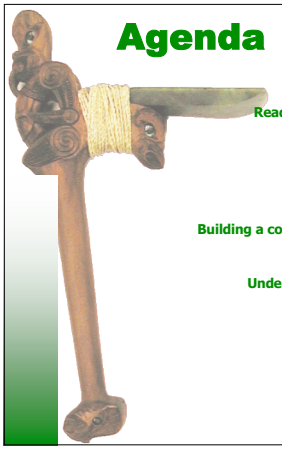
Greenstone: Open source software for building digital library collections

Ian H. Witten

Computer Science Department
Waikato University
New Zealand

http://greenstone.org
http://nzdl.org



Agenda

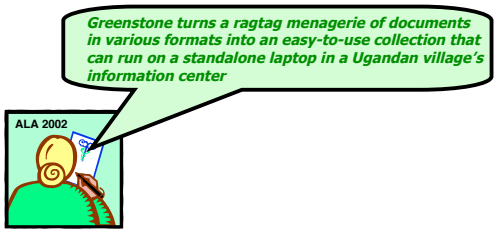
Overview
What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections

Librarian interface
Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)

Advanced stuff
Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done

Reaching out
Serving and acquiring OAI
DSpace and METS
Greenstone3

What we wanted



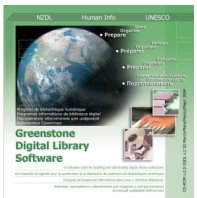
What we wanted

- "Collections" of digital material
- Individualized, depending on metadata etc
- Up to several Gb of text …
- … + associated images, movies, whatever
- Fully searchable
- Served on WWW, or published on removable media
- Run anywhere, on any computer
- Fully internationalized
- Non-exclusive: documents and metadata in any format
- Non-prescriptive: standard and non-standard metadata

What we got: Greenstone

Access	❖Accessible via any Web browser ❖Server runs on anything (all Windows + Unix + Mac) ❖Collections can be published on CD-ROM/DVD ❖Trivial to install ❖GUI interface for building and publishing collections
Searching/ browsing	❖Collection-specific ❖Full-text and fielded search ❖Flexible browsing facilities ❖Metadata-based (Dublin Core recommended) ❖Creates all access structures automatically
Extensible	❖Plugins — new document, metadata formats ❖Classifiers — new metadata browsers
Multi-*	❖Multilingual: Documents *and* interfaces ❖Multimedia: image, video, audio collections exist ❖Multiformat: Documents *and* metadata

UNESCO: Distributing Greenstone DL software



Sustainable development

"Give a man a fish, feed him for a day
Teach a man to fish, feed him for life"

Greenstone software on CD-ROM

- **GNU licensed**
- **Fully documented … in English/French/Spanish/Russian**
- **Language interfaces … Arabic Chinese Czech … Thai Turkish**
- **Unix/Windows/Mac OS-X**
- **Trivial to install**
- **GUI interface for gathering, enriching, building …**
- **Serve collections on Web or write them to CD-ROM**
- **Document formats: HTML, Word, PDF, PS, plain text, e-mail**
- **Metadata formats: XML, DC, OAI, MARC, …**

download from *http://greenstone.org*

Greenstone facts

Distribution
- Open source: Gnu GPL
- Distributed via SourceForge since: No
- Average downloads: 5000/month sinc
- Humanitarian CD-ROMs produced: 30
- Distribution for each one: 5000/year

International
- Languages for interface: 38
- Languages for full software + manuals
- Countries represented on email lists:
- UNESCO training courses in: Bangalore, Almaty, Dakar, Suva, ..

UN Agencies
- UNESCO, Paris ("Information for All" p
- FAO, Rome (Info Management Resour
- UNU, Japan (CD-ROM collections of UI

Technical centers
- University of Waikato, New Zealand
- Indian Institute of Sciences, Bangalore
- University College, London
- University of Cape Town, South Africa
- University of Lethbridge, Canada

Arabic
Armenian
Bosnian
Croatian
Deutsch (German)
English
Español (Spanish)
Farsi
Finnish
Français (French)
Galician
Georgian
Hebrew
Hindi
Indonesian
Italiano
Kannada
Māori
Nederlands (Dutch)
Serbian
Thai
Turkish
Ukrainian
português-BR (Brasil)
português-PT (Portugal)
Česky (Czech)
Ελληνικά (Greek)
Русско (Russian)
Қазақ (Kazakh)
日本語 (Japanese)
汉语 (Chinese)

Sample collections at greenstone.org

International	Argentina Human Rights Commission	Argentina
	Tasmania State Library	Australia
	Peking University Digital Library	China
	Gresham College, London	England
	University of Applied Sciences, Stuttgart	Germany
	Association of Indian Labour Historians, Delhi	India
	Indian Institute of Management, Kozhikode	India
	Indian Institute of Science, Bangalore	India
	Vimercate Public Library, Milan, Italy	Italy
	Netherlands Institute for Scientific Information Services	Netherlands
	Philippine Government Information Network	Philippines
	Mari El Republic, Russia	Russia
	Slavonski Brod Public Library, Slovenia	Slovenia
	Vietnam National University	Vietnam
	Welsh Books Council	Wales
U.S.	Auburn University, Alabama	
	Detroit Public Library	
	Hawaiian Electronic Library	
	ibiblio project, University of North Carolina	
	Illinois Wesleyan University	
	LeHigh University, Pennsylvania	
	New York Botanical Garden	
	University of California at Riverside	
	University of Chicago Library	
	University of Illinois	
	Texas A&M University	
	Washington Research Library Consortium	

Standards

Metadata
- Can use any metadata set, Dublin Core supplied
- Plugins for

XML	**Refer**
MARC	**OAI**
CDS/ISIS	**METS**
ProCite	**DSpace**
BibTex	

- METS can be used as Greenstone's internal representation

Serving
- Web
- Can publish Greenstone collections on CD-ROM
- Can publish Greenstone collections on OAI
- Export collections to METS
- Export collections to DSpace (*ready for DSpace's batch import program*)

Documents
- Plugins for

PDF	**ZIP**	**Images** (any format: GIF, JPEG, TIFF ...)
PostScript	**Excel**	**MP3**
Word, RTF	**PPT**	**Ogg Vorbis**
HTML	**Email**	**UnknownPlug** (e.g. for audio, MPEG, Midi)
Plain text	**Source code**	
Latex	**RealMedia**	

What is open-source software?

- "The **basic idea behind open source** is very simple: When programmers can read, redistribute, and modify the source code for a piece of software, the software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing."

  from *www.opensource.org*
- Anyone can redistribute the software, even for a fee
- Source code must always be available

The power of open source: Greenstone uses ...

Ghostscript	**Interpreter for Adobe Postscript documents (Postscript plugin)**
Kea	**Keyphrase extraction program (to generate metadata)**
pdftohtml	**Converter for PDF documents (PDF plugin)**
rtftohtml	**Converter for RTF documents (RTF plugin)**
TextCat	**Detects languages and document encodings**
wvWare	**Converter for Word documents (Word plugin)**
Xlhtml	**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
❖ Lucene	Alternative, incremental, full-text search engine
❖ GCC	C/C++ compiler
❖ CVS	Version control system
❖ Perl	Used for plugins etc
❖ Apache	Web server used by many Greenstone installations

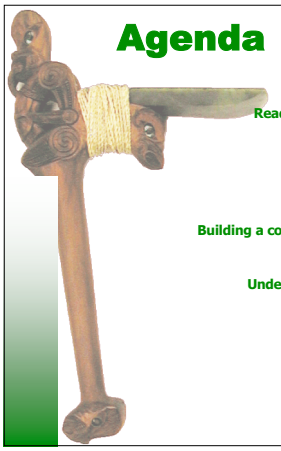
Example

Humanity Development Library

for sustainable development and basic human needs

- ❖160,000 pages
- ❖30,000 images
- ❖800 books
- ❖430 magazines
- ❖340 kg
- ❖US$20,000
- ❖CD-ROM
- ❖US$1
- ❖Win3.1x upward
- ❖Stand-alone
- ❖*and* intranet server
- ❖Web browser user interface

