Open Source Web Content Management Technologies for Libraries

Dr. M.G. Sreekumar
UNESCO Coordinator, Greenstone Support, South Asia
Librarian & Head, CDDL, IIM Kozhikode

Pre-Conference Tutorial
February 25, 2009

Agenda

- The Current Information Landscape
- Open Source Overview
- The OS Treasure Trove
- Categories of Open Source Software for Libraries
- Open Source Digital Library Systems
- Open Source Suite from PKP, SFU
- KOHA
- Greenstone
- DSpace
- PKP OAI Harvester
- Drupal Content Management System
- Discussion
Foreword

- Demand for improved information and knowledge management solutions - universities, enterprises and institutions
- Digital Libraries gaining increasing social attention, academic and research interest
- Need for Integrated access to disparate information resources
- Key challenge - how to create online information environments facilitating internal content publishing and single point access to internal/external information sources
- Latest DL technologies Vs Traditional libraries and knowledge management
- Options before us – Proprietary Vs Open Standards / Open Source Software
- Fortunately we have a plethora of Open Source Solutions available for Library applications

The Current Environment

- Fascinating times in the history of libraries, information systems and electronic publishing
- Possibilities of building large-scale services
  - Collections are in digital formats and
  - Retrieved over networks
- Materials are stored on computers
- Network connects the computers to personal computers on the users' desks
- In a complete digital library, nothing need ever reach paper
Feel of the Hour
Need of the Hour

Future Libraries?

What is a library and what should it be in 2012, 2020 and beyond…

- What does the academic library of the future look like?
- Where do its walls begin and end?
- On campus?
- On our desktop?
- At home?
- Does it still have a function as a separate and distinct space?
- Or has it become the first step to an all-virtual future?
- Libraries have never been more interesting, difficult and challenging…
Challenges of the Day

- Relevance of Libraries in the Google Era
- Retention of Users, especially the New Gen
- Proliferation of Content
- Diverse Datastreams - Content Categories, Publication Types
- Multimedia, Polymedia, Multiformats
- Collection Building – Acquisition, Subscriptions, Licensing...
- Copyright, Intellectual Property, Fair Use...
- Technology Complexities, Infrastructure Issues
- Publishers’ Stringent Policies / Monopolies
- Integration of legacy systems and the new genre

Information Strategy Tips

- Context = Scenarios, Paradigms
- Constant = Change
- Technology = Facilitate, and NO intimidate
- Information = The Big Picture - Landscape
- Content = Aggregate, Integrate
- Service = Markup, Market
- Capital = Human, Tacit, Values, and Users
Factors of Change

- Electronic Publishing
- Internet & Web
- Intranets
- Digital Libraries
- Knowledge Management
- E-Information: Usage and User Behavior

Experiences? Lessons? Impact? Implications?

IM: Key Goals

- Develop and manage a dynamic, unified information resource base (content repository) that gathers and organizes relevant internal and global information resources, based on a taxonomy of information needs of the enterprise, and make these available for learning and informed decision making.
IM : Key Goals...

- Support different manifestations of information sources – implicit/explicit, print/digital, local/remote, free/commercial, etc.

- Support for delivering personalized information services to staff, both on demand and in anticipation.

User Generated Internet Content:
- Blogs
- YouTube
- MySpace

And the same is TRUE with Scholarly Communication too!!
Top Tech Trends in IT / LIS

- Web 2.0 / Library 2.0
- Blogs / RSS Feeds / Wikis / Podcasts / Webcasts
- Open Source Software, Open Standards, Open URL
- User Tagging, Automated Tagging
- Web OPACs, and Interface Design
- Seamless Integration / Aggregation
- OA -> OAP + OAA
- Open Resource Discovery Tools - Google Scholar
- E-Books, E-Journals, E-Resources
- Harvesting, Federation, Metasearching
- Digital Rights Management

<table>
<thead>
<tr>
<th>L 1.0</th>
<th>L 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closed stacks</td>
<td>Open stacks</td>
</tr>
<tr>
<td>Collection development</td>
<td>Library suggestion box</td>
</tr>
<tr>
<td>Preorganized ILS</td>
<td>User tagging</td>
</tr>
<tr>
<td>Walk-in services</td>
<td>Globally available services</td>
</tr>
<tr>
<td>“Read-only” catalog</td>
<td>Amazon-style comments</td>
</tr>
<tr>
<td>Print newsletter mailed out</td>
<td>Team-built blog</td>
</tr>
<tr>
<td>Easy = dumb users</td>
<td>Easy = smart systems</td>
</tr>
<tr>
<td>Limited service options</td>
<td>Broad range of options</td>
</tr>
<tr>
<td>Information as commodity</td>
<td>Information as conversation</td>
</tr>
<tr>
<td>Monolithic applications</td>
<td>Flexible, adaptive modules</td>
</tr>
<tr>
<td>Mission focus is output</td>
<td>Mission focus is outcome</td>
</tr>
<tr>
<td>Focus on bringing ‘em in</td>
<td>Focus on finding the user</td>
</tr>
<tr>
<td>ILS is core operation</td>
<td>User services are core</td>
</tr>
</tbody>
</table>
Penetration of E-Content in Libraries

