Home Page

3.14 Building Collection

There are essentially three different ways of building collections:

- ◆ The librarian interface
- ◆ The Collector
- ◆ Building from the command line

3.15 Greenstone Librarian Interface

The following steps are involved in creation of digital collection.

Creation of New collection

- i). Launch the librarian interface under Windows by selecting Greenstone Digital Library from the Programs section of the Start menu and then choose Librarian Interface.(Figure 2)

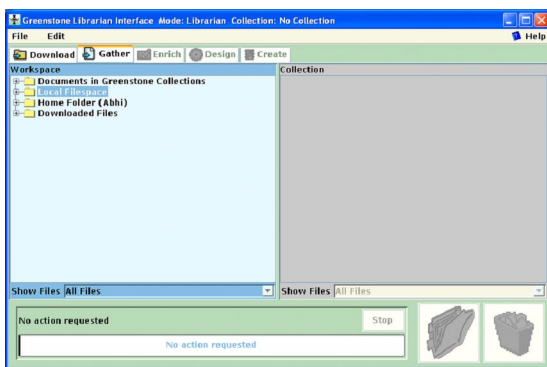


Figure 2: Librarian Interface

- ii). Select New from the File menu in the horizontal menu bar at the top of the window. It needed several fields to fill out for your collection, and then the following screen will appear. (Figure 3)

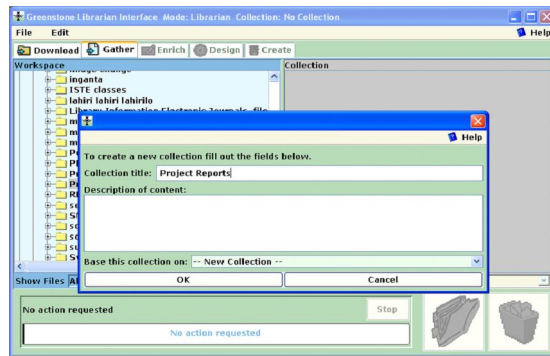


Figure 3: Create a new collection

Selection Metadata

- iii). Software provides different types of metadata on the screen as shown in figure 4. We can select the International standard metadata.

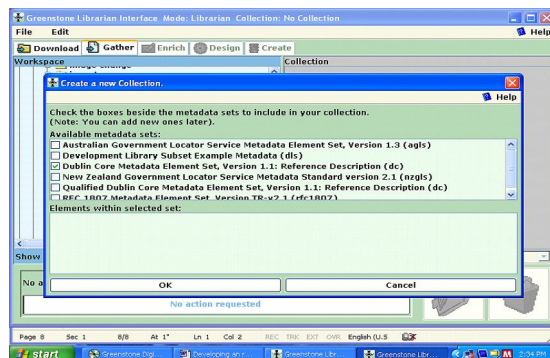


Figure 4: Selection of Metadata

Gathering

- iv). Add some of your own documents that are not in the Demo or DLS collections. Close the Greenstone Collections folder in the left-hand panel and double-click the Local Filespace folder. Navigate to a directory that contains

some documents i.e. Word or HTML or other files. Drag a few of these into the right-hand panel to include them in your collection. (Figure 5)

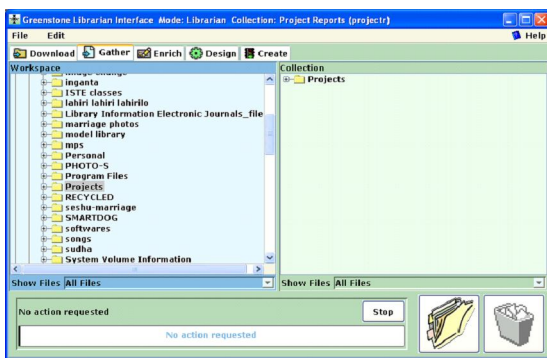


Figure 5: Gathering area

Enrich

v). Having gathered several files into the collection, now enrich them with additional information called 'metadata'. This section explains how metadata is created, edited, assigned and retrieved, and how to use external metadata sources. (Figure 6)

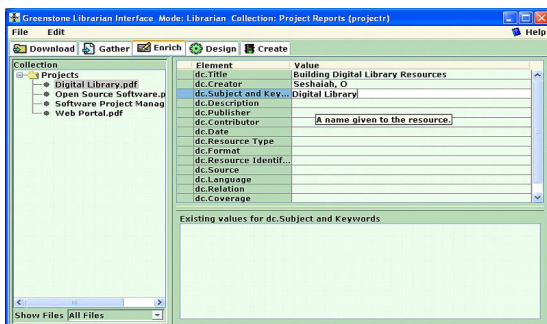


Figure 6: Enriching

Design

vi). Once the files are marked up with metadata, you next decide to make it visible to the users as greenstone collection. What kind of information is searchable? What ways are

provided to browse through the documents? What languages are supported? Where do the buttons appear on the page? These things can be customized. (Figure 7)