Global Help Project, Antwerp (+ UN agencies)



Agenda

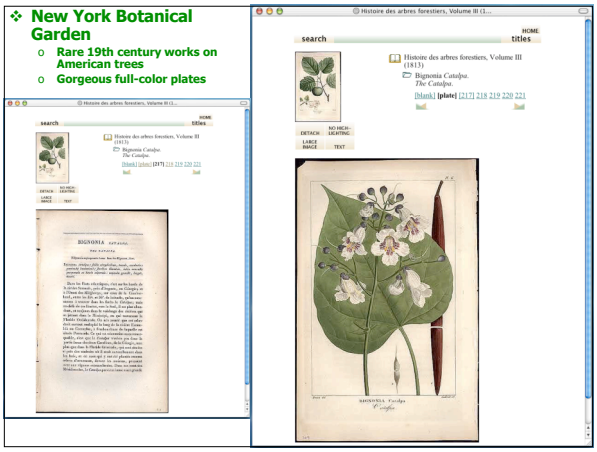
Overview
What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections

Librarian interface
Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)

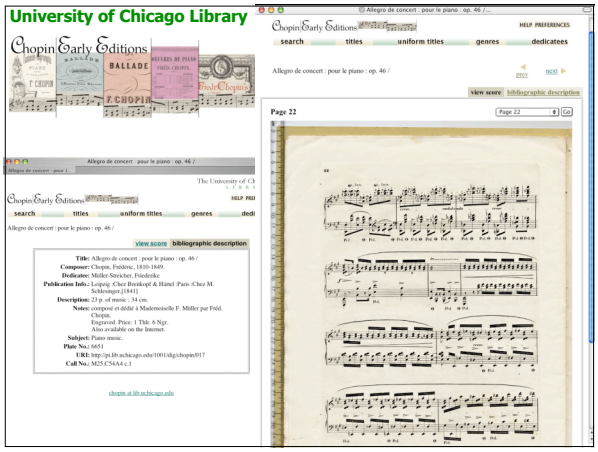
Advanced stuff
Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done

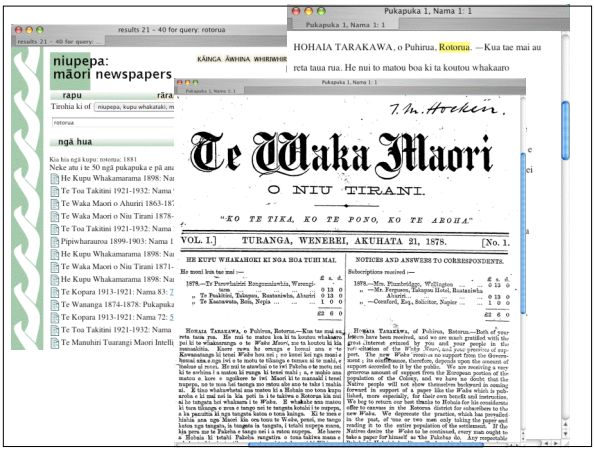
Reaching out
Serving and acquiring OAI
DSpace and METS
Greenstone3

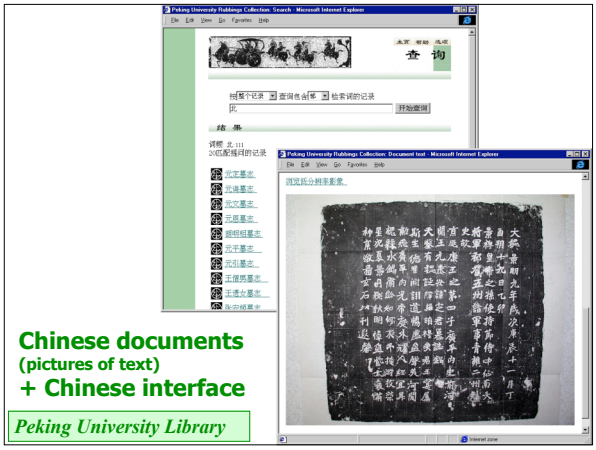
❖ **New York Botanical Garden**
- Rare 19th century works on American trees
- Gorgeous full-color plates



University of Chicago Library

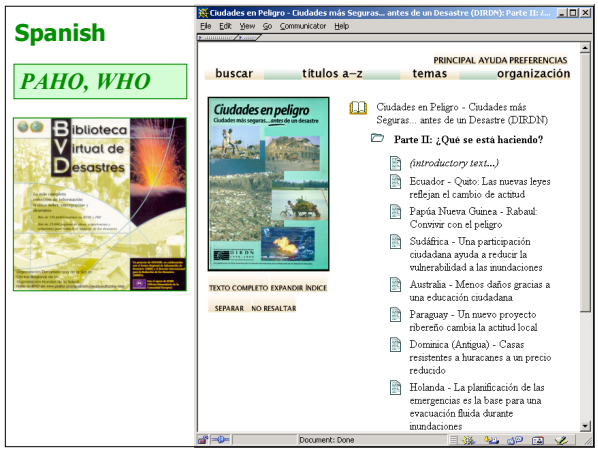





Chinese documents
(pictures of text)
+ Chinese interface
Peking University Library

Chinese
(Chinese & English interfaces)
Classic Chinese literature

French
UNESCO, Paris
SAHEL point DOC
UNESCO - SAHEL point DOC - CD-ROM 1.0


Spanish
PAHO, WHO

Russian
Mari El Republic
http://gov.mari.ru/gsdl

Agenda
Overview
What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections
Librarian interface
Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)
Advanced stuff
Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done
Reaching out
Serving and acquiring OAI
DSpace and METS
Greenstone3

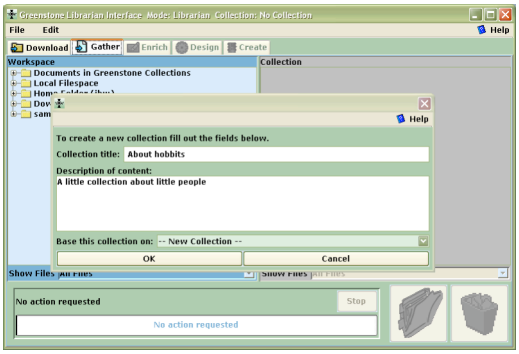
The Greenstone Librarian Interface (GLI)

- Building collections
- Interactive Java program
- Runs on anything
- Build a collection on the computer you are on
- … plus new applet version
- Includes metadata editor

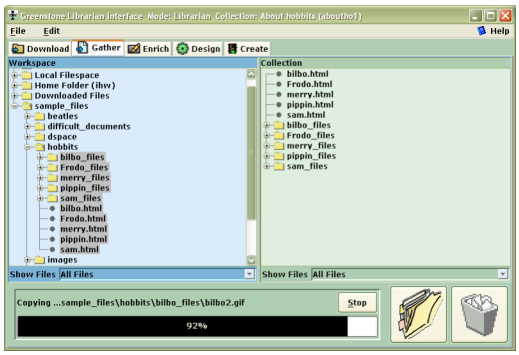
Caveat: cannot deal with such huge metadata collections as Greenstone can

(Tutorial exercise: small collection of HTML files)
Invoke GLI: build a small collection of HTML files
- **Gather**
- **Create**
- **Look at extracted metadata**
- **Set up shortcut in the Librarian interface**