**PUBLICATION TYPES**
- E-Books, E-Journals...
- Aggregated Scholarly E-Journal Databases
- Databases, CBT/ WBT
- Portals, Vortals...
- Value added services
- Preprints, Eprints, E-Documents....

**DOCUMENT FORMATS**
- ASCII, RTF, HTML, SGML, Postscript, PDF, Proprietary, Native Application Formats
- Images, Graphics
- Audio
- Video
- XHTML, ASP, PHP, XML ...
Internally Generated

Externally Sourced

Open Access

Information Landscape [print/digital]

Lib 2.0 & Patron 2.0

Social Computing & Social Software

Books, eBooks, Journals, eJournals, Databases, Patents, Reports, Online Resources...

OA Journals, OA Archives, Scholarly Articles, ePrint Archives, ETDs, eCoursewares

Social Computing & Social Software

OA Journals,
OA Archives, Scholarly Articles, ePrint Archives,
ETDs, eCoursewares

Shift in Approaches

Traditional

Automated

Dig. Library

Limited/ Rigid

Improved

Efficient/ Flexible

AACR2
CCC
CC / LCCS
DDC / UDC
Thesauri/LCSH

AACR2
ISO 2709
CCF
MARC
Thesauri

Metadata
DCMI -- W3C
EAD, TEI, DTD
METS, MODS,
Z39.50
MARC21
OAI-PMH
What Distinguishes a DL?

- Site Neutrality (3 in 1 Access-Anytime, Anywhere by Anyone Access)
- Open Access
- Greater variety and granularity of information
- Sharing of information ‘Sharium’
- Up-to-date ness
- Always available (365*7*24)
- New forms of rendering (New Genre)

Digital Libraries: An Overview

Diagram showing the relationship between Digital Libraries and related fields such as Computing, Networking, Content, Collections, Services, and Community.
What is open source software?

- In the phrase *open source*, *source* refers to source code, the human-readable computer code which is the origin, or source, of the computer application.
- *Open* refers to the terms of access to that computer source code.
- So *open source* software is software for which the source code is freely available.
Advantages and Disadvantages

**Advantages**
- Mostly issued under an internationally accepted License
- Access to source code and ability and right to modify it
- Right to redistribute modifications to benefit wider community
- Free
- Excellent support networks
- Large and enthusiastic user base

**Disadvantages**
- Limited or no accountability
- Informal and unaccountable support channels
Different Open Source Licenses

- GNU GPL ("General Public License")
- GNU Lesser GPL
- BSD License
- Mozilla Public License
- IU Open Source License
- **And more...**
A Good Starting Point

- oss4lib: Open Source Systems for Libraries
  - [www.oss4lib.org](http://www.oss4lib.org)
**Open Source Resources**

- [Open Source Initiative](#)
- [GNU](#)
- [SourceForge](#)
Categories of Open Source Software

- **Operating Systems**
  - Linux, Free / Open BSD, Open Solaris…

- **Programming Languages**
  - Perl, PHP, Python

- **Applications**
  - Apache, Tomcat, emacs, grep, MySQL, sendmail, ssh

Open Source Software for DLs

- Linux, Apache, Tomcat, PHP, Perl, DLXS, ImageMagick, Unreal Media Server, Greenstone, DSpace, ePrints, FEDORA, CDSWare, MySQL, Darwin Streaming Server, emacs, CVS, Webalizer, LibXML, LibXSLT, Saxon, and more!
Some categories of open source library software

- Library-oriented search engines
  - Cheshire, Pears, dbWiz…
- Z39.50 toolkits
  - ZetaPerl (Perl), JAFER (Java), YAZ (C/C++), Mercury
  - Z39.50 Client…
- MARC parsers
  - MARC.pm (Perl), MARC4J (Java)
- MarcEdit
- Image processing
  - ImageMagick, tiffinfo/tiffdump

Some categories of Open Source Library software / Open Standards

- Portals
  - MyLibrary
- OAI service providers and data providers
  - PHP OAI Data Provider
  - Lots! See www.openarchives.org
- METS tools
  - Page turners, toolkits, more: see www.loc.gov/mets/
- Dublin Core
  - http://dublincore.org
Web Server

