

Building Digital Libraries Using Greenstone Open Source Software

Dr. M.G. Sreekumar
(UNESCO Coordinator, Greenstone Support for South Asia)
(<http://greenstonesupport.iimk.ac.in>)

Visiting Professor
Department of Information Science
Faculty of Computer Science and Information Technology
University of Malaya, Kuala Lumpur, Malaysia
(<http://fsktm.um.edu.my>); (mgsree@um.edu.my)

Contents

- | | |
|---|---------------------------------------|
| 1. Introduction | 2. Digital Libraries: Overview |
| 3. Digital Library Features | 4. Digital Library Software |
| 5. Digital Library Objectives
and Workflow | 6. Software Selection |
| 7. Digital Library Development | 8. Greenstone Fact Sheet |
| 9. User base | 10. Multi-lingual Support |
| 11. Training | 12. E-Mail Support |
| 13. Greenstone Features | 14. Installation |
| 15. Collection Building,
Configuration | 16. Helpline, Archives |

1. Introduction

Libraries today buy, subscribe, license and accumulate information in an unprecedented array of content categories or publication types, and in a rapidly proliferating mix of formats (digital as well as print). There is a great deal of cultural divide and philosophical deviation between the traditional information resources being handled by libraries for centuries now and the new genre of electronic and digital information being sourced and accessed. In the traditional paradigm, the books and journals bought and subscribed to by the libraries were naturally owned by them, allowing them to make the best use of the resources within the 'fair use' clause or principle. Whereas in the electronic publishing scenario all the traditional belief, approach and understanding about the digital

documents that the library purchase / subscribe to, have a world of difference. Libraries get only a license to use the electronic information (books, journals, databases, softwares etc.) while purchasing, and even this license is issued only for a prescribed period of time. Librarians at same time, have the professional responsibility to assure uninterrupted as well as perpetual access to the information subscribed to by the library. Issues of copyright, intellectual property, and fair use are very much important to libraries [Orsdel, 2002].

In the current practical library setting there is an amazing penetration of digital information through a variety of publication forms such as books (published as such or issued as accompaniment), journals, portals, vortals, reports, CBTs, WBTs, cases, databases etc. The penetration level of electronic information in the special libraries and libraries belonging to centers of higher learning are supposed to be 70% as against their print counterparts. To make matters more complex the vast array of different formats, standards and platforms in which documents are published, pose a multiplicity of threats to the librarian who is supposed to be the custodian and service provider of these information products once it has found its way into the library. As librarians, we are sometimes the stewards of unique collections too.

2. Digital Libraries

Digital Libraries (DL) are now emerging as a crucial component of global information infrastructure, adopting the latest information and communication technology. Digital Libraries are networked collections of digital texts, documents, images, sounds, data, software, and many more that are the core of today's Internet and tomorrow's universally accessible digital repositories of all human knowledge. According to the Digital Library Federation (DLF, USA - <http://www.dlf.org>), "Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities".

In India currently the concept 'Digital Library' is being practiced by and large loosely or even confused by many information systems. It is therefore imperative that the concept is properly understood so that there is no ambiguity while we progress with the work of designing or developing a digital library which is fully justified in the technical sense of the word. It is important that embarking on a digital library project is something which will take away substantial amount of time, energy, manpower and of course the hard earned money being pumped into it – be it for system development or towards development and maintenance of the collection, in a meaningful way. There is consensus all over that there exists a very large quantum of digital information, scholarly as well as trade, which are scattered and distributed throughout the Net and also being stored in

numerous other databases and repositories spread across the world. Also there is an unprecedented technology support and availability of infrastructure for digital libraries.

3. DL Features

Digital libraries offer new levels of access to broader audiences of users and new opportunities for library and information science field to advance both theory and practice [Marchionini, 1998]. They contain information collections predominantly in digital or electronic form. Electronic publications have some special management requirements as compared to printed document. They include infrastructure, acceptability, access restrictions, readability, standardization, authentication, preservation, copyright, user interface etc.

Digital libraries do enable the seamless integration of the scholarly electronic information, help in creating and maintaining local digital content, and strengthen the mechanisms and capacity of the library's information systems and services. They increase the portability, efficiency of access, flexibility, availability and preservation of digital objects. Digital Libraries can help move the nation towards realizing the enormously powerful vision of 'anytime, anywhere' access to the best and the latest of human thought and culture, so that no classroom, individual or a society is isolated from knowledge resources. Digital library brings the library to the user, overcoming all geographical barriers [ICDL, 2004].

4. DL Software

Undoubtedly it is essential to have a robust and flexible digital collections management and presentation software for creating and delivering digital collections. The preservation of digital objects is currently intimately tied to software that presents those objects. Complete preservation of complex digital objects, especially, is likely to require preservation of the software needed to use those objects. [Borgman, 1996]. The complexity of the situation is that digital library technologies and contents are not static. Continual evolution and investment are required to maintain the digital library. Commercial digital library products are comprehensive and extensible enough to support this evolution, but in many cases they are beyond the reach of most of the libraries in India. Some of the popular commercial DL software in the Indian libraries are VTLS (<http://www.vtls.com>) from the international market and ACADO (<http://www.transversalnet.com/acado/index.htm>) as an Indian initiative. The latter is definitely less costlier when compared but still striving its best to get a critical mass of users. The whole lot of associated issues include initial purchase fee, licensing fee, upgrade fee, annual maintenance contracts (AMCs) and so on. The best available choice for the librarian now is to turn to an Open Source Software (OSS). OSS has grown tremendously in scope and popularity over the last several years, and is now in

widespread use. The growth of OSS has gained the attention of research librarians and created new opportunities for libraries [Frumkin, 2002]. OSS is close to our hearts primarily for their free (or almost free) availability and the broad rights it awards to the consumer. According to Stallman and others at OSS, 'Free Software' uses the 'free' from 'freedom', not the one from 'free beer' [http://www.opensource.org/docs/definition_plain.html].