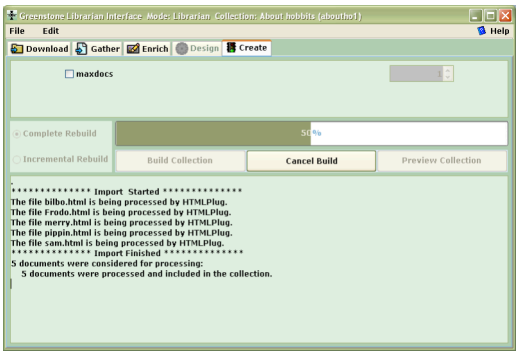
Create a new collection



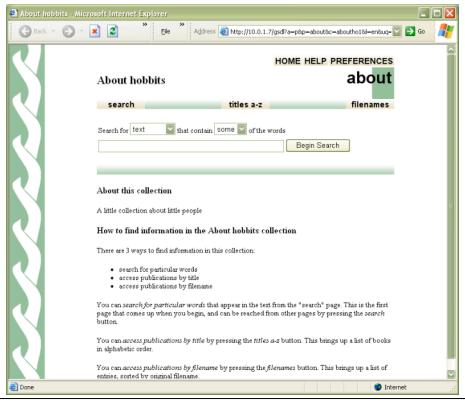
Gather: Gather the files together



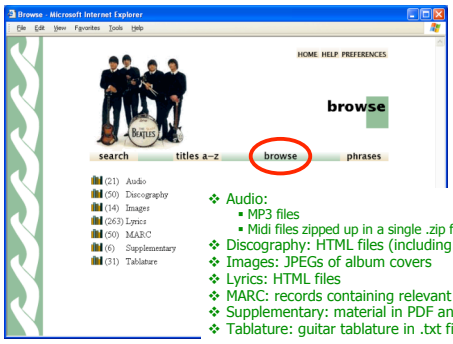
Create: Build the collection



Preview: admire the result

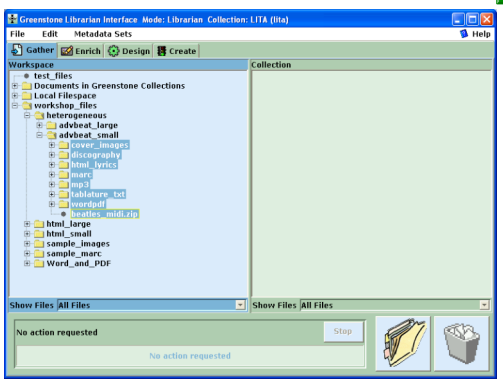


An example: Beatles collection

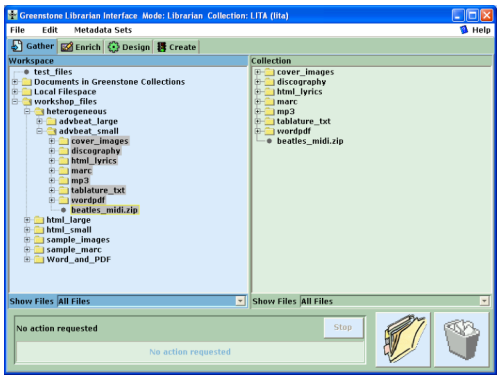


- Audio:
  - MP3 files
  - Midi files zipped up in a single .zip file
- Discography: HTML files (including many images)
- Images: JPEGs of album covers
- Lyrics: HTML files
- MARC: records containing relevant bibliog records
- Supplementary: material in PDF and Word files
- Tablature: guitar tablature in .txt files

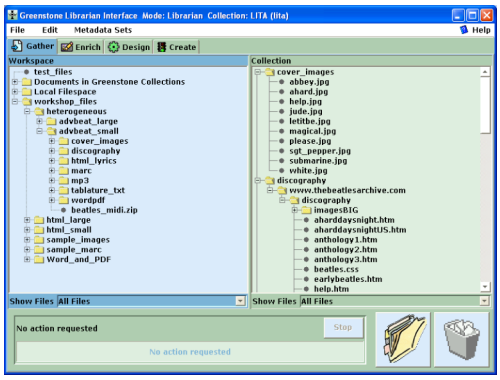
Building the Beatles collection



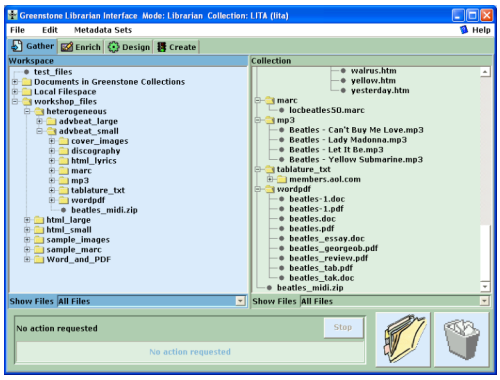
Gather: Gather the files together



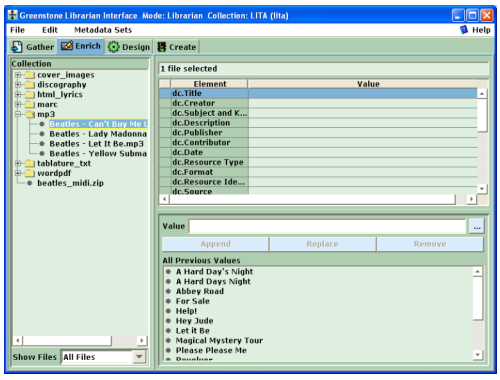
The ragtag menagerie of documents



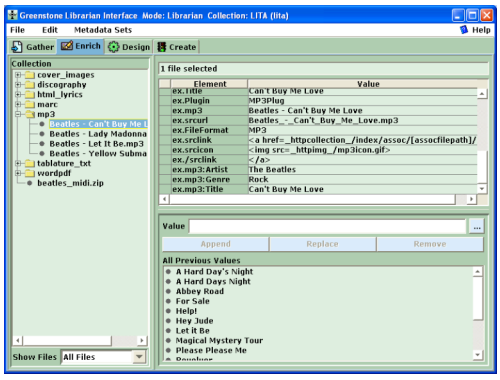
The ragtag menagerie of documents



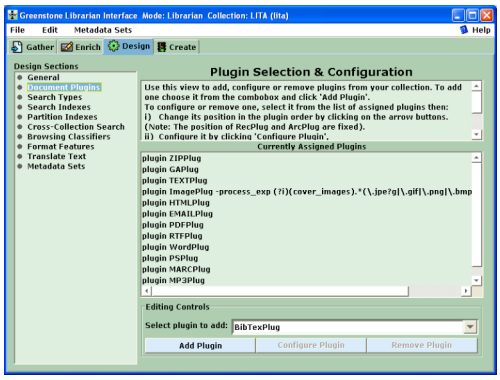
Enrich: Add metadata (if you like)



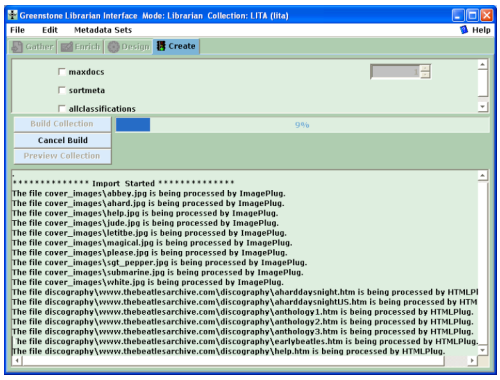
Enrich: Extracted metadata from MP3Plug



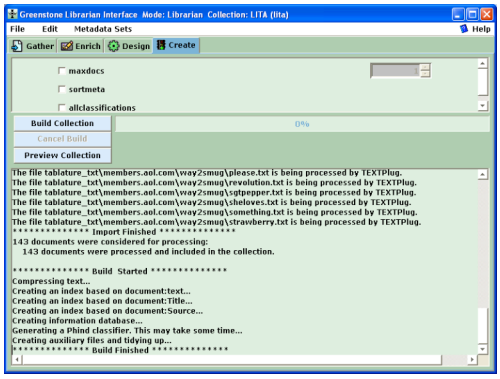
Design: Here are the plugins (and much more)



Create: Building the collection

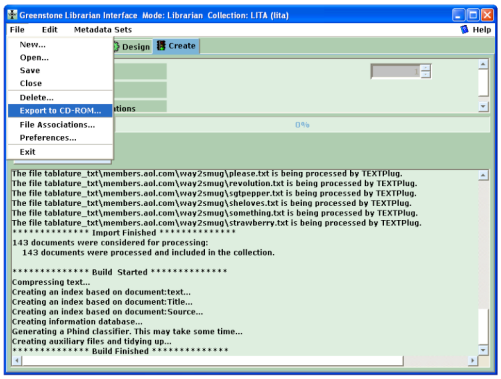


Create: It's built – preview it?