- Apache
- Lots in Java! see at…
  - http://java-source.net/open-source/web-servers

Database Management Systems (DBMS)

- MySQL
- PostgreSQL
- mSQL
- CDS/ISIS, Win/ISIS, GenISIS etc.
Web Server-Side Scripting

- PHP
  - Architecture
    - Linux, Apache, MySQL, PHP (LAMP)
    - Windows, Apache, MySQL, PHP (WAMP)

Web Services

- Apache Tomcat Web Container/Service
- Apache Cocoon Content Framework/Service
- Apache Ant Build Tool
Integrated Library Management System (ILMS)

Managing legacy systems

- KOHA
  - http://www.koha.org/
- Evergreen
- Emilda
- OpenBiblio
- phpMyLibrary
- NewGenLib
  - http://www.verussolutions.biz/

Server Log Analysis

- Webalizer
  - http://www.webalizer.org/
Z39.50 Protocol
for online/remote Search & Retrieval
(http://www.loc.gov/z3950/)

- Interoperability standard (ANSI/NISO Standard) and a software which facilitates cross-database/archives search
- A client-server protocol for searching and retrieving information from remote computer databases
Serials Manager

- **CUFTS**
  - http://researcher.sfu.ca/cufts

Citation Manager

- **Citation Manager** (from PKP, Simon Fraser University, Canada)
  - Bibliographic Management
    (http://researcher.sfu.ca/cm)
Link Resolving

- **GODOT** - Electronic (Online) Resources Management
  - [http://researcher.sfu.ca/godot](http://researcher.sfu.ca/godot)

Open Journal Publishing

- **OJS**
  - [http://pkp.sfu.ca/ojs](http://pkp.sfu.ca/ojs)
Open Conference Systems

- **OCS**: Conference workflow automation
  - [http://pkp.sfu.ca/ocs](http://pkp.sfu.ca/ocs)

Open URL Systems

- **Open URL 1.0**
  - [http://www.oclc.org/research/software/openurl/default.htm](http://www.oclc.org/research/software/openurl/default.htm)
Open Digital Libraries

- Greenstone
- DSpace
- Eprints
- FEDORA etc.

Open Access Archives / IRs

- DSpace
- Eprints
- FEDORA
- CDSWare
- Greenstone etc.
Learning Management Systems (LMS)

- E-learning Systems
  - Moodle
  - Manhattan etc.

Content Management Systems (CMS)

- Drupal
- Joomala
- MediaWiki
- Wordpress
Open Archives Harvester

- Harvester
  - http://pkp.sfu.ca/harvester
Federated Searching

- **dbWiz** : PKP Project
  - [http://researcher.sfu.ca/dbwiz](http://researcher.sfu.ca/dbwiz)
- **Google Custom Search**
Social Computing/Software

- Blogs
- Tags
- Wikis
- RSS
- Feed Aggregation etc.
Open Courses

- WWW.W3SCHOOLS.COM
- Open Courseware http://ocw.mit.edu/
Integrated Library Management Systems (ILMS)
Library Automation is the Foundation to all Library Modernization Activities

- Proprietary Solutions are Expensive
- HighTech ILMS Packages are very much available in the OSS Domain
- Selection of the Software is to be done judiciously as the success or failure may have lasting impacts
- For this tutorial, KOHA ILMS is showcased

Features of ILMS

- Separates software functions into discrete programs called modules, and integrated then into a unified interface.
- **Acquisitions** (ordering, receiving, and invoicing materials),
- **Cataloging** (classifying and indexing materials),
- **Circulation** (lending materials to patrons and receiving them back),
- **Serials** (tracking magazine and newspaper holdings), and the
- **OPAC** (public search/retrieval interface for users). Each patron and item has a unique ID in the database that allows the ILS to track its activity.
Features of KOHA

- Koha is an open-source Integrated Library System (ILS).
- It supports global standards including MARC 21 bibliographic format;
- Z 39.50 information retrieval protocol;
- Web-centric architecture (no additional software/utility is required at the client side).
- Provides tremendous freedom for customization.
- Has all the modules of LMS (Acquisition, Cataloguing, Circulation, OPAC, Membership Management, System Administration, Serial Control, etc.)
- Web based OPAC system (allows the public to search the catalogue in the library, from campus and at home).

Features of KOHA...