“OSS is software for which the source code is available to the end-user. The source code can be modified by the end-user. The licensing conditions are intended to facilitate continued re-use and wide availability of the software in both commercial and non-commercial contexts. The cost of acquisition to the end-user is often minimal. According to the proponents of OSS, 'Open source is a development methodology; free software is a social movement'. There are number of other notable features to OSS. Firstly, it has no secrets and the innards are available for anyone to inspect. It is not privately controlled and hence likely to promote open rather than proprietary formats. It is typically maintained by communities rather than corporations and hence bug fixes and enhancement are often frequent and free. It is usually distributed free of charge (developers make their money from support, training, and specialist add-ons; not marketing). It is also essential to clear up some of the misunderstandings about OSS. Open source software may or may not cost money. The cost of ownership often bears little relation to the cost of acquiring a piece of software. 'Public domain' is something different. Open source software has a copyright holder and conditions of legal use. Open source software does not mandate exclusivity. One can use open source programs under Windows. Also one should not choose software solely on the basis of open source. Interoperability and open standards for data are equally important” [OSS Watch, 2005].

According to Altman, for the library fraternity there are other set of reasons too for preferring OSS over commercial software. Long term preservation, assurance of privacy, provision for auditing, facilitating community resources, and conformity to open standards are hallmarks of OSS. Since commercial software is usually distributed only as a binary that will run only on a single hardware platform (and often only under a single version of a particular operating system) commercial software is very difficult to preserve over the long run without developing hardware emulation (and possibly OS 'emulation', as well). OSS, in contrast, can often be recompiled, or at least ported, to new hardware and operating systems [Altman, 2001]. In order to get a picture about the availability of OSS for digital library applications, it is encouraged to visit the directories of OSS projects, such as GNU [<http://www.gnu.org/>] and Sourceforge [<http://www.sourceforge.net/>] open source directory which lists over fifty-thousand projects, and the numbers continue to grow.

5. DL Objectives and Workflow

The primary objective of a digital library is to enhance the digital collection in a substantial way, by strategically sourcing digital materials, conforming to copyright permissions, in all possible standards/formats so that scalability and flexibility is guaranteed for the future and advanced information services are assured to the user community right from beginning. The digital library should also be able to integrate and aggregate the existing collections and services mentioned above with an outstanding client interface. This implies that the digital library system should also have a strong collection interface capable of embracing almost all the popular digital standards and formats and software platforms, in line with the underlying digital library technologies in vogue. This is crucial in the case of multimedia integration, which is again important as we planned to also host a digital audio and video library as part of the core library collection. Emphasis should also be given to maximise the efficiency and effectiveness of the information access and retrieval capabilities of the system by deploying Resource Description Framework [RDF] supplemented with popular descriptive metadata standards. The Internet also possesses, in addition to its mammoth proprietary information base, an invaluable wealth and a vast collection of public domain information products such as databases, books, journals, theses, technical reports, cases, standards, newsletters etc., scattered and distributed across the world. This treasure should also be explored to its maximum for collection building, based on the source and quality. Standard workflow patterns are to be identified for the system which include 'content selection', 'content acquisition', 'content publishing', 'content indexing and storage', and 'content accessing and delivery'. The system should also concern about such related issues, viz., preservation, usage monitoring, access management, interoperability, administration and management etc.

It is always desirable to have crosswalks between the digital catalogue of the library (OPAC) and the digital library, as the OPAC in most cases, acts as a stepping stone for effective information discovery in the library. It also facilitates a healthy bridging between the traditional and the digital library. MARC or any of its variant forms is the desired bibliographic standard recommended for the OPAC, for want of interoperability. Dublin Core [DCMI], MODS (Metadata Object Description Schema) or METS (Metadata Encoding and Transmission) are the recommended metadata format for the digital collection, and XML is the desired encoding scheme [XML]. The XML encoding schemas and the related DTDs (Document Type Definition) strengthen the digital library on strong footing and the XSL (Extensible Stylesheet Language) transformations acts as dynamic gateways between the diverse data streams and the HTML front-end.

6. Selection of the DL Software

The software selection based on set parameters is an uphill task, as the technology itself was still emerging only. In general, what is desirable is a system that is flexible enough to

fit the current digital information system as above and to accommodate future migration. It should be robust in technical architecture as well as the content architecture. The system should address all major digital library related issues such as 'design criteria', 'collection building', 'content organisation', 'access', 'evaluation', 'policy and legal issues' including 'intellectual property rights'. That the system should be in a position to embrace almost all predominant and emerging digital object formats and capable of supporting the standard library technology platforms, should be the major focus. It should provide two important user interfaces: a public user interface for presentation and a metadata creation interface for administration. The system should also provide a powerful search engine and the interface should be easy to navigate and there should be provision for customisation.

There are many digital library softwares available, proprietary as well as open source, and most of them conform to international standards. As mentioned earlier, VTLS and ACADO are the commercial ones available and popular in the Indian market. Some of the popular Open Source Softwares for digital libraries, which are in use internationally, are 'DSpace', 'Dienst', 'Eprints', 'Fedora', 'Greenstone' etc. In line with the subject thrust of this paper, the Greenstone features are discussed in this paper.

7. Developing Digital Libraries using Open Source Software

Digital libraries do enable the creation of local content, strengthen the mechanisms and capacity of the library's information systems and services. They increase the portability, efficiency of access, flexibility, availability and preservation of content. A state-of-art Digital Library shall give a real boost to the library's modernization activities and its endeavours to launch innovative digital information services to the user community. Once the information is made digital, it could be stored, retrieved, shared, copied and transmitted across distances without having to invest any additional expenditure. Value added and pinpointed information at the click of the mouse will become a reality if there is a Library Portal to provide access to the invaluable collection hosted by the Digital Library.