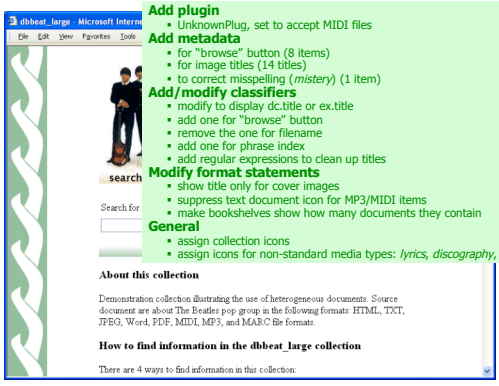


Previewing the collection

Export the collection to CD-ROM?



A (slightly) enhanced collection



Add plugin
- UnknownPlug, set to accept MIDI files

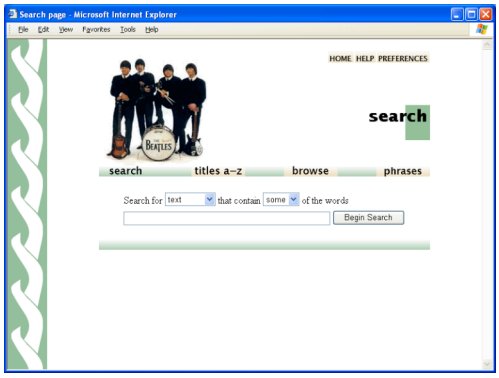
Add metadata
- for "browse" button (8 items)
- for image titles (14 titles)
- to correct misspelling (*mistery*) (1 item)

Add/modify classifiers
- modify to display dc.title or ex.title
- add one for "browse" button
- remove the one for filename
- add one for phrase index
- add regular expressions to clean up titles

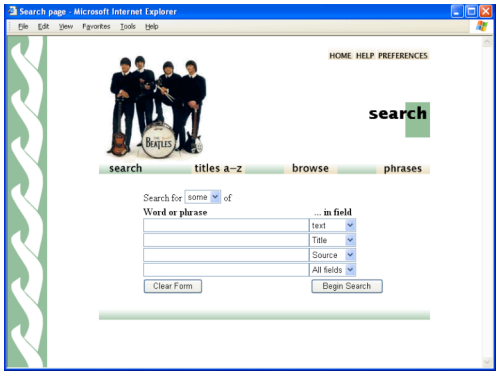
Modify format statements
- show title only for cover images
- suppress text document icon for MP3/MIDI items
- make bookshelves show how many documents they contain

General
- assign collection icons
- assign icons for non-standard media types: *lyrics*, *discography*, etc

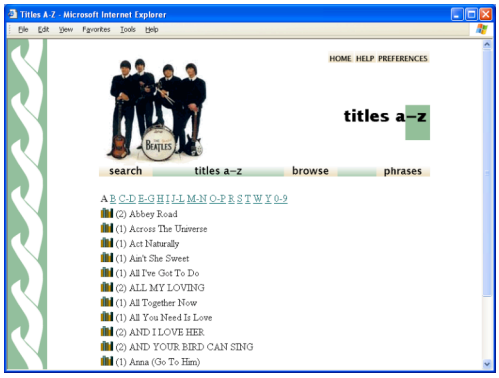
Full-text search



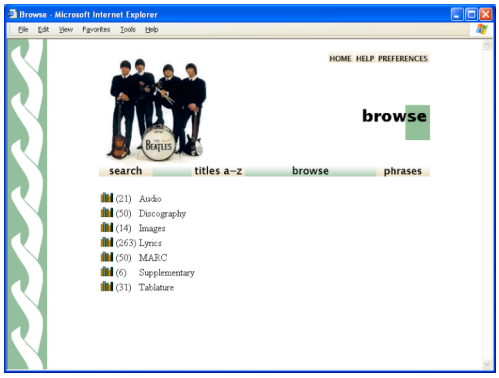
Form-based search



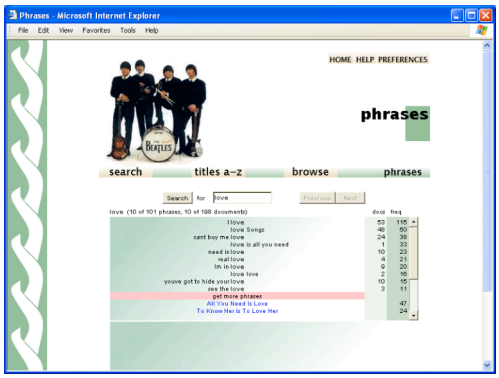
Browsing titles



Browsing document types



Hierarchical phrase browser



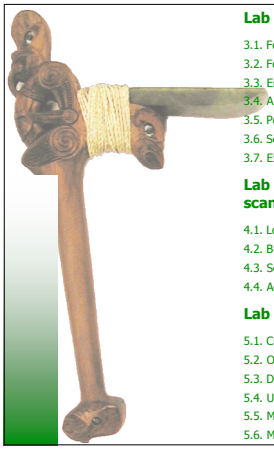
The workshop

Lab 1: Installing, browsing, building

1.1. Working with a pre-packaged collection (UNAIDS)
1.2. Installing Greenstone
1.3. Updating a Greenstone installation
1.4. Building a small collection of HTML files
1.5. A large collection of HTML files—Tudor
1.6. A collection of Word and PDF files—Part A
1.7. Enhanced Word document handling
1.8. Downloading files from the web

Lab 2: Adding metadata—and using it

2.1. A collection of Word and PDF files—Part B
2.2. A simple image collection
2.3. Enhanced collection of HTML files—Tudor
2.4. A bibliographic collection—Part A
2.5. CDS/ISIS collection
2.6. Editing metadata sets



Lab 3: Advanced collection configuration

3.1. Formatting the Word and PDF collection
3.2. Formatting the HTML collection—Tudor
3.3. Enhanced PDF handling
3.4. A bibliographic collection—Part B
3.5. Pointing to documents on the web
3.6. Section tagging for HTML documents
3.7. Exporting a collection to CD-ROM/DVD

Lab 4: Two examples: multimedia and scanned images

4.1. Looking at a multimedia collection
4.2. Building a multimedia collection
4.3. Scanned image collection
4.4. Advanced scanned image collection

Lab 5: Interoperability

5.1. Customization: macro files and stylesheets
5.2. Open Archives Initiative (OAI) collection
5.3. Downloading over OAI
5.4. Use METS as Greenstone's Internal Representation
5.5. Moving a collection from DSpace to Greenstone
5.6. Moving a collection from Greenstone to DSpace

News flashes

Getting stuff in

- New and flexible facilities for "paged images"
- Audio or video? Plugins for MP3, Ogg Vorbis, RealMedia
- Metadata? Plugins for MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI
- Keep metadata as a single file or "explode" it into individual entries
- CDS/ISIS plugin (plus new, great documentation)
- Import your collection as METS, or from DSpace
- Plugin for the OpenDocument format

- LOM? Use the LOM plugin
- CONTENTdm? Use the CONTENTdm plugin
- Using the Download panel to interactively download from OAI, or Z39.50, or SRW
- Greenstone converts XML-based formats (OAI, METS) on the fly using XSLT (e.g. for Fedora)