- The software is UNICODE compliant. The creation and retrieval of Indic script based documents is possible.
- Export/Import and backup/restoration facilities are available.
- Includes features of fourth generation Library Management Software (LMS).
- Runs on Linux, Unix, Windows and MacOS platform.
- Koha uses Apache Web server, MySQL as backend RDBMS and Perl (Programming language for server-side scripting).
- All these softwares are open source.
KOHA on Windows

Installation

(XP/2000 Server)

OS / Software Requirements

- Windows XP/2000/Server
- Webserver- Apache
- Database- MySQL
- Programming Language - Perl
Installation

- Apache-2.0.63
  http://www.apache.org/dist/httpd/binaries/win32/apache_2.0.63-win32-x86-no_ssl.msi

- MySQL-4.1

- ActivePerl-5.8.8.22-MSWin32-x86
  http://downloads.activestate.com/ActivePerl/Windows/5.8/ActivePerl-5.8.8.22-MSWin32-x86-280952.msi

- Koha (2.2.9-W32-R1)
  http://www.koha.rwjr.com/downloads/Koha2.2.9-W32-R1.EXE

Click on Apache Source and Run
Server Details …

- **Network Domain** - localdomain
- **Servername** - localhost
- **Administrators Email** - your email-ID
- **Install for all Users** on Port 80
- Select Set up Type as “Typical”
- Destination Folder – **Default**
- Go ahead and **Install**
- Once the Apache installation wizard completes its task, you will find a Feather Icon in your System Tray
You can now Start/Stop Apache

Open: services.msc

[Image of the Run dialog box with the option to open services.msc]

[Image of the Services (Local) window with Apache listed]
Installing MySQL

Click on MySQL Setup
Create a Folder “MySQL” in C:\

Welcome to the Setup Wizard for MySQL Server 5.0

The Setup Wizard will install MySQL 5.0 release 5.0.3b on your computer. To continue, click Next.

WARNING: This program is protected by copyright law.

Setup Type:
Choose the setup type that best suits your needs.

- **Typical**
  - Common program features will be installed. Recommended for general use.

- **Complete**
  - All program features will be installed. (Requires the most disk space.)

- **Custom**
  - Choose which program features you want installed and where they will be installed. Recommended for advanced users.
You will be asked for MySQL.Com Sign Up which you can Skip by selecting Skip Sign-Up

Select Standard Configuration
MySQL Server Instance Configuration Wizard

Configure the MySQL Server 4.1 server instance.

Processing configuration ...

- Prepare configuration
- Write configuration file (C:\mysql\my.ini)
- Start service
- Apply security settings

Configuration file created.
Windows service MySQL installed.
Service started successfully.
Security settings applied.

Press [Finish] to close the Wizard.

Finish
Installing Perl

Click on Active Perl Setup
Create a Folder “usr” in C:\
Installing KOHA

Click on Koha Source File.

Welcome to Koha V2.2.9 on Windows Setup program. This program will install Koha V2.2.9 on Windows on your computer.

It is strongly recommended that you exit all Windows programs before running this Setup Program.

Click Cancel to quit Setup and close any programs you have running. Click Next to continue with the Setup program.

Supporting Software found on this Computer:
- Apache Version 2.2.9
- MySQL Version 4.1.22
- ActiveState Perl Version 5.8.8 Build 917
View Release Notes

RELEASE NOTES
-------------
Koha is the first Open Source Integrated Library System.
Koha is a feature-rich open-source ILS, developed initially in New Zealand by
Kairos Communications Ltd and first deployed in January of 2003 for Horowhenua
Library Trust. It is currently maintained by a team of software providers and
library technology staff from around the globe.
Koha 2.2.9 is more than 250,000 lines of code
(see: http://koha.sourceforge.net/doc/), contributed by about 50 different
developers (48, translators).

More than 100 libraries are registered as users on wiki.koha.org, and we are sure
that at least 300 libraries use the software. There are 2 projects derived from Koha.

WARNINGS
-------
* The interface is tested only with Mozilla/Firefox. Should work.

Choose Installation Type

Please select type of installation:

- **Full**: Choose Full if you want to install all options and want the install to
automatically configure Apache, MySQL,
and Perl to run Koha. (Existing
configuration files will be backed up and
then changed.)

- **Custom**: Choose Custom if you want to choose
what options are installed and which
configuration files are changed. You will
have to manually configure any options
you don't install.
Backup Replaced Files

This installation program can create backup copies of all files replaced during the installation. These files will be used when the software is uninstalled and a rollback is requested. If backup copies are not created, you will only be able to uninstall the software and not roll the system back to a previous state.

Do you want to create backups of the replaced files?

- Yes
- No

Please select the directory where the replaced files will be copied.

Backup File Destination Directory

\%\user\Koha2.9\backup

Browse...

Start Installation

You are now ready to install Koha 2.2.9 on Windows.