World over there is increasing appreciation of the Open Access movement and the Open Source Software philosophies and for many libraries it is a chosen decision, be it technical or financial reasons, not to go for a proprietary digital library software. One needs to evaluate some of the popular Open Source Software for digital libraries, which are in use internationally. 'Dienst', 'Eprints', 'Fedora', 'Greenstone' etc. are among the candidates for the preferred software. Obviously **Greenstone** outscores the group as a general purpose digital library software from the point of view of a multi-publication type, multi-format, multi-media and a multi-lingual practical digital library [Greenstone]. And once finalized, it could be formally adopted as the software for creating the digital library.

The **Greenstone Digital Library Software (GSDL)** is a top of the line and internationally renowned Open Source Software system for developing digital libraries, promoted by the New Zealand Digital Library project research group at the University of Waikato, led by Dr. Ian H. Witten, and is sponsored by the UNESCO. Greenstone software uses three more additional associated softwares namely, Java Run Time Environment (JRE), ImageMagick and Ghostscript. The software suite is available at the open source directory 'Sourceforge.Net'.

8. Greenstone Fact Sheet (www.greenstone.org)

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 in the form of a fully-searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open-source, multilingual software, issued under the terms of the GNU General Public License. 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."

8.1 Technical Features

8.1.1 Platforms. Greenstone runs on all versions of Windows, and Unix, and Mac OS-X. It is very easy to install. For the default Windows installation absolutely no configuration is necessary, and end users routinely install Greenstone on their personal laptops or workstations. Institutional users run it on their main web server, where it interoperates with standard web server software (e.g. Apache).

8.1.2 Interoperability. Greenstone is highly interoperable using contemporary standards, It incorporates a server that can serve any collection over the Open Archives Protocol for Metadata Harvesting (OAI-PMH), and Greenstone can harvest documents over OAI-PMH and include them in a collection. Any collection can be exported to METS (in the Greenstone METS Profile, approved by the METS Editorial Board and published at <http://www.loc.gov/standards/mets/mets-profiles.html>), and Greenstone can ingest documents in METS form. Any collection can be exported to DSpace ready for DSpace's batch import program, and any DSpace collection can be imported into Greenstone.

8.1.3 Interfaces. Greenstone has two separate interactive interfaces, the Reader interface and the Librarian interface. End users access the digital library through the Reader interface, which operates within a web browser. The Librarian interface is a Java-based graphical user interface (also available as an applet) that makes it easy to gather

material for a collection (downloading it from the web where necessary), enrich it by adding metadata, design the searching and browsing facilities that the collection will offer the user, and build and serve the collection.

8.1.4 Metadata formats. Users define metadata interactively within the Librarian interface.

These metadata sets are predefined:

- Dublin Core (qualified and unqualified)
- RFC 1807
- NZGLS (New Zealand Government Locator Service)
- AGLS (Australian Government Locator Service)

New metadata sets can be defined using Greenstone's Metadata Set Editor. "Plug-ins" are used to ingest externally-prepared metadata in different forms, and plug-ins exist for

- XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace, METS

8.1.5 Document formats. Plug-ins are also used to ingest documents. For textual documents, there are plug-ins for

- PDF, PostScript, Word, RTF, HTML, Plain text, Latex, ZIP archives, Excel, PPT, Email (various formats), source code

For multimedia documents, there are plug-ins for

- Images (any format, including GIF, JIF, JPEG, TIFF), MP3 audio, Ogg Vorbis audio, and a generic plug-in that can be configured for audio formats, MPEG, MIDI, etc.

9. User base

9.1 Distribution. As with all open source projects, the user base for Greenstone is unknown. It is distributed on SourceForge, a leading distribution centre for open source software.

Distributed via SourceForge since:	Nov 2000
Average downloads since then:	4500/month
Currently running at:	4500/month
Proportion of downloads that are documentation:	60%
Proportion of downloads that are software:	40%

Of these, 80% are Windows binaries
15% are Linux binaries
5% are source

9.2 Greenstone Example Collections: Examples of public Greenstone collections (see <http://www.greenstone.org> for URLs) can be found at:

- Association of Indian Labour Historians, Delhi
- Auburn University, Alabama
- California University at Riverside
- Chicago University Library
- Detroit Public Library
- Gresham College, London
- Hawaiian Electronic Library
- Illinois Wesleyan University
- Indian Institute of Management
- Kyrgyz Republic National Library
- LeHigh University, Pennsylvania
- Mari El Republic, Russia
- National Centre for Science Information, Bangalore, India
- Netherlands Institute for Scientific Information Services
- New York Botanical Garden
- Peking University Digital Library
- Philippine Research Education and Government Information Network
- Secretary of Human Rights of Argentina
- Slavonski Brod Public Library, Slovenia
- State Library of Tasmania
- Stuttgart University of Applied Sciences
- Texas A&M University Center for the Study of Digital Libraries
- University of Illinois
- University of North Carolina ibiblio project
- Vietnam National University
- Vimercate Public Library, Milan, Italy
- Washington Research Library Consortium
- Welsh Books Council

UN agencies with an interest in Greenstone include

- UNESCO, Paris
 - Sponsors distribution of the Greenstone software as part of its Information for All programme

- Food and Agriculture Organization (FAO), Rome
 - The Information Management Resource Kit uses Greenstone as the (only) example of digital library software in the Digitization and Digital Libraries self-instructional module (<http://www.imarkgroup.org>)
- Institute for Information Technology in Education (IITE), Moscow
 - Have commissioned an extensive course on Digital libraries in education that uses Greenstone for all the practical work
 - United Nations University (UNU), Japan
 - Two CD-ROM collections of UNU material have been produced

