Managing Digital Collections with Greenstone Digital Library Software

Dr Venkata Rao Pokuri
<pvrao@mgsipap.org>
Outline

- Digital Collections
- Digital Library
- Open Source Software
- Greenstone Digital Library Software (GSDL)
- Installation of GSDL
- Using Librarian Interface (GLI)
- Using User Interface
- Practical examples
Electronic Document

- When the content (text, picture, audio-video etc.) is created or stored electronically.
- An Electronic/Digital Document
  - Unit of work or Complete work
    - Electronically stored book is Electronic Document
    - Each chapter is Electronic Document
    - A picture embedded in the chapter
Digital Collections

- Print documents as electronic documents
- Existing and future electronic documents
- Electronic document format
- Electronic Collections
Digitization

- Hard copies into electronic copies
  - Preserving the hard copies
  - Accessing the hard copies electronically
  - Reusing the hard copies electronically
Electronic Document Format

- Format adds special characteristics to the Electronic Document/Digital Document
  - Identify / Distinguish
  - Share
  - Web enabled
  - Convert
  - Compress
Electronic Document Format

- **Text format**
  - Applicable to documents where content is largely textual.

- **Image format**
  - Used to create/store the digital image of text pages, photographs etc.

- **Audio & Video format**
  - Used for capturing, storing & processing audio and video

- **Animation format**
Electronic Document Format

Text format

- Simple Text File Format (TXT)
- Rich Text Format (RTF)
- Portable Document Format (PDF)
- Hypertext Markup Language (HTML)
- Microsoft Word Document (DOC)
Electronic Document Format

**Image format**
- Graphics Interchange Format (GIF)
- Joint Photographers Expert Group (JPEG)
- Portable Network Graphics (PNG)
- Tagged Image File Format (TIFF)
- Bit Map Picture (BMP)
Electronic Document Format

Audio format

- Wave (Wav)
- Musical Instrument Digital Interface (MIDI)
- Real Audio (RA)
- Motion Picture Experts Group (MPEG) Audio Layer 3 (Mp3)/(Mp4)
- Windows Media Audio (WMA)
Electronic Document Format

Video format
- Motion Picture Experts Group (MPEG)
- Real Video (RA/RAM)
- Audio Video Interleave (AVI)
- Windows Media Video (WMV)
- QuickTime (MOV)
- Flash Video (FLV)
Electronic Document Format

Animated format

- Graphics Interchange Format (GIF)
- Audio Video Interleave (AVI)
- Moving Picture Experts Group (MPEG)
- Shockwave Flash (SWF)
- Quick Time (QT)
Digital Collections

- Existing electronic documents
- Future electronic documents
Managing Electronic Collections

- Electronic collections
- Electronic format
- Indexing & Retrieval
- Storage
- Access
- Delivery
- Archive
“A digital library is an organised and focused collection of digital objects, including text, images, video and audio, along with methods for access and retrieval, and for selection, creation, organization, maintenance and sharing of the collection”
Digital Library: Key Components

- Selection and Acquisition
- Organization of documents
- Indexing and storage
- Search and retrieval
- Access and delivery
- Compliance to Standards
Digital Library Software

- A key requirement for development and deployment
  - Commercial Software
  - Open Source Software (OSS)
Open Source Software

- Ability to run, change, and improve the software
- Redistribute it in modified or unmodified form
- Collaborative software method
- Source Code is made accessible
- Driver of innovation
Open Source Software

- Proprietary Software
  - development behind walls
  - Source Code
  - Vendor
  - Reliability
  - Minimal freedom to customize
  - License fee
Open Source Software

- Source Code developed over the Internet (Public View)
- Reliability -Peer Review
- No dependency on Vendor
- No license fee
- Flexibility to customize
- User-centric development
OSS Applications

- Operating System
  - Linux
- Application Software
  - Desktop: OpenOffice
  - Web Browser: Firefox
- Many other applications...
Open Source Software

SourceForge.net is the world’s largest Open Source software development website. SourceForge.net provides free hosting to Open Source software development projects with a centralized resource for managing projects, issues, communications, and code.