Press the Next button to begin the installation or the Back button to reenter the installation information.
KOHA Configuration & Administration

Librarian Interface

http://intranet
KOHA OPAC
[User Interface]
http://opac

It works!

Opac Screen
KOHA Demo

- Intranet
- Opac
- Downloading MARC Records from Library of Congress
- Importing MARC Records into Koha
- Search / Retrieval in the Opac

Koha Manuals & Must Visits

- Koha Home - http://www.koha.org
- A comprehensive manual on Koha 3.01.x (Linux) is available at:
  - https://sites.google.com/a/liblime.com/koha-manual/
- For Windows, it is available at
- Documentation - http://www.kohadocs.org
- Discussion forums - http://koha.org/community/mailing-lists.html
Greenstone: Open source Software for Building Digital Library Collections

What are digital libraries for?

- Knowledge/content management
  - Manage and access internal information assets
- Scholarly communication, education, research
  - E-journals, e-prints, e-books, data sets, e-learning
- Access to cultural collections
  - Cultural, heritage, historical & special collections, museums, biodiversity
- E-governance
  - Improved access to government policies, plans, procedures, rules and regulations
- Archiving and preservation
- Many more …
DL Software: Alternatives

- What are your expectations?
- Develop local web-based application?
- Commercial DL solution?
- Adopt open source software?
  - Greenstone
  - Eprints
  - DSpace
  - Fedora…

Digital Library Technologies

- Interoperability
- Unified interface for heterogeneous libraries
- Metadata mapping across different libraries
- OAI-compliant data and service providers
- Multilingual digital libraries
- Scalable digital library architectures
- Publication tools
- Searching tools
DLs: Workflows and Processes

- Content selection
- Content acquisition
- Content publishing
  - Metadata preparation
  - Content loading
- Content indexing & storage
- Content access & delivery
- Preservation
- Access management
- Usage monitoring and evaluation
- Networking and interoperation
- Maintenance

DL Software: Key requirements

- Document types (book, journal article, lecture …)
- Document formats (text, PDF, Word, PS, …)
- Content acquisition (online and offline)
  - Metadata description, content tagging
  - Content uploading
- Indexing and retrieval
  - Structured/ full text indexing
  - Automatic metadata extraction
- Storage
  - Data compression
  - Efficient storage for metadata
  - Efficient location of metadata and documents
- Access and delivery
  - Structured search, browse, hierarchical browsing
  - CD-ROM distribution
DL Software: More requirements

• Scaling up – for large collections
• Multilingual support
• Access management and security
• Usage monitoring and reporting
• Standards compliance
  – XML, Dublin Core, Unicode
• Interoperation
  – OAI, Z39.50 compliance, MARC21…
Complete DL Systems

- Greenstone
- DSpace
- Eprints
What is the Greenstone software?

- Software suite for building, maintaining, and distributing digital library collections
- Comprehensive, open-source
- Developed by New Zealand Digital Library Project at the University of Waikato
- Distribution and promotion partners:
  - UNESCO
  - Human Info NGO, Belgium
  - NCSI, Bangalore; UCT, Cape Town; Dakar, Senegal; Almaty, Kazakhstan; …
  - You!

Greenstone Features

- Supports creation and management of collections by administrator(s)
- Web interface for search and retrieval
  - Customizable metadata
  - Supports full text search of content
- Extensive document filters
  - Word, Excel, PowerPoint, PDF, ...
  - Can extract metadata from documents
- Many ways to build a collection, including:
  - Local files
  - Retrieve web sites
  - Retrieve objects via OAI-PMH
# Greenstone Features...

<table>
<thead>
<tr>
<th>Open Source Philosophy</th>
<th>Full-text mirroring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interfacing &amp; Content Delivery via Web</td>
<td>Text Level Penetration</td>
</tr>
<tr>
<td>Multi S/W Platform</td>
<td>Data Compression</td>
</tr>
<tr>
<td>Multi Lingual Support</td>
<td>Password protection</td>
</tr>
<tr>
<td>Multi Formats</td>
<td>Administrative Functions</td>
</tr>
<tr>
<td>Structured Metadata in XML using DC</td>
<td>Concurrent &amp; Dynamic Content Development</td>
</tr>
<tr>
<td>Metadata Extraction</td>
<td>Uniform Presentation</td>
</tr>
<tr>
<td>Searching &amp; Browsing</td>
<td>Publishing on CDROMs</td>
</tr>
<tr>
<td>Plug-ins for Documents</td>
<td>International Presence</td>
</tr>
</tbody>
</table>