Humanitarian collections. Greenstone is used by Human Info NGO in Belgium to produce collections of humanitarian information and distribute them on CD-ROM widely throughout the developing world. (For more information, contact Michel Loots mloots@humaninfo.org)

- - Number of humanitarian collections: approx 35-40
 - Annual distribution of each one: approx 5,000 copies

10. Languages

One of Greenstone's unique strengths is its multilingual nature. The reader's interface is available in the following languages:

- Arabic, Armenian, Bengali, Catalan, Croatian, Czech, Chinese (both simplified and traditional), Dutch, English, Farsi, Finnish, French, Galician, Georgian, German, Greek, Hebrew, Hindi, Indonesian, Italian, Japanese, Kannada, Kazakh, Kyrgyz, Latvian, Maori, Mongolian, Portuguese (BR and PT versions), Russian, Serbian, Spanish, Thai, Turkish, Ukrainian, Vietnamese

The Librarian interface and the full Greenstone documentation (which is extensive) is in:

- English, French, Spanish, and Russian.

11. Training

Training is a bottleneck for widespread adoption of any digital library software.

Greenstone's Waikato site <http://www.greenstone.org>; the Greenstone Wiki <http://greenstone.sourceforge.net/wiki/index.php/GreenstoneWiki>, and the Greenstone Support for South Asia <http://greenstonesupport.iimk.ac.in> give many training materials and guidance on the software.

Many international training courses have been run.

- UNESCO
 - has sponsored training courses in Bangalore (2002 and 2003), Almaty (2003), Senegal (2004), Suva (2004) and Kozhikode (2006)
- Self-study courses
 - FAO and UNESCO IITE have produced training material on Greenstone in the Digitization and Digital Libraries self-instructional module available at <http://www.imarkgroup.org>.
- Digital Library conferences
 - There have been Greenstone tutorials (on several occasions) at all major digital library conferences: JCDL, ECDL, ICADL, ICDL
 - Librarian conferences
 - There have been Greenstone workshops and presentations at LITA, DLF, ALA Annual Conference
- Payson Institute, Tulane University
 - has run courses that use Greenstone collections as a resource in locations in Africa (Burkina Faso, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Ghana, Rwanda, Senegal, Sierra Leone, Togo) and Latin America (Argentina, Bolivia, Colombia, Ecuador, Guatemala)
- Others
 - There have been several Greenstone courses in India (e.g. Kozhikode, Poona), some in Canada (Vancouver, Calgary, Edmonton), one in Cuba (Havana).

12. E-mail support

There are many E-Lists and E-Groups available for Greenstone support. For subscribing to the main Greenstone lists, visit <https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users> for User's List (greenstone-users-request@list.scms.waikato.ac.nz) and <https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-devel> for Developer's list. There is also an E-List for supporting the South Asian Greenstone users greenstonesupport@iimk.ac.in.

-

Number of people on Greenstone email lists:	600
Number of countries represented:	70
Number of messages (excluding spam):	150/month

13. Greenstone : Features

The salient features of Greenstone are basically taken from two of the official publications of the software development team appeared in D-Lib Magazine during the year 2001 [Witten, 2001] and 2003 [Witten, 2003]. Greenstone builds collections using almost popular and standard digital formats such as HTML, XML, Word, Post Script, PDF, RTF, JPG, GIF, JPEG, MPEG etc. and many other formats which include audio as well as video. It is provided with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt entirely automatically. The system is extensible: software "plug-ins" accommodate different document and metadata types. Greenstone incorporates an interface that makes it easy for people to create their own library collections. Collections may be built and served locally from the user's own web server, or (given appropriate permissions) remotely on a shared digital library host. End users can easily build new collections styled after existing ones from material on the Web or from their local files (or both), and collections can be updated and new ones brought on-line at any time. The Greenstone Librarian Interface (GLI) is a Java based GUI interface for easy collection building. Greenstone software runs on a wide variety of platforms such as Windows, Unix / Linux, Apple Mac etc. and provides full-text mirroring, indexing, searching, browsing and metadata extraction. It incorporates an interface that makes it easy for institutions to create their own library collections. Collections could be built and served locally from the user's own web server, or (given appropriate permissions) remotely on a shared digital library host. The other set of features include OAI plug-in (introduced since the 2.40 version) and DCMI compliance, UNICODE based multi-lingual capabilities and a user-friendly multimedia interfacing

[Unicode]. Further more, it has a powerful search engine 'Managing Gigabyte' Plus-Plus ('MG' PP) and metadata-based browsing facilities. A very interesting feature of Greenstone is its exhaustive set of well documented and articulated manuals (<http://www.greenstone.org/cgi-bin/library?e=p-en-docs-utfZz-8&a=p&p=docs>) such as 'Installer's Guide', 'User's Guide', 'Developer's Guide', and 'From Paper to Collection' a document describing the entire process of creating a digital library collection from paper documents. This includes the scanning and OCR process and the use of the "Organizer". There is one more interesting documentation 'Inside Greenstone Collections' which clarifies most of the trickier parts of using Greenstone, especially dealing with configuration file for the collection in question.

The primary objective of any digital library will be to enhance the digital collection in a substantial way, by strategically sourcing digital materials, conforming to copyright permissions, in all possible standards / formats so that scalability and flexibility is guaranteed for the future and advanced information services and are assured to the user community right from beginning. The digital library has to be planned in such a way that it will integrate and aggregate the existing collections and services with an outstanding user interface. Accordingly, necessary strategies are to be adopted towards working out the digital library system. This implies that the digital library system should have a strong collection interface capable of embracing almost all the popular digital standards, digital formats and software platforms, in line with the underlying digital library technologies in vogue. This is crucial in the case of multimedia integration, which is again important as it is planned to host digital audio and video library as part of the core library collection.