Working on stuff

- Work on a remote collection with the Librarian applet
- There's also a Java Remote Building program for this
- Cross-collection searching …

Getting stuff out

- Export your collection as METS, or for DSpace
- Serve it over OAI, or over Z39.50

Support for Institutional Repositories*

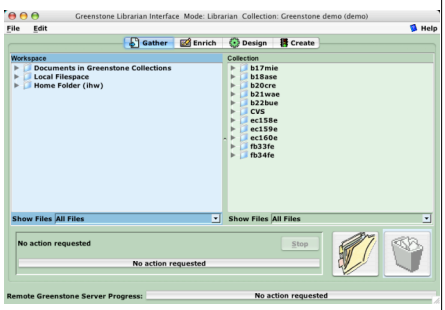
- Add documents to collections incrementally
- (Can use Lucene as its search engine instead of the old MG)
- DSpace-like web submission interface (with flexible workflows)

Future-proofing

- Greenstone3?
  It's coming, but everything is compatible
  —and your readers will never notice the difference

News flash Applet version of GLI

- **Collection on remote Greenstone server**
- **Edit it locally**
  - **gather**
  - **enrich**
  - **design**
  - **create**
- **Authentication**
  - **uses admin facility**
- **Locking**

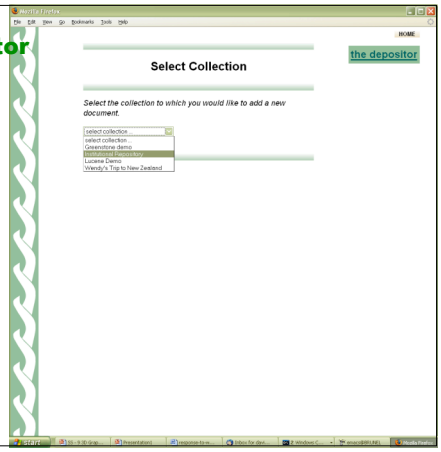


News flash CONTENTdm lookalike

http://puka.cs.waikato.ac.nz/cgi-bin/library?a=p&p=home&c=contentdm

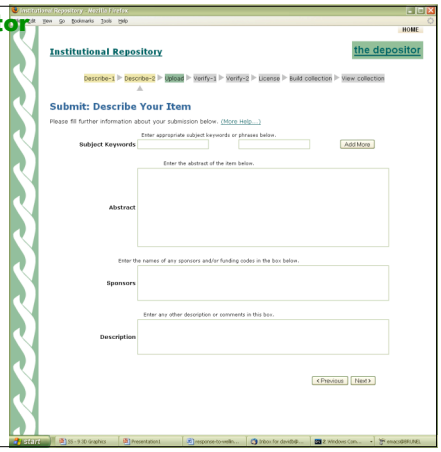
DSpace

News flash
The Depositor



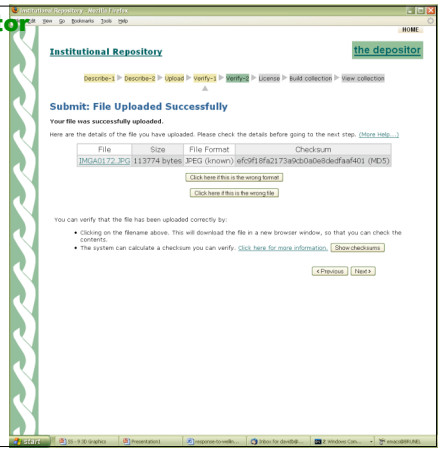
The Depositor

The Depositor

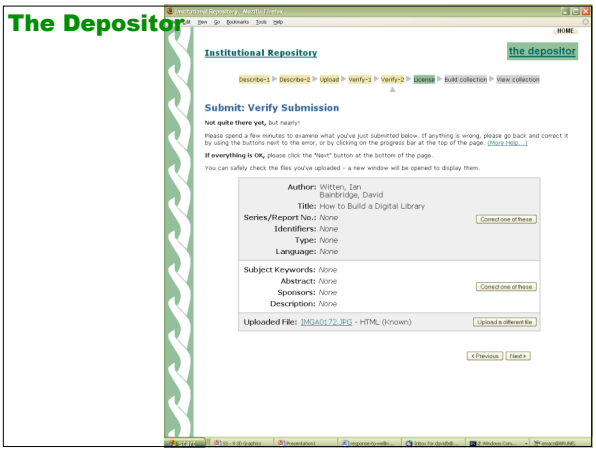


The Depositor

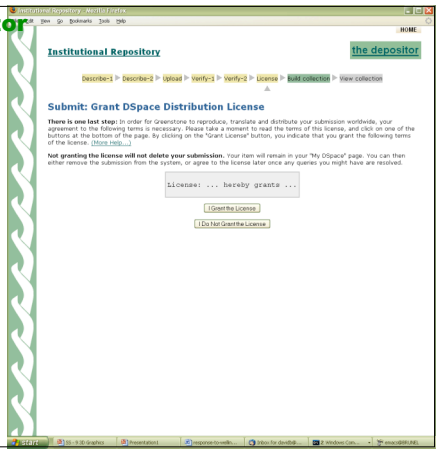
The Depositor



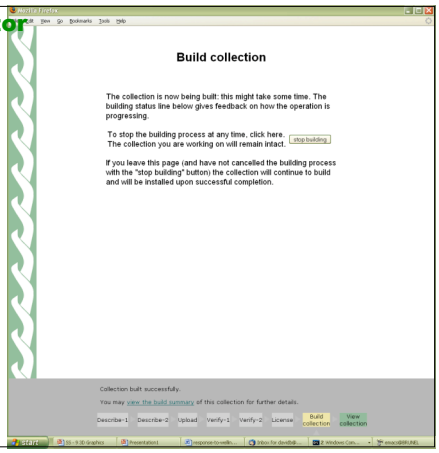
The Depositor



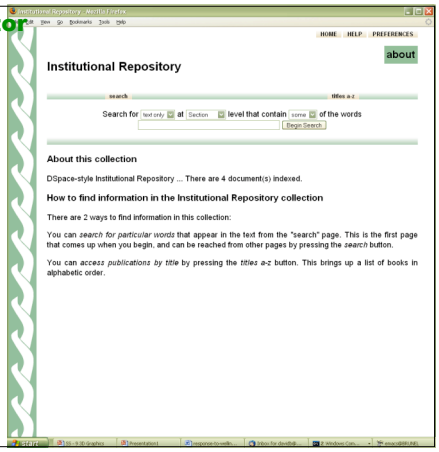
The Depositor



The Depositor



The Depositor



Agenda

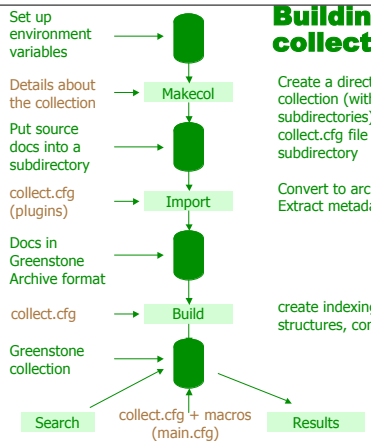
Overview
- What does Greenstone do?
- Greenstone facts; standards
- Reader's Interface: examples of collections

Librarian interface
- Build a collection in 30 sec (Hobbits)
- Build a multimedia collection (Beatles)
- Adding and using metadata
- Browsing classifiers, search indexes
- Building a collection manually (for masochists only)

Advanced stuff
- Under the hood: collection configuration file
- Customizing with macros
- Personalizing your home page
- Different interface languages
- Examples of what others have done