# Greenstone Features contd...

<table>
<thead>
<tr>
<th>Easy Installation</th>
<th>Full-text mirroring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy Maintenance</td>
<td>Text Level Penetration</td>
</tr>
<tr>
<td>Content Development (3 alternate ways)</td>
<td>Data Compression</td>
</tr>
<tr>
<td>Predominantly GLI now - since (V. 2.41)</td>
<td>Password protection</td>
</tr>
<tr>
<td>Hierarchy Structure</td>
<td>Administrative Functions</td>
</tr>
<tr>
<td>Interface Customization</td>
<td>Concurrent &amp; Dynamic Content Development</td>
</tr>
<tr>
<td>– Front Page Design, Header for the Digital Library, Collection Icon, Cover Images</td>
<td>Uniform Presentation</td>
</tr>
<tr>
<td>Collection Configuration (Collect.cfg) File</td>
<td>Publishing on CDROMs</td>
</tr>
<tr>
<td>Scalability, Flexibility</td>
<td>International Presence</td>
</tr>
<tr>
<td>Interoperability (Crosswalk), OAI Compliance</td>
<td>Lifeline: Listserv / E-Group / Archives</td>
</tr>
</tbody>
</table>
The power of open source: Greenstone uses ...

- **Ghostscript**: Interpreter for Adobe Postscript documents (Postscript plugin)
- **Kea**: Keyphrase extraction program (to generate metadata)
- **pdf2html**: Converter for PDF documents (PDF plugin)
- **rtf2html**: Converter for RTF documents (RTF plugin)
- **TextCat**: Detects languages and document encodings
- **xvWare**: Converter for Word documents (Word plugin)
- **xhtml**: Converter for Excel/Powerpoint documents (plugins)
- **XML::Parser**: Parses XML documents, used to read and write Greenstone’s internal XML document format

and ...

- **MG**: Creates compressed full-text indexes and performs searches
- **GDBM**: Database used for metadata etc
- **wget**: Downloading pages from the Web when creating collections
- **YAZ**: Client and server implementation of Z39.50
- **Stemmer**: English language stemmer

---

- **GCC**: C/C++ compiler
- **CVS**: Version control system
- **Perl**: Used for plugins etc
- **Apache**: Web server used by many Greenstone installations
- **OAI-PMH**: OAI Performance
Example Greenstone collections

• Rapid growth in use
• International – Many Countries…China, Germany, India, UK, USA, Russia, Malaysia, Singapore… – Almost all countries/Continents
• Increasing activity on Greenstone mailing list
• Promotion by UNESCO – “deployment of DL’s for sharing public domain information”
• Wide variety of DL collections have been developed in several languages
  – historical, educational, cultural, and research

Greenstone Technology

- Runs on Windows (back to 3.1), Linux, Mac OS X, Unix
- Written in C++, Perl, and Java
- Uses MG/MG++ search engine
- Several different Web and Java/Swing user interfaces for various functions
- Web interface for user access
Greenstone on Windows

Version 2.81rc

Greenstone & Associated Softwares

- Greenstone 2.81 (http://www.greenstone.org)
- Java Runtime Environment (JRE) (http://java.sun.com)
- ImageMagick (http://www.imagemagick.org)
- Ghostscript (http://www.cs.wis.edu/~ghost/)
- Module for CD-ROM Publishing (http://www.greenstone.org)
- Additional Language Pack (http://www.greenstone.org)
Installing Greenstone

Softwares/Files Required

1. Java Runtime Environment (JRE)
2. ImageMagick
3. Ghostscript
4. Greenstone now comes with “ImageMagick” and “Ghostscript” bundled and hence they don’t need to be installed
5. Greenstone 2.81
6. Module for CD-ROM Publishing
7. Additional Language Pack
Installing... Java Runtime Environment (JRE)

Step 1. Check and Remove any Java Presence

Step 2. Locate the jre-1_5_0_05-windows-i586-p and Click to Install
Greenstone Installation

Version 2.81rc

Language

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

- [ ] English
- [ ] Français (French)
- [ ] Español (Spanish)
- [ ] Deutsch (German)
- [ ] русский язык (Russian)
- [ ] 中文 (Chinese)
Greenstone 2.81rc installer

Greenstone 2 is software for building and sharing digital libraries.

It is recommended that you uninstall any previous installations of Greenstone 2 before running this installer.

Greenstone 2 is licensed under the GNU General Public License v2

GNU GENERAL PUBLIC LICENSE
Version 2, June 1991

Copyright (C) 1988, 1991 Free Software Foundation, Inc.
675 Mass Ave, Cambridge, MA 02139, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your own free program.