14. Greenstone Installation

The GNU Public License version Greenstone can be downloaded from 'http://www.greenstone.org' or 'http://sourceforge.net/index.php'. You can download the binaries for Linux or Windows. The associated softwares such as Java Runtime Environment (JRE) and the Imagemagick also to be downloaded. A graphical tool is used for collection building and configurations and customization. This is called the Greenstone Librarian Interface (GLI) and it requires the Java Runtime Environment (JRE). The latest version pertaining to Volume 2 release of Greenstone is V.2.74.

Click on "gsdl-2.74-win32 1.exe". The Install Shield Wizard will begin the installation. Accept all the term of license agreement by clicking on <Yes> button. Click on <next> to install GSDL in the default folder, which is C:\program files\greenstone. Choose the type 'Local Library'. By default, Local Library is highlighted. Set the Admin Password as "admin" (you can later change it). Installation wizard now starts copying the required files from the GSDL folder. Click on the Finish button to finish GSDL installation. To check whether your installation is proper, Click on 'Start→Programs→Greenstone Digital Library→Greenstone Digital Library'. Click on Enter Library in the 'Dialog Box' and Your Browser should display The GSDL Homepage.

Now, install 'Imagemagick' software which is available at '<http://www.imagemagick.org>'. 'Imagemagick' is a free software suite to create, edit, and compose bitmap images. It can read, convert and write images in a large variety of formats. Using Imagemagick, images can be cropped, colors can be changed, various effects can be applied.

15. Collection Building and Configuration

Greenstone used to have three modes of collection building, viz., Command Line, Web Interface and the GLI (Greenstone Librarian Interface), until recently. Progressing with version 2.4x., the GLI got strengthened as well as popularized, the Web Interface mode has been withdrawn temporarily. The GLI based collection building is quite easy and simple a method. Collection developers can activate the GLI software and use the 'Gather', 'Enrich', 'Design', 'Create' panel for making collection.

1. The '**Gather**' Panel facilitates putting the relevant files from the 'workspace' to the 'collection building' area. The 'Enrich' Panel explains how metadata is created, edited, assigned and retrieved, and how to use external metadata sources. Help for this is provided in the GLI Interface. The 'Design' Panel facilitates customising your interface, once your files are marked up with metadata. Using the Gather Panel, you can specify the fields that are searchable, allow browsing through the document, facilitate the languages that are supported, and provide the buttons that are to appear on the page. Help for this is provided in the GLI Interface. The Create Panel facilitates creation of your collection.

To build a typical collection, say 'MyTest' collection, first go to 'File' section, select 'New' and then give the collection name as 'MyTest'. Select OK from the panel and then you will get another panel popped up where you will select the appropriate Metadata Set. You may also give the description about the collection here. By default, the system will prompt Dublin Core metadata set. Click on OK button and you will get the collection create panel made ready for accepting the file(s).

The 'Gather' Panel is activated now. From the 'Workspace' provided, identify the document to be put in the collection by locating it in the local folder. Drag and drop the file to the Collection Area using the mouse. The necessary 'plugin' for the creation of the collection is to be tick marked and enabled in the 'Design' panel, which is the next step in the collection building process. If the collection has objects for which 'plugins' are not provided in the default set, a new dialog box for adding the required plugin will appear and it has to be the added to the default set.

2. Go to the '**Enrich**' panel and give necessary values for the Dublin Core element sets.

Manage Metadata Sets - This feature allows you to add, configure and remove the Metadata Sets in your collection and what Elements they contain.

3. Design Panel

The next step is to give necessary values and arguments for the '**Design**' panel which include [Note: GLI Design Panel's own language is used below i. to x., for want of clarity and to avoid any ambiguity in usage]:

i. *General Options* - In this section, give the e-mail address of the 'collection creator', 'collection maintainer', 'collection title' (will be supplied by the system), collection folder (will be supplied by the system), Image file location for the Collection icon and the Image file location for the Document icon. Click on the Tick mark for making this collection publicly available.

ii. *Document Plugins* - This section facilitates adding, configuring or removing plugins from your collection. To add one, choose it from the combobox and click 'Add Plugin'. To configure or remove one, select it from the list of assigned plugins and then: i) Change its position in the plugin order by clicking on the arrow buttons. (Note: The position of RecPlug and ArcPlug are fixed). ii) Configure it by clicking 'Configure Plugin', iii) Remove it by clicking 'Remove Plugin'. Plugins are configured using a pop-up design area with a scrollable list of arguments. Enable arguments and enter or select values as necessary.

iii. *Search Types* - Defining the search type is an advanced feature, only available when enabled (by checking the 'Enable Advanced Searches' box). Once enabled, further controls for selecting and changing the order of search types become available. See the 'Search Type Selection and Ordering' section of the 'Design' Panel for more information on this.

iv. *Search Indexes* - The required number of searchable indexes the collection must have, is to be selected here. To add a new index, enter a unique name for the index, select material/metadata is to be indexed, and click 'Add Index'. If you wish to add all of the available sources so as to have indexes built on them, then click 'Add All'.

v. *Partition Indexes* - This feature help to refine index creation. This facility is disabled in the GLI mode.

vi. *Cross-Collection Search* - This feature facilitates cross-collection searching, where a single search is performed over several collections, as if all the collections were one. Specify (Tick Mark) the collections to include in a search by clicking on the appropriate collection's name in the list below. The current collection will automatically be included. [Note : If the individual collections do not have the same indexes (including sub

collection partitions and language partitions) as each other, cross-collection searching will not work properly. The user will only be able to search using indexes common to all collections].

vii. *Browsing Classifiers* - This feature allows the AtoZ browsing of the collection and by default it takes the 'Dublin Core . Title'. You can more data elements in the AtoZ classify list as deem fit for the collection using this feature.

viii. *Format Features* - The web pages you see when using Greenstone are not pre-stored, but are generated 'on the fly' as they are needed. Format commands are used to change the appearance of these generated pages. Some are switches that control the display of documents or parts of documents; others are more complex and require html code as an argument. To add a format command, choose it from the 'feature' list. If a True/False option panel appears, select the state by clicking on the appropriate button.