Registered Projects: 171,169  Registered Users: 1,885,845

Know Java? Google is looking for Java experts. Jobs in Bangalore/Hyderabad. Apply! www.Google.co.in/jobs

Open Source Software Free Open Source Software. Analysis Info, Reports, White Papers & More!
OSS for Libraries

- Database Management Systems
- Library Management Systems
- Digital Library Software
- Institutional Repository Software
- Content Management Software
- Many other applications...
Open Source Software

MARC module for Drupal 5

Submitted by codhub on Sun, 2008-01-06 23:57. MARC | OPAC | PHP

As seen on drupal.org: "This module provides a way to map data in the MARC record to Drupal content types and import sets of MARC records. The module currently supports mapping to the default node fields, taxonomy fields and CKX text fields... By importing a library's MARC records directly into Drupal as nodes, you can easily recreate your library's catalog in a rich social environment."

Very cool to see something like this pop up into another popular CMS (Scribio being the prior example).
Open Source Software
Open Source Software

New Koha Extensions Manager Site Released

The new contribs site is a place where you can find plugins, like template sets, browser plugins, migration tools and command line utilities for Koha.

Check it out at http://contribs.koha.org

Latest News

> Koha 2.2.9 Released 21 May 2007
> New Extensions Manager Website Launched 17 Jan 2007
> Koha 2.2.7 with new 4.09.59 client released 27 Nov 2006

Quick Links

> Documentation  Bugzilla
> Development Wiki  Koha at Savannah
> Discussions  ViewCVS
> Koha for Windows  Chat Information
Digital Library Software

- Greenstone Digital Library Software
  www.greenstone.org

- Kete
  http://kete.net.nz
Kete Digital Library Software

Kete.net.nz

Blog | Themes

Search: Go

Kete Resources on this Site:

- The Downloads page and the new Installation Guide
- the list of available Vendors for support of Kete
- the Kete Documentation, with discussion for getting answers to your questions and wiki-style collaborative editing
- the Training Handbook for creating Kete content
- the Blog for Kete announcements, news, and other articles
- the Themes basket to find how you can use themes to change the look of your kete, download one from the repository, or add a theme to share with the community
- the list of sites that run on Kete, add yours!
- have an problem with Kete or want to request a feature? Read how to raise your issue in "How to add a ticket"
Greenstone Digital Library Software (GSDL)

- Open Source Software promoted by New Zealand Digital Library Project
- Produced at the University of Waikato, New Zealand
- Developed and distributed in cooperation with UNESCO and Human Info NGO, Belgium
- Issued under the GNU General Public License
GSDL: Introduction

- Building digital libraries
- Distributing digital collections
- Organizing information
- Publishing on the Internet
- CD-ROM Collection
GSDL - Objective

The aim of the software is to empower users, particularly, in universities, libraries and other public service institutions, to build their own digital libraries.
GSDL: Download

- Current Version 2.82 (16 June 2009)
- This distribution includes:
  - Pre-built demonstration collection
  - Developing new collections
- Available for free to download from
  - www.greenstone.org
  - www.sourceforge.net
GSDL: Download

Latest Distribution (2.82) Release Notes
The main Greenstone download, suitable for most users. Includes everything you need to create and host Greenstone collections.
- For Windows: Greenstone-2.82-win32.exe (62MB)
- For MacOS: Greenstone-2.82-MacOS-intel.dmg (36MB)
- For GNU/Linux: Greenstone-2.82-linux (57MB)

Latest Source Distribution (2.82) Release Notes
The Source Distribution is a complete source distribution of Greenstone. The Sourcecode Top-up is used to add the source code to a binary distribution of Greenstone (above).
- Source Distribution: (zip, 55MB) (tar.gz, 52MB)
- Sourcecode Top-up: (zip, 26MB) (tar.gz, 24MB)

Nightly Release Snapshots
Every night, Greenstone snapshot releases are automatically generated for windows, linux, and mac. They are made available at our snapshots page.
GSDL - Features

Platforms:

- Runs on all versions of Windows, Unix, Mac OS-X.
- End users can install the software on desktops
- Institutional users run it on their main web server, where it interoperates with standard web server software (Apache).
GSDL - Features

Interface:

- End users access the digital library collections through the User Interface, which operates within a web browser.
- The Librarian Interface is a Java-based graphical interface
GSDL - Features

Metadata Format:

- New metadata sets can be defined using Greenstone’s Metadata Set Editor.
- The pre-defined metadata sets in GSDL are:
  - Dublin Core (DC)
  - RFC 1807
  - New Zealand Government Locater Service (NZGLS)
  - Australian Government Locator Service (AGLS)
GSDL - Features