(footnotes)
Greenstone’s Interfaces
Digital Library (User + Librarian) Librarian Interface (GLI)

Invoking Greenstone on Browser
Opening Greenstone on Browser

Digital Library Server

Greenstone Digital Library

Greenstone Digital Library Software

GREENSTONE DIGITAL LIBRARY

version 2.81rc

Enter Library

Press this button to begin using the library

Collections

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 is designed and distributed under the cooperation with UNESCO and the Human Info NGO. It is open-source software, available from https://greenstone.org in the terms of the GNU General Public License.

The aim of the software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Digital libraries are rapidly emerging as a way to disseminate and access information in the fields of education, science and culture around the world, and particularly in the developing countries. It is expected that this software will encourage the effective deployment of digital libraries in diverse environments and support the public domain.

This software is developed and distributed as an international cooperative effort established in August 2001 among partners.

New Zealand Digital Library Project at the University of Waikato

Greenstone software was one of the projects supported by the Communication Sub-Commission of the New Zealand National Commission for UNESCO as part of New Zealand's contribution to UNESCO's programme.
Opening the GLI
GLI Functions

- Establish new collection (or work on old)
- Select files to include in collection (Gather)
- Enrich files with metadata (Enrich)
- Select Plugins, Indexes, Classifiers (Design)
- Build Collection (Create)
- Format and Control Display (Format)
- Customize Appearance
- Preview Collection
Collection Building…

- Greenstone used to have three modes of collection building, viz., **Command Line, Web Interface and the GLI** (Greenstone Librarian Interface).
- Progressing with version 2.4x, the GLI got strengthened as well as popularized.
- 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’, ‘Format’ and ‘Create’ panel for making collection.

Customization

- 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.
Customizing with macros

- Let you customize presentation
- Present pages in different languages
- Print variables into the page text (e.g. number of search hits)

- Macro files
  - stored in `greenstone2/macros` folder
  - each file defines one or more “packages”
    (A “package” is a group of macros)
  - loaded on startup
    (note difference between Local and Web Library)
  - listed in `etc/main.cfg`

- Collection-specific macros
  - Stored in `greenstone2/collect/mycol/macros/extra.dm`
  - Or include argument `[c=collectionname]` for each macro
Collection configuration

• Collection configuration file determines content conversion, extraction and building of indexes and browsing structures
  – indexes, classifiers, plugins

• Presentation of search/browse results and collection interface is determined by “format” strings and “macros”
Greenstone Support for South Asia has been launched

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.

This support effort is coordinated by the CCDC at IIM Kozhikode in collaboration with the Greenstone development team in order to ensure effective promotion and support for Greenstone in South Asia.

Greenstone Support for South Asia

Examples at www.greenstone.org

Greenstone Support for South Asia

greenstonesupport.iimk.ac.in
DSpace is a groundbreaking digital institutional repository that captures, stores, indexes, preserves, and redistributes the intellectual output of a university’s research faculty in digital formats.”

- Developed jointly by MIT Libraries and Hewlett-Packard
- Licensed under BSD distribution license
- [www.dspace.org](http://www.dspace.org)
DSpace

- Supports submission of, management of, and access to digital content
  - Formats: text, images, audio, video
- Organized based on organizational needs of a large university
  - Communities and collections
DSpace Data Model

- Community
- Collection
- Item
- Bundle
- Bitstream
- Bitstream Format
- Dublin Core Record

DSpace Features

- Digital preservation
  - Persistent IDs, support levels for different file formats
- Access control
- Versioning
- Search and retrieval
  - Based on qualified Dublin Core metadata
- OAI-PMH data provider
  - To support metadata harvesters
DSpace Technology

- OS: Unix or Linux
- Written in Java
- PostgreSQL relational database
- Provides complete Web user interface, but Java APIs available

DSpace Architecture
DSpace Software / Utilities

1. **Java SDK 1.4.2**  
   http://java.sun.com/javase/downloads/index_jdk5.jsp

2. **Apache Maven 2.0.54**  
   http://maven.apache.org/download.html

3. **Tomcat 5.0.28**  
   http://tomcat.apache.org/download-55.cgi

4. **Apache Ant 1.6.5**  
   http://ant.apache.org/bindownload.cgi

5. **PostgreSQL 8.0.2**  
   http://mirror.tomato.it/ftp/pub/PostgreSQL/binary/v8.0.2/win32/

6. **DSpace 1.5x / 2.x**  
   http://downloads.sourceforge.net/dspace/dspace-1.5.0-release.zip

Invoking DSpace

- Start the Tomcat Service
- Start PostgreSQL Service
DSpace JSP User Interface

Welcome to DLIS Open Access Archive

Department of Library and Information Science
Welcome to DLIS Archive

Search DSpace

Enter some text in the box below to search DSpace.