Reaching out
- Serving and acquiring OAI
- DSpace and METS
- Greenstone3

Building a collection

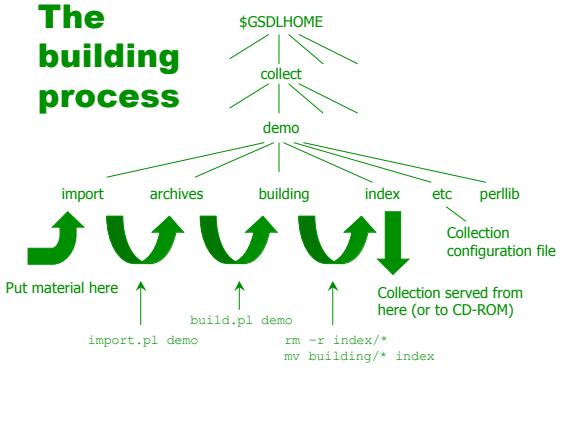


Create a directory for the collection (with subdirectories), put collect.cfg file in "etc" subdirectory

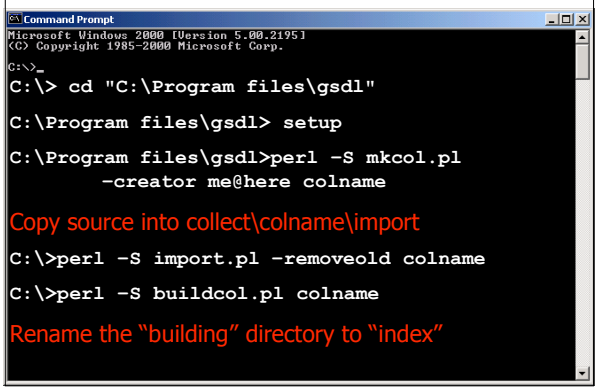
Convert to archive format
Extract metadata

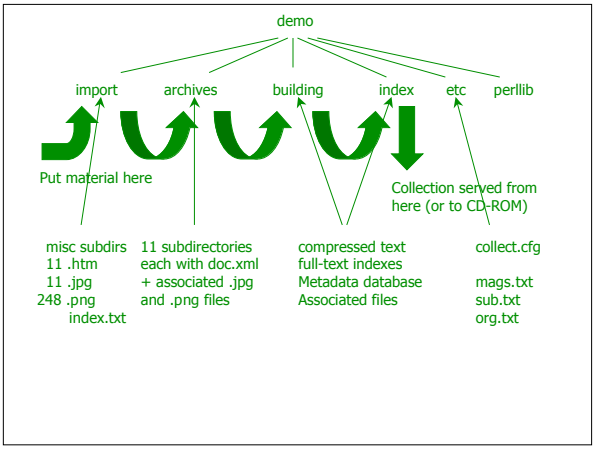
create indexing & browsing structures, compress …

The building process

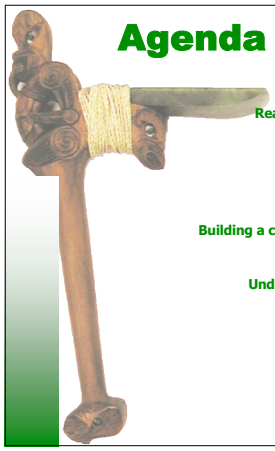


The building process





Agenda



Overview
- What does Greenstone do?
- Greenstone facts; standards
- Reader's Interface: examples of collections

Librarian interface
- Build a collection in 30 sec (Hobbits)
- Build a multimedia collection (Beatles)
- Adding and using metadata
- Browsing classifiers, search indexes
- Building a collection manually (for masochists only)

Advanced stuff
- Under the hood: collection configuration file
- Customizing with macros
- Personalizing your home page
- Different interface languages
- Examples of what others have done

Reaching out
- Serving and acquiring OAI
- DSpace and METS
- Greenstone3

Under the hood: Collection configuration file

- ❖name, icon, etc
- ❖description
- ❖email of creator
- ❖search indexes
- ❖plugins
- ❖classifiers

how to format

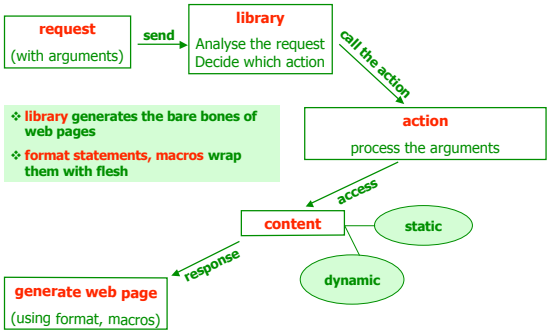
- ❖documents
- ❖query results
- ❖classifiers

```
creator         sjboddie@cs.waikato.ac.nz
maintainer      sjboddie@cs.waikato.ac.nz
public          true
beta            true

indexes         section:text section:Title document:text
defaultindex    section:text

plugin          GAPlug
plugin          ArcPlug
plugin          RecPlug

classify        Hierarchy -hfile sub.txt -metadata Subject -sort Title
classify        HDLList -metadata Title
classify        Hierarchy -hfile org.txt -metadata Organization -sort Title
classify        List -metadata Howto

format          SearchVList "<td valign=top>[link][icon][/link]</td>
                <td>{If}{[parent(All': '):Title],[parent(All': '):Title]: }
                [link][Title][/link]</td>"
format          CL4VList    "<br>[link][Howto][/link]"
format          DocumentImages true
format          DocumentText "<h3>[Title]</h3>\\n\\n<p>[Text]"

collectionmeta collectionname  "greenstone demo"
collectionmeta collectionextra "This is a demonstration collection for the
                Greenstone digital library software.\nIt contains a small
                subset (11 books) of the Humanity Development Library"
collectionmeta iconcollectionsmall "/gsdl/collect/demo/images/demosm.gif"
collectionmeta iconcollection      "/gsdl/collect/demo/images/demo.gif"
collectionmeta .section:Title      "section titles"
collectionmeta .document:text      "entire books"
collectionmeta .section:text       "chapters"
```

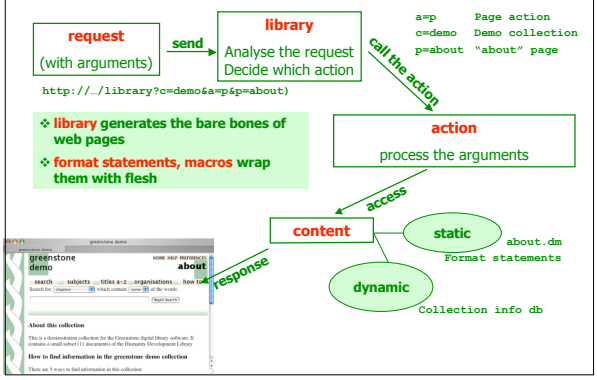
Alter configuration

❖ Add full-text index of titles	indexes document:Title
❖ ... or authors	indexes document:Creator
❖ Add alphabetic author browser	classify AZList –metadata Creator
❖ Include Word documents	plugin WordPlug
❖ Include PDF documents	plugin PDFPlug
❖ Separate index for each language	languages en fr es
❖ Extract acronyms and add list	plugin PDFPlug –extract_acronyms
❖ Import OAI metadata	plugin OAIPlug
❖ Extract phrase hierarchy and add browser	classify Phind
❖ Alter the format of any of the above	format ...
❖ Restrict collection's interface langs	format PreferenceLangs en|fr|es
❖ Change default interface language	cgiarg shortname=1 argdefault =fr

Generating web pages



Generating web pages



Customizing with macros

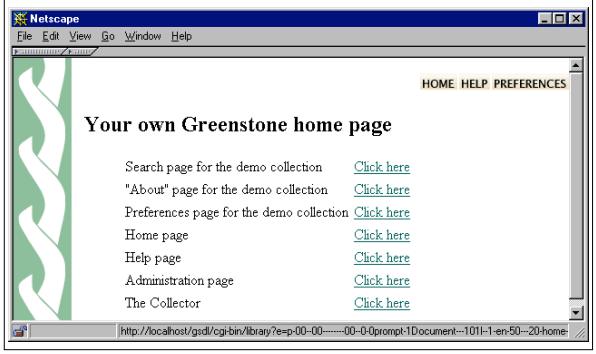
- let you customize presentation
- present pages in different languages
- print variables into the page text
  (e.g. number of search hits)