For example, to get the Cover Image displayed in the document while building the collection, go to the 'Choose Features' dropdown box and enable the 'DocumentIMages', i.e., make its value to True.

ix. *Translate Text* - Use this feature to review and assign translations of text fragments in your collection. The translated text will appear in a different box in the browser.

x . *Metadata Sets* - This feature allows you to add, configure and remove the Metadata Sets in your collection and what Elements they contain.

4. Now go to the '**Create**' panel and click on the '**Build Collection**'. Greenstone will start creating the collection. You can see the built collection by clicking on the 'Preview Collection'.

Please remember you have to save your collection development process from time to time. It is not mandatory that you need to comply with the entire set of formalities for a building a collection in a single stretch. You can do it in different sessions too. What is important is saving the sessions from time to time. In the GLI mode of collection building, the various panels to be used are illustrated in Figure 1.

5. Format Panel

i. General - This section explains how to review and alter the general settings associated with your collection. First, under the "Format" tab, click "General". Here some collection wide metadata can be set or modified, including the title and description entered when starting a new collection. First are the contact email addresses of the collection's creator and maintainer. The following field allows you to change the collection title. The folder that the collection is stored in is shown next, but this cannot be altered. Then comes the

icon to show at the top left of the collection's "About" page (in the form of a URL), followed by the icon used in the Greenstone library page to link to the collection. Next is a checkbox that controls whether the collection should be publicly accessible. Finally comes the "Collection Description" text area as described in "Creating a New Collection".

ii. Search - This section explains how to set the display text for the drop down lists on the search page. Under the "Format" tab, click "Search". This pane contains a table listing each search index, index level (for MGPP or Lucene collections), and index or language partition. Here you can enter the text to be used for each item in the various drop-down lists on the search page. This pane only allows you to set the text for one language, the current language used by GLI. To translate these names for other languages, use the Translate Text part of the Format view (see "Translate Text" feature in the Format panel).

iii. Format Features - The web pages you see when using Greenstone are not pre-stored, but are generated 'on the fly' as they are needed. Format commands are used to change the appearance of these generated pages. Some are switches that control the display of documents or parts of documents; others are more complex and require html code as an argument. To add a format command, choose it from the 'feature' list. If a True/False option panel appears, select the state by clicking on the appropriate button.

For example, to get the Cover Image displayed in the document while building the collection, go to the 'Choose Features' dropdown box and enable the 'DocumentIMages', i.e., set its value to True.

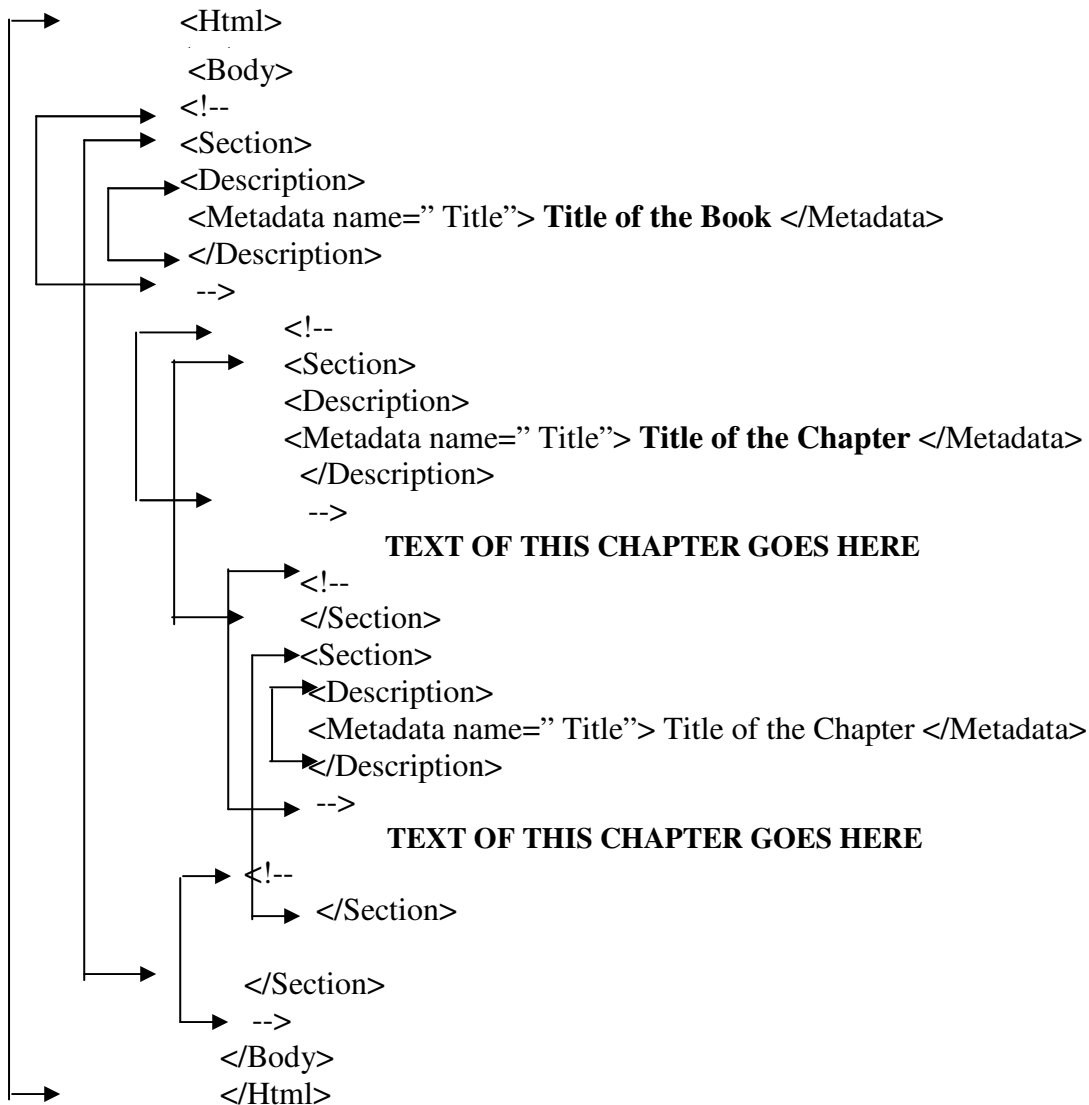
iv. Translate Text - Use this feature to review and assign translations of text fragments in your collection. The translated text will appear in a different box in the browser.