Plug-ins:
- Used to ingest externally prepared metadata in different forms
- Plug-ins exist for - XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace, METS.
- Plug-ins are also used to ingest documents in different formats.
- For textual documents, the plug-ins are: PDF, Postscript, Word, RTF, HTML, Plain Text, Latex, ZIP Archives, Excel, PPT, Email, Source Code.
- For multimedia, the plug-ins are: Images (GIF, TIFF, JPG, JIF), MP3 Audio, MPEG, MIDI, etc.
GSDL - Features

Interoperability:

- Can harvest documents over OAI-PMH (Open Archives Protocol for Metadata Harvesting) and include them in a collection.
- Any collection can be exported to METS 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.
GSDL - Features

Multilingual:

- Uses UNICODE, which facilitates building, searching and browsing in any UNICODE compliant language

- The Users’ interface is available in a number of languages, including Bengali, Hindi and Kannada
GSDL - Features

Multimedia:

- Allows creation of non-textual multimedia such as audio, video, and pictures accompanied by textual description or metadata
- Allows searching and browsing of multimedia
GSDL - Features

Indexing:

- Can build indexes from full text documents and also metadata associated with these documents.

- The indexes can be created for various metadata fields either automatically extracted or manually assigned.
GSDL - Features

Search and Browse:

- Provides a variety of search options as full text based indexing is made possible.
GSDL: Summary of features

- Easy Installation and easy Maintenance
- Content Development - Predominantly GLI
- Interface Customization
  - Front Page Design, Header for the Digital Library, Collection Icon, Cover Images
- Collection Configuration (Collect.cfg) File
- Scalability, Flexibility
- Interoperability (Crosswalk), OAI Compliance
GSDL : Summary of features

- Accessible via any Web browser
- Server runs on Windows and Linux
- Collections can be published on CD-ROM
- Full-text and fielded search
- Flexible browsing facilities
- Metadata-based (Dublin Core)
- Collection-specific
GSDL: Summary of features

✓ Hierarchical phrase browsing supported
✓ Creates all access structures automatically
✓ Plugins — new document, metadata format
✓ Classifiers — new metadata browsers
✓ Multilingual: Documents and Interfaces
✓ Multimedia: video, audio collections
GSDL - Installation (Windows)

- Version 2.81
- To run Greenstone under any 32 bit Windows environment (that is, Windows 95/98/Me/NT/2000/XP etc.).
- Standard Windows installer program, simply download the file and double-click it to install
GSDL : Software Required

- Greenstone Digital Library Software
  http://www.greenstone.org
  (Release 2.81 onwards JRE & ImageMagick are included in the GSDL installer)

- Web Browser
  (Internet Explorer/Netscape/Mozilla)
  http://www.msn.com;
  http://www.netscape.com
  http://www.mozilla.com
GSDL - Installation
GSDL - Installation

Please choose your preferred language (Note: This installer has not been fully translated into all these languages yet)

- English
- Français (French)
- Español (Spanish)
- Deutsch (German)
- русский язык (Russian)
- 中文 (Chinese)
GSDL - Installation
GSDL - Installation

In order to access certain parts of the Greenstone interface you need a password. A user with the username 'admin' will be created for you with the password you provide. The password must be between 3 and 20 characters long.

Admin Password: admin
GSDL - Installation
GSDL - Installation
GSDL - GLI

- The installation of the GLI is integrated with the installation of GSDL.
- Java based interface for building digital library collections
- Very user-friendly approach to building digital collections.
- Targeted at four different levels of user - Assistant Librarians, Librarians, Library System Specialists and Expert users.
- Supports seven basic activities such as Download, Gather, Enrich, Design, Format, Build and Preview the collection.
GSDL - GLI
GSDL-GLI
GSDL - GLI

- **Download** - Downloading files and websites from the Internet
- **Gather** - Gathering the source documents that will comprise the collection and associate appropriate metadata element set.
- **Enrich** - Assigning metadata for each source document
- **Design** - Specifying collection configuration in terms of indexes, classifiers, etc.
- **Create** - Creating the collection
- **Format** - Designing the appearance of the Collection
- **Build and Preview** - Building and Previewing the created collection
GSDL-GLI: Download
GSDL - GLI: Gather
GSDL - GLI: Enrich

![Greenstone Librarian Interface](image)