❖ Macro files
- stored in gsdl/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 gsdl/collect/mycol/macros/extra.dm
- Or include argument [c=collectionname] for each macro

Personalizing your home page

C:\Program Files\gsdl\etc\main.cfg change *home.dm* to *yourhome.dm*



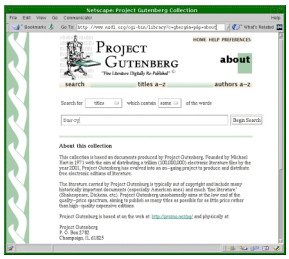
yourhome.dm

```
package home
_content_ {
<h2>Your own Greenstone home page</h2>
<ul>
<table>
<tr valign=top><td>Search page for the demo collection<br></td>
    <td><a href="_httpquery_&c=demo">Click here</a></td></tr>

<tr><td>"About" page for the demo collection</td>
    <td><a href="_httppageabout_&c=demo">Click here</a></td></tr>

<tr><td>Preferences page for the demo collection</td>
    <td><a href="_httppagepref_&c=demo">Click here</a></td></tr>

<tr><td>Home page</td>
    <td><a href="_httppagehome_">Click here</a></td></tr>

<tr><td>Help page</td>
    <td><a href="_httppagehelp_">Click here</a></td></tr>

<tr><td>Administration page</td>
    <td><a href="_httppagestatus_">Click here</a></td></tr>

<tr><td>The Collector</td>
    <td><a href="_httppagecollector_">Click here</a></td></tr>

</table>
</ul>
}
```

Macros used in home.dm

*_httppagehome_*	name of the home page
*_httppagehelp_*	… the help page
*_httppagestatus_*	… the administration page
*_httppagecollector_*	… the Collector page
httpquery&c=demo	search page for the demo collection
httppageabout&c=demo	about page for the demo collection
httppagepref&c=demo	preferences page for the demo collection
_content_{ … }	defines a macro called *_content_* contains HTML, but '{', '}', '\', and '_' must be escaped with a backslash
_header_{ … }	HTML page header (contains squirly bar)
_footer_{ … }	HTML page footer
main.cfg	contains list of macros, replace *home.dm* by *yourhome.dm* and put it in the *macros* directory

about.dm



❖ can contain conditional statements
- _If_ …

❖ can take arguments
- a=p, l=de etc.

```
package about

##################################
# about page content
##################################

_pagetitle_{_collectionname_}

_content_{
<center>
_navigationbar_
</center>
_query:queryform_
<p>_iconblankbar_
<p>_textabout_
_textsubcollections_
<h3>_help:textsimplehelpheading_</h3>
_help:simplehelp_
}

_textabout_ {
<h3>_textabcol_</h3>
_Global:collectionextra_
}
```

Macro names begin and end with underscore

Content defined in curly brackets

Content is plain text, HTML, other macros, Javascript, applet links

Customization hints

- To change the look and feel of Greenstone
  - edit the *base* and *style* packages
- To change the query page
  - edit *query.dm*
- To change the Preferences page
  - edit *pref.dm*

Different interface languages

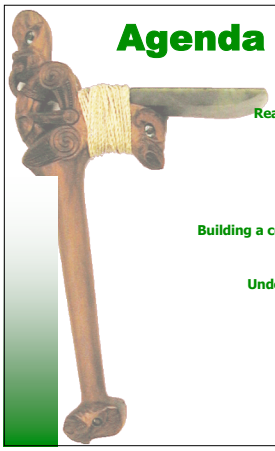
english.dm

```
#----------------------------------------------------------
# text macros
#----------------------------------------------------------

_textimageabout_ {About page}
_textimagehelp_ {Help page}
_textimagepref_ {Preferences page}
_textimagegreenstone_ {Greenstone Digital Library Software}

_textimagesearch_ {Search for specific terms}
_textimageCreator_ {Browse alphabetical list of authors}
_textimageTitle_ {Browse alphabetical list of titles}
_textmatches_ {Matches }
_textbeginsearch_ {Begin Search}

_texticonsmalltext_ {View this section of the text}
_texticondetach_ {Open this page in a new window}
_texticonexpandtoc_ {Expand table of contents}
_texticonexpandtext_ {Display all text}
_texticonhighlight_ {Highlight search terms}

_textlanguage_ {Interface language: }
_textsimplemode_ {simple query mode}
_textadvancedmode_ {advanced query mode (allows boolean searching u
```

chinese.dm

```
#----------------------------------------------------------
# text macros
#----------------------------------------------------------

_textimageabout_ [l=zh] {主页}
_textimagehelp_ [l=zh] {帮助页}
_textimagepref_ [l=zh] {选项页}
_textimagegreenstone_ [l=zh] {绿石数字化图书馆软件}

_textimagesearch_ [l=zh] {特定范围检索}
_textimageCreator_ [l=zh] {按作者名字的阿拉伯字母顺序浏览 }
_textimageTitle_ [l=zh] {按题名的阿拉伯字母顺序浏览}
_textmatches_ [l=zh] {匹配 }
_textbeginsearch_ [l=zh] {开始查询}

_texticonsmalltext_ [l=zh] {  浏览该部份文本}
_texticondetach_ [l=zh] {  打开一个新的窗口显示本页}
_texticonexpandtoc_ [l=zh] {  扩展书录表}
_texticonexpandtext_ [l=zh] {  显示所有文本}
_texticonhighlight_ [l=zh] {  查询项以粗体显示}

_textlanguage_ [l=zh] {界面语言：}
_textsimplemode_ [l=zh] {简单查询方式}
_textadvancedmode_ [l=zh] {高级查询方式(可以使用 !、&、| 和括号来布尔查询)}
```

Agenda



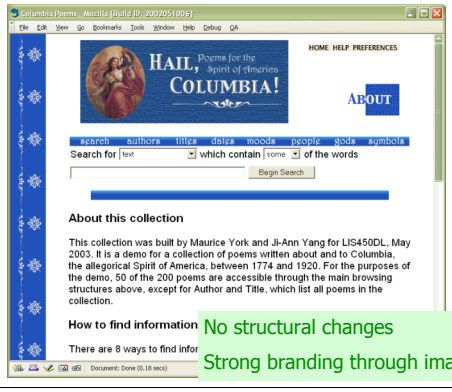
Overview
- What does Greenstone do?
- Greenstone facts; standards
- Reader's Interface: examples of collections

Librarian interface
- Build a collection in 30 sec (Hobbits)
- Build a multimedia collection (Beatles)
- Adding and using metadata
- Browsing classifiers, search indexes
- Building a collection manually (for masochists only)

Advanced stuff
- Under the hood: collection configuration file
- Customizing with macros
- Personalizing your home page
- Different interface languages
- Examples of what others have done