v. Cross-Collection Search - This feature facilitates cross-collection searching, where a single search is performed over several collections, as if all the collections were one. Specify (Tick Mark) the collections to include in a search by clicking on the appropriate collection's name in the list below. The current collection will automatically be included. [Note : If the individual collections do not have the same indexes (including sub collection partitions and language partitions) as each other, cross-collection searching will not work properly. The user will only be able to search using indexes common to all collections].

vi. Collection Specific Macros - Under the "Format" tab, click "Collection Specific Macros". This view shows the contents of the collection's extra.dm macro file. This is where collection specific macros can be defined. To learn more about macros, see Chapter 3 of the Greenstone Developer's Guide.

15.1 Hierarchy Structure

To create indexes for section and sub-section, the pre-requisite is that the document should be in HTML format. Therefore your collection files in other formats like PDF, Word, etc. are first to be converted into HTML format. Also in the Collection Configuration file (for GLI, in the Design Panel, in the Document Plugin section, while configuring the Arguments in the HTML Plugin, click and enable the 'description_tags'), the HTML plugin has to be modified to 'plugin HTMLPlug -description_tags'. Corresponding changes have to be made in the 'indexes' and the 'collectionmeta' lines. Obviously now the Source File has to be edited as a HTML file structure. For the section and sub sections, you need to edit the source file as follows, giving XML tags as comments in the body of the HTML file.



15.2 Customization of User Interface (MyLibrary)

In order to change the look and feel of the Greenstone user interface, you need to work on the Collection Configuration (Collect.cfg) files. Customising the User Interface requires a certain degree of knowledge on HTML and some level of Web Designing skills are prerequisites for this.

i. *Collect.cfg* - This is the collection configuration file. You can find this file in the “Program Files\Greenstone\collect\etc” directory. Details on how to create this file can be found in the Developer’s Guide, “1.5 Collection configuration file” and “2.3 Formatting Greenstone output”.

ii. *Macro files* - Macro files have an extension ‘.dm’. All macro files are stored in the “macros” directory. Details on how to create macros and macro files can be found in the Developer’s Guide “2.4 controlling the Greenstone user interface”.

iii. *Image files* - All images files can be found in the ‘Program Files\Greenstone\images’ directory.

iv. *Main.cfg* - This file contains a list of all macro files used for the User Interface. If you created a new ‘.dm’ file, you need to add it to this file. The main.cfg file is stored in the “Program Files\Greenstone\etc” directory.

v. *Getting the Cover Image* - For you to get the Cover Image of your input document, you need to put the image file and the source file (document) into a single folder. They both should bear the same name also. While building the collection, Greenstone will take both the files to “Program Files\Greenstone\collect<collection name>\archives\Hash”. The collection thus built will display the Cover Image along with the document. Also in the Design Panel, in the Document Plugin section, while configuring the Arguments for the HTML Plugin, give the custom argument as ‘cover_image’.

vi. *Getting the Collection Icon* - Click on Design panel ->General Option -> URL to home page icon (Browse for image and locate it).

vii. *Getting Header Image for the Digital Library* - To get the header image which says MyLibrary banner in the DL head, create the graphic file (preferably a GIF file), name it as ‘gsdlhead.xxx’ and then replace it with the file available in ‘Program Files\Greenstone\images.’

viii. *Deep Level Customization* - By default, Greenstone’s collection icon area is a matrix grid (the N X 3 format). You can change the collection icon area by editing the ‘_content_ macro’ in ‘home.dm’. You will need to remove the ‘_homeextra_ macro’ (this

is the N x 3 table that the Greenstone C++ code automatically creates for you) and can then put whatever design customization you want into this area. You will need to put the icons and links to the collection yourself.

You can also achieve high end customization by replacing the 'home.dm' with 'yourhome.dm' in the \greenstone\etc folder.