Communities in DSpace

Choose a community to browse its collections.
- Computer Science and Information Technology

DSpace XML User Interface

Manakin

DSpace/Manakin Repository

Welcome to the new Manakin interface to the DSpace digital repository. DSpace is a digital service that collects, preserves, and distributes digital materials. Repositories are important tools for preserving an organization’s legacy; they facilitate digital preservation and scholarly communication.

Search DSpace

Enter some text in the box below to search DSpace.

Communities in DSpace

Select a community to browse its collections.
- Computer Science and Information Technology
WAMP

- Windows
- Apache
- MySQL
- PHP

Installing WAMP Server

- Click and Install
- Give the default installation location to C:\wamp
- Click the WAMP Icon at the System Tray
- Test for Local Host in the Browser – You will get the WAMP Page
OAI Harvester

- **Open Archives Harvester**
- The PKP 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.
Installing PKP Harvester

- Download PKP Harvester from http://pkp.sfu.ca/?q=harvester
- Unzip the .gz and the .tar files to C:\wamp\www\harvester2
- Open the file index.php in the Web Browser - http://localhost/harvester2/index.php
- Give the necessary parameters/data as required
- Click on “Install Harvester”
- Check for successful installation/debug
- The OAI Harvester is now available at http://localhost/harvester2/

Adding and Managing Archives

- Download PKP Harvester from http://pkp.sfu.ca/?q=harvester
- Unzip the .gz and the .tar files to C:\wamp\www\harvester2
- Open the file index.php in the Web Browser - http://localhost/harvester2/index.php
- Give the necessary parameters/data as required
- Click on “Install Harvester”
- Check for successful installation/debug
- The OAI Harvester is now available at http://localhost/harvester2/
The image below shows the main administration page:

Site Administration

Site Management
- Site Settings
- Language
- Cron tasks
- Plugins
- Pending Tasks

Archives
- Add Archive
- Manage Archives

Administrative Functions
- System Information
- Export User Data

[Image of the administration page]
Drupal
Content Management System

www.drupal.org
Database Created
Give “Install.php” at Localhost
Drupal Core

- Drupal Versions
- Hook
- Themes
- Nodes
- Blocks
- Menus
- URL’s
- Taxonomy
- Users
Westlake Porter Public Library

http://www.westlakelibrary.org

Monterrey Technology Library

http://biblioteca.mty.itesm.mx
School Library System of Genesee Valley Boces

http://sls.gvboces.org

Hamline University Law Library

http://law.hamline.edu/library/law-library.html
Drupal & Libraries Trends

- Libraries are using it to distribute the management of website content
- Drupal is being used to empower students, faculty, and staff to create their own content
- Most libraries who have done one website in Drupal have had positive results and are expanding their Drupal efforts
- Many smaller projects are being created by librarians
- Drupal is often used to present new and different views of library content
- Commonly used for library staff intranets
Five Modules Librarians Should Know About

1. Bibliography Module
   - Manage & display lists of scholarly publications
   - bibTex and Endnote Import/Export
   - http://drupal.org/project/biblio

2. Marc Record
   - Creates a library catalog by importing Marc records
   - http://drupal.org/project/marc

3. Book Review
   - Publish book reviews
   - Fields for unlimited authors, booktitle, cover image, publisher, copyright, isbn, # of pages, price, synopsis, table of contents and the review text
   - http://drupal.org/project/bookreview

4. Millennium
   - Searches Innovative Interfaces’ Millennium WebOpac for MARC records
   - Creates a Drupal-based library catalog
   - http://drupal.org/project/millennium

5. Z39.50 Search
   - Searches the world’s libraries and presents results
   - http://drupal.org/project/z3950
Open Web Design

Open Web Design is a community of designers and site owners sharing free web design templates as well as web design information. Helping to make the Internet a better place.

1000s of Designs

Open Web Design competes: Prize 2000 Cash!
Deadline: May 11, 2008 @ 12:00pm ET
Please see the contest information area for details. Themes for this contest is "Open Design." Deadline for entries will be Sunday, May 11, 2008. The judges will be you, and our partner The Web Design Group. For examples of submissions, see Open Web Design.
DL - Hardships

- Copyright Issues
- Technology Complexities
- Infrastructure Issues
- Publications/Formats – Diverse Datastreams
- Digital Objects/Formats - Multiple
- Publishers’ Policies – Stringent, Inconsistent
Major Tasks

- Content identification (internal / external)
- Content Creation
- Content Collation/Signposts
- Organisation
- Updation
- Retrieval / Dissemination
- User Training
- Archiving

DIGITAL LIBRARY ARCHITECTURE
Thank you Questions?