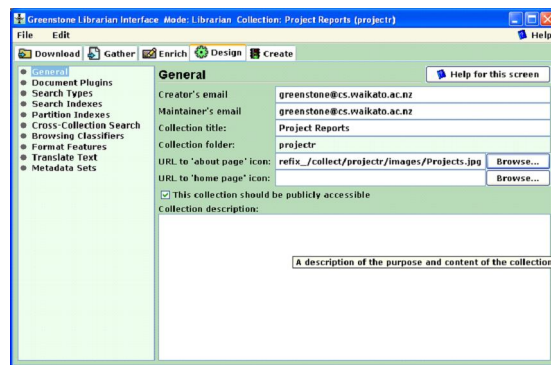


Figure 7: Designing

Create

vii). Click the Create tab to leave the Design mode and create your new collection. Click the Build Collection button at the bottom. While the computer is building the collection you will receive some feedback on what it is going on. (Figure 8)

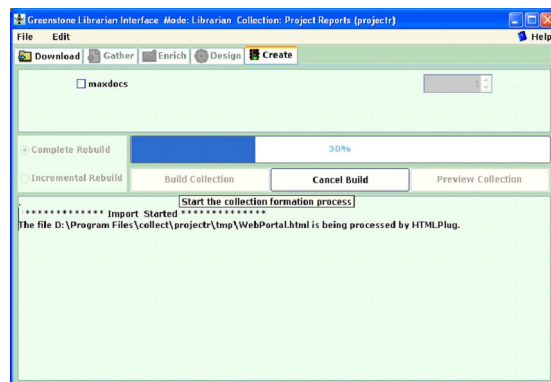


Figure 8: Creating Collection

viii). When it has finished, click the Preview tab to view the collection from the librarian interface. (Figure 9)

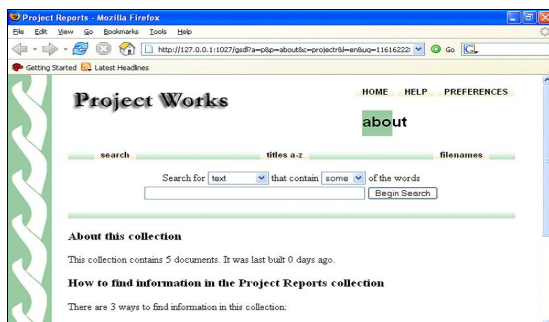


Figure 9: Collection Preview

You will also find when you visit your Greenstone home page that the collection has been installed as one of the regular collections. (Figure 10)

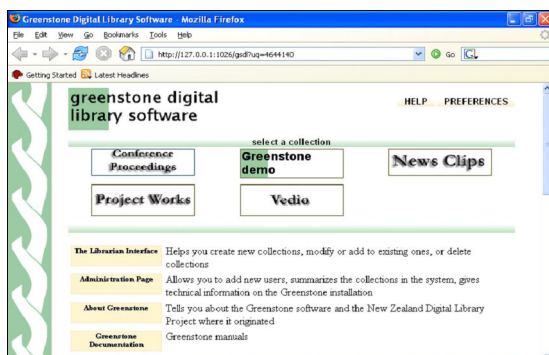


Figure 10: Different collections

3.16 The Collector

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

4. Conclusion

This is the age of information explosion. It demands Institutional Librarians to organize and provide right information to the right user at the right time. To fulfil this task they build their institutional

repositories. Many open source software are available to build digital repositories. GSDL software is one of the open source software. It can build institutional repository effectively.

Acknowledgement

The author is thankful to A. Hariharan, President, SALIS (Society for the Advancement in Library and Information Science) and Mr. Y.Venkateswara Rao, Lecturer in English, PBRVITS for their encouragement in writing this article.

References

1. Parashuram Kattimani S et al. Newsclips Management of Institutional Repositories. In SALIS-2006, edited by Harish Chandra et al. Paper presented at the National Conference on Initiatives in Libraries and Information Centres in the Digital Era, 8-10 June 2006, Coimbatore, SALIS, Chennai, 2006 pp.147-150.
2. www.greenstone.org
3. Sharad Kumar Sonkar et al. Application of Greenstone Digital Library (GSDL) Software in Newspapers Clippings. DESIDOC Bulletin of Information Technology, 2005, 25(3), 9-17.

About Authors

Mr. O Sessaiah, Librarian, PBR Visvodaya Institute of Technology and Science, Kavali.

Mr. Ch.Veeranjaneyulu, Librarian, Geethanjali Institute of Technology and Science, Gangavaram, Nellore (D.t)