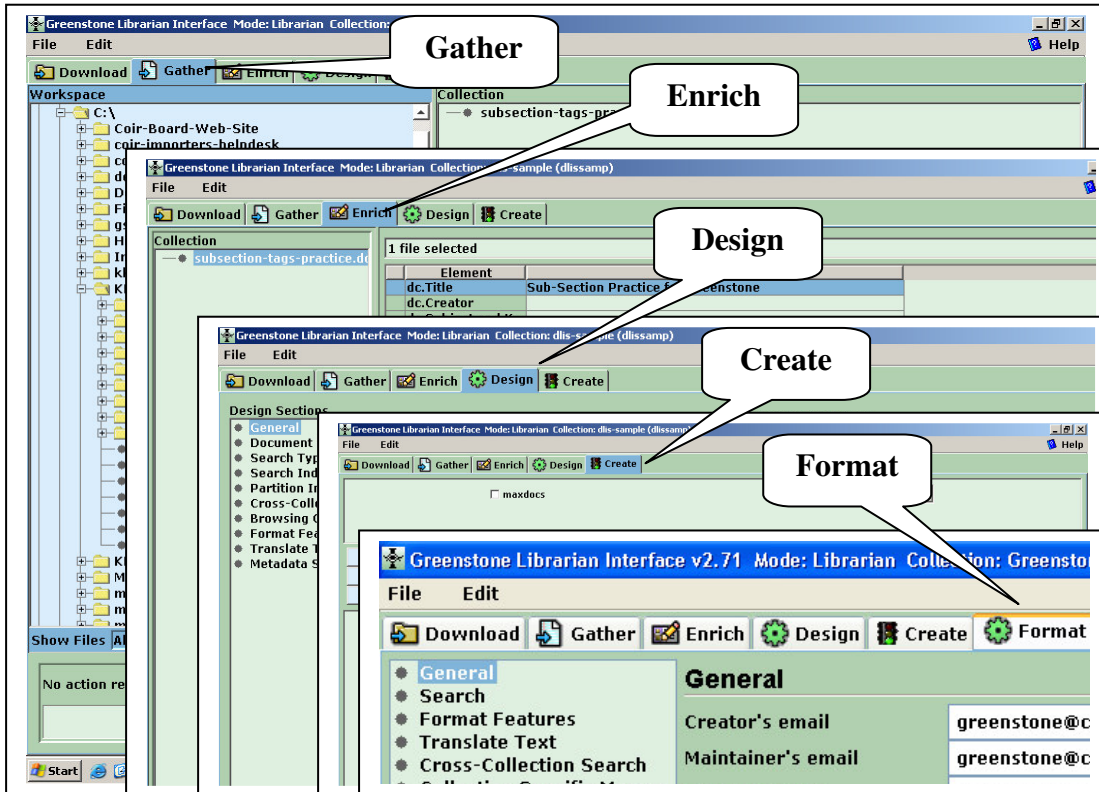


Fig. 1. GLI Panels for Gather, Enrich, Design, Create and Format

16. GSDL : Helpline, Archives

Greenstone’s E-Mail list is a very useful and active listserv which shares and clarifies user experiences and stories dealing with real life situations. To subscribe or unsubscribe to the list via the World Wide Web, visit “<https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users>” or, via email, send a message with subject or body 'help' to “greenstone-users-request@list.scms.waikato.ac.nz”. Greenstone has started one more List recently, for the Greenstone 3 Version (the latest Beta version) user group, and the details are available at “<https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone3>”.

UNESCO has initiated a Greenstone support organization for South Asia in 2006, supported by a group of experts in the region, and it is coordinated by IIM Kozhikode <http://greenstonesupport.iimk.ac.in>. The site is rich with many of the Greenstone support materials. In addition, an E-list greenstonesupport@iimk.ac.in offers online support to professionals on Greenstone.

For those looking for quick solutions for their real-time or on-the-job trouble shooting while using the software, 'Greenstone Archives' is a treasure house. It is a database of the email messages circulated in the List, and is searchable. The mails generated from the List and its threads are archived and made available for the user community. The archive is available at "<http://www.sadl.uleth.ca/nz/cgi-bin/library?a=p&p=about&c=gsarch-e>". This is the major list used worldwide for Greenstone and the content of the messages is usually global in nature. Developers and Greenstone users can avoid a great deal of unwanted labour by carefully going through the archive before they start working on problem solving, or before shooting a mail to the List.

References

1. Orsdel, Lee Van; Born, Kathleen. 2002
Doing the Digital Flip.
Library Journal, 127 (7): 51-55.
2. OCLC Report on Five-Year information format trends. 2003
<<http://www.oclc.org/reports/2003format.htm>>
3. Marchionini, G. 1998
Research and development in digital libraries. Allen Kent (Ed.)
Encyclopedia of Library and Information Science, 63: 259-279.
4. ICDL 2004. <www.teriin.org/events/icdl/background.htm>
5. Sreekumar M.G. and Sunitha T. 2005
Essential Strategies and Skill Sets Towards Creating Digital Libraries Using Open Source Software.
[Proceedings of NAACLIN 2005, DELNET, Bangalore, India].
6. Borgman, Christine L. 1996
Social Aspects of Digital Libraries, pp170
[Proceedings of the first ACM international conference on Digital libraries
Bethesda, Maryland, United State, March 20-23, Organised by Association of Computing Machinery]

7. Frumkin, Jeremy (ED). 2002
Special Issue: Open Source Software
Information Technology and Libraries 21(1)
8. Stallman, Richard
<http://www.opensource.org/docs/definition_plain.html>
9. OSS Watch <http://www.oss-watch.ac.uk/talks/2003-09-24-csg/index.xml.ID=body.1_div.37>
10. Altman, M. 2001
Open Source Software for Libraries: from Greenstone to the Virtual Data Center and Beyond.
IASSIST Quarterly. Winter : 1-7.
11. GNU (GNU's Not Unix!)
<<http://www.gnu.org/> (13 June, 2005)>
12. SourceForge.Net (world's largest Open Source software development website)
<<http://www.sourceforge.net/>>
13. RDF (Resource Description Framework)
<<http://www.w3c.org/RDF>>
14. DCMI (Dublin Core Metadata Initiative)
<<http://dublincore.org>>
15. Greenstone Digital Library Software
<<http://www.greenstone.org>>
16. Witten, Ian H. et al. 2001
Greenstone : Open-Source Digital Library Software
D-Lib Magazine, 7 (10): 1-16.
17. Witten, Ian H. 2003
Examples of Practical Digital Libraries : Collections Built Internationally Using Greenstone
D-Lib Magazine 9 (3): 1-15.
18. Unicode Consortium
<<http://unicode.org>>
19. IIMK Digital Library

<<http://intranet.iimk.ac.in/cgi-bin/library>>