<table>
<thead>
<tr>
<th>Element</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.Title</td>
<td></td>
</tr>
<tr>
<td>dc.Creator</td>
<td></td>
</tr>
<tr>
<td>dc.Subject and Key...</td>
<td></td>
</tr>
<tr>
<td>dc.Description</td>
<td>The extent or scope of the content of the resource</td>
</tr>
<tr>
<td>dc.Publisher</td>
<td></td>
</tr>
<tr>
<td>dc.Contributor</td>
<td></td>
</tr>
<tr>
<td>dc.Date</td>
<td></td>
</tr>
<tr>
<td>dc.Resource Type</td>
<td></td>
</tr>
<tr>
<td>dc.Format</td>
<td></td>
</tr>
<tr>
<td>dc.Resource Identif...</td>
<td></td>
</tr>
<tr>
<td>dc.Source</td>
<td></td>
</tr>
<tr>
<td>dc.Language</td>
<td></td>
</tr>
<tr>
<td>dc.Relation</td>
<td></td>
</tr>
<tr>
<td>dc.Coverage</td>
<td></td>
</tr>
</tbody>
</table>

No Metadata Element Selected
<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>The name given to the resource by the creator or publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creator</strong></td>
<td>The person responsible for the intellectual content of the resource</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>The Topic of the resource</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>A textual description of the content of the source</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>The Entity responsible for making the resource available</td>
</tr>
<tr>
<td><strong>Contributor</strong></td>
<td>A person or organization (other than the Creator) who is responsible for making significant contributions to the intellectual content of the resource</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>A date associated with the creation or availability of the resource</td>
</tr>
</tbody>
</table>
GSDL-GLI: Enrich (DC)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>The nature or genre of the content of the resource</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>The physical or digital manifestation of the resource</td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td>An unambiguous reference that uniquely identifies the resource within a given context</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>A reference to a second resource from which the present resource is derived</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>The language of the intellectual content of the resource</td>
</tr>
<tr>
<td><strong>Relation</strong></td>
<td>A reference to a related resource, and the nature of its relationship</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>Spatial locations and temporal durations characteristic of the content of the resource</td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Information about rights held in the resource</td>
</tr>
</tbody>
</table>
GSDL - GLI: Design
GSDL - GLI: Create
GSDL - GLI: Build

**Complete Rebuild**
- 83%

**Incremental Rebuild**
- Build Collection
- Cancel Build
- Preview Collection

**Import Started**
The file Workshop.jpg is being processed by ImagePlug.

**Import Finished**
1 document was considered for processing:
1 document was processed and included in the collection.

**Build Started**
Compressing text...
Creating an index based on document:text...
Creating an index based on document:Title...
Creating an index based on document:Source...
Creating information database...
GSDL - GLI: Preview
GSDL-GLI: Format
Greenstone User Interface

- Access the collection through its user interface.
- Searching the collections: Plain Search and Form-Based Search
- Browsing the documents by metadata fields
- View complete content
- Highlight matching search terms or not
- Detach the document for viewing in a different window.
- The Preference setting allows the user to change the language of the Greenstone interface.
Greenstone User Interface

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...
Greenstone User Interface

About this collection

This is a demonstration collection for the Greenstone digital library software. It contains a small subset (11 documents) of the Humanitarian and Development Libraries.

How to find information in the Greenstone demo collection

There are 5 ways to find information in this collection:
- search for particular words
- access publications by subject
Greenstone User Interface
Greenstone User Interface
Greenstone User Interface
Greenstone User Interface
Greenstone User Interface
GSDL - Customization

- Greenstone Pages are controlled by macro files (available in Greenstone → macros), images, and CSS style sheets (available in Greenstone → images)
- The user interface look-and-feel can be altered using "macros" written in a simple macro language
- Home page can be customized by editing “yourhome.dm” in main.cfg
GSDL - Customization

- Macro files
  - stored in `gsdl/macros` folder
  - each file defines one or more “packages”
    (A “package” is a group of macros)
  - listed in `etc/main.cfg`

- Collection-specific macros
  - Stored in `gsdl/collect/mycol/macros/extra.dm`
GSDL - Customization

- Collection configuration file
  - gsdl/collect/mycol/etc/collect.cfg
- Images
  - gsdl/images
- Macro files
  - gsdl/macros/*.dm
- List of macro files
  - gsdl/etc/main.cfg
GSDL: Example Collections

- development library subset
- bibliography collection
- bibliography supplement
- msword and pdf demonstration
- MARC example
- simple image collection
- greenstone archives
- OAI example
- garish
- (Documented) OAI demo collection
- CDS/ISIS example
- authentication and formatting demo
GSDL: Example Collections
GSDL - Support

- www.greenstone.org
- http://greenstone.org
- Greenstone Mailing List
  https://list.scms.waikato.ac.nz/mailman/listinfo/greenstone-users
- Greenstone Support Network India - Asia Region (greenstonesupport@iimk.ac.in)
GSDL Archives

About this collection

This is a collection of email messages from the Greenstone mailing list archives. The collection includes 10053 messages from the beginning of the mailing list in April 2000. It is updated nightly with new messages from the previous day. The mailing list is used for communicating with the entire Greenstone team, therefore the content of the messages is usually global in nature. The mailing list is also a good way of getting help with problems - someone on the team will probably be able to help you.
Welcome to GreenstoneWiki: Documentation for Greenstone

Are you a Greenstone user? Looking for documentation to help you build better collections?? This is the place for you!

Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NSC. To learn more about it or to download the software, visit greenstone.org or read the Greenstone Fact Sheet.

This documentation was prepared by the Greenstone team at the University of Waikato in New Zealand. If you’re a Greenstone user, you can help us with the documentation project. Here’s how

Contents [hide]
1. Documentation
   1.1 Manuals
   1.2 FAQ
   1.3 General documentation
   1.4 Tutorial exercises
GSDL - Support

The Greenstone Digital Library Software (http://www.greenstone.org) is an Open Source package for building and distributing Digital Libraries, which has been developed by the University of Waikato, New Zealand. Greenstone is supported by the UNESCO and the Human Info NGO Belgium for spreading the benefits of this software to developing countries. Greenstone is a powerful and flexible software which is of great potential interest to libraries and information centers and other public and private institutions in South Asia.
GSDL - Examples

Some of Indian DLs using GSDL

- Archives of Indian Labour, Noida
- Indian Institute of Management Kozhikode
- Indian Institute of Science, Bangalore
- Indian Institute of Technology, Bombay
- M G S Institute of Public Administration, Punjab, Chandigarh
- Many...
GSDL - Example DLs
GSDL - Example DLs
GSDL - Example DLs

Late Dr. T.B. Rajashekar's Academic Contributions is a digital collection comprising of his academic works. During his illustrious career spanning over 25 years, late Dr. Rajashekar wrote many articles, which includes among other things, research papers, invited talks, tutorials, special lectures and training materials. The topics covered in his articles includes Digital Libraries, Information Management, Content Management, Open Access Initiatives, Ontology etc. Please note that the collection includes only those articles, which could be traced.

The initiative to collate and publish late Dr. T.B. Rajashekar's work was taken up by NCSI-NET after his tragic and untimely death on 3rd June, 2005. The CD-Rom version of the collection comprising of his works was released on his first death anniversary, 3rd June, 2006.
GSDL - Example DLs

Indian Institute of Science Publications Database
Bangalore - 560012, India

Select a collection

Home  About  Recent Publications  Add Publications  Suggestions

Conference  Reports  Journals

Search IISc publication database – Microsoft Internet Explorer

File  Edit  View  Favorites  Tools  Help

Back  Forward  Reload  Stop  Refresh  Search  Favorites  Media

Address: http://vidya-mapsk.ncsi.iisc.ernet.in/cgi-bin/library/a=p&y=searh
GSDL - Example DLs

Indian Institute of Management Kozhikode

Digital Library

E Journals
Information Technology
Business
Economics
Generalia
Management
Psychology
Statistics
Journal alert

Videos
Environmental Science
Classics
Finance
Marketing
Sociology
IIMK Publications
News Archive
GSDL - Example DLs

**Electronic Theses and Dissertations**

**IIT Bombay**

select a collection

- Doctoral Theses
- Masters Dissertations

**Doctoral Theses**

We have a good number of theses on our database providing access to bibliographic information and an abstract since 1965 till to-date. Click on Doctoral Theses to search and browse bibliographic details and the abstract.

**Masters Dissertations**
GSDL Examples...

- Could be yours...

- ... your library!
GSDL Demo

- Demo on Collection building...
  - Creating a simple multi format collection
  - Creating a simple image collection
  - Creating a CD-ROM of an independent collection
Discussion...

Thank you for your attention!