Reaching out
- Serving and acquiring OAI
- DSpace and METS
- Greenstone3

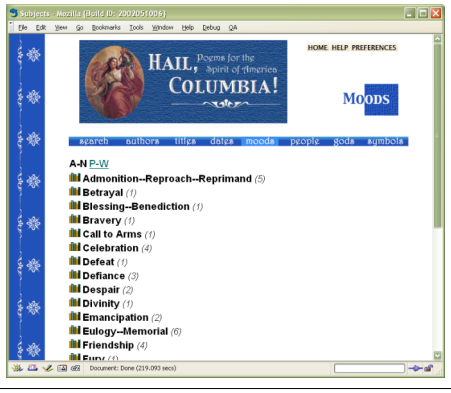
Illinois student project



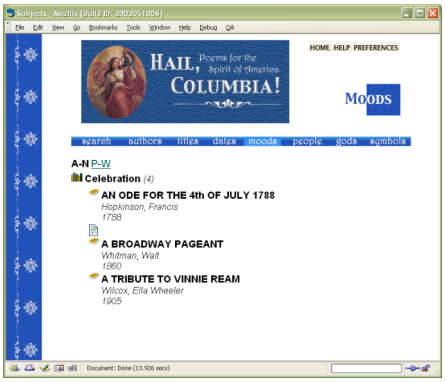
No structural changes

Strong branding through images

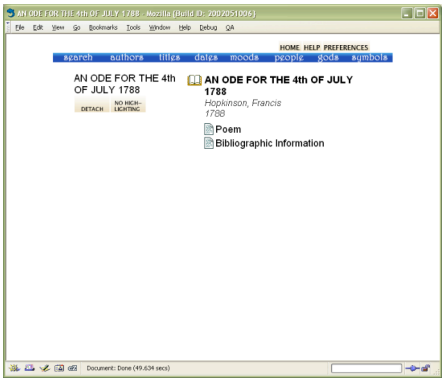
Illinois student project



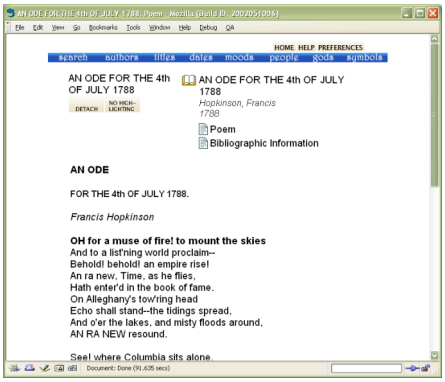
Illinois student project



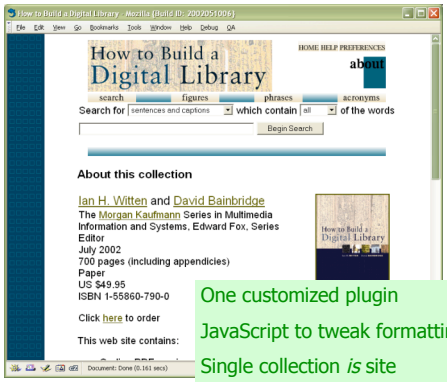
Illinois student project



Illinois student project

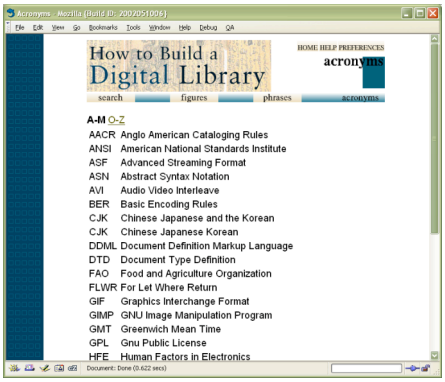


"How to build" web site

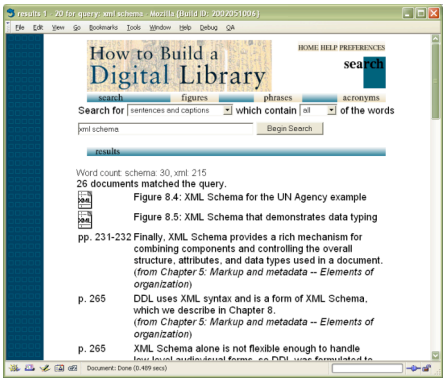


- One customized plugin
- JavaScript to tweak formatting
- Single collection *is* site

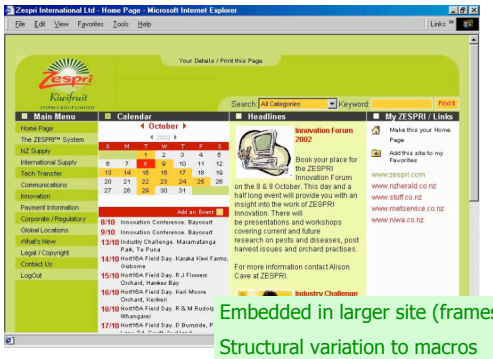
"How to build" web site



"How to build" web site



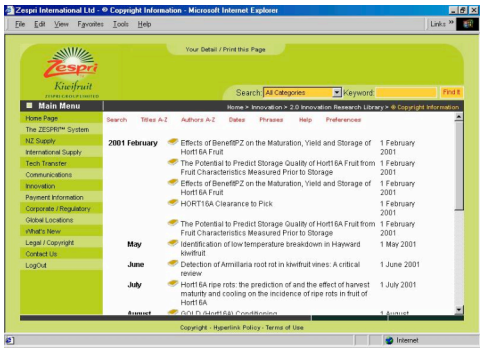
Zespri innovation library



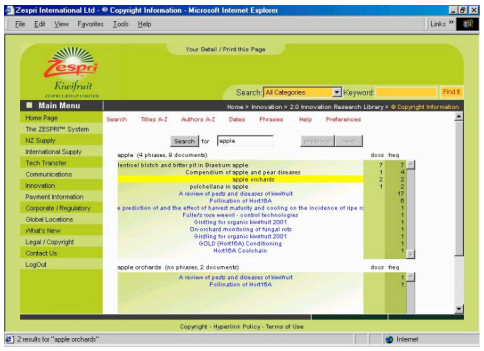
Embedded in larger site (frames)

Structural variation to macros

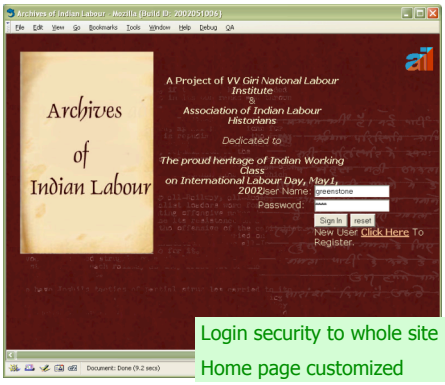
Zespri innovation library



Zespri innovation library



Archives of Indian labour

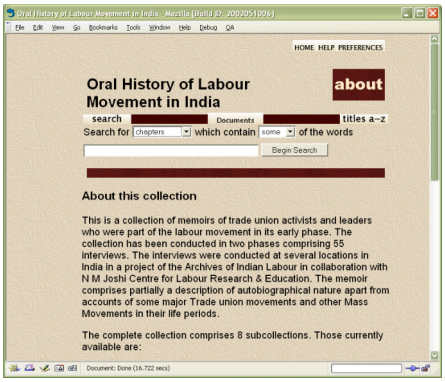


Login security to whole site

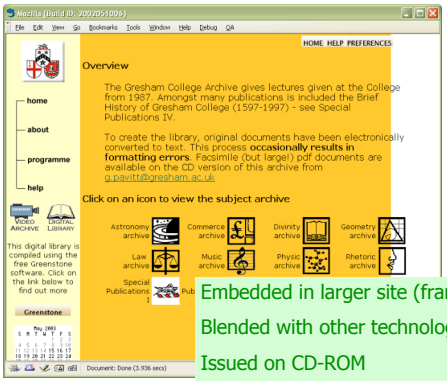
Home page customized

Archives of Indian labour

Archives of Indian labour



Gresham College archive

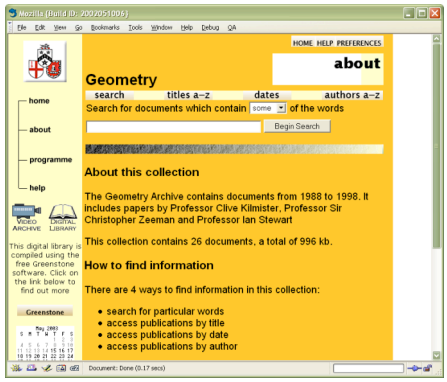


Embedded in larger site (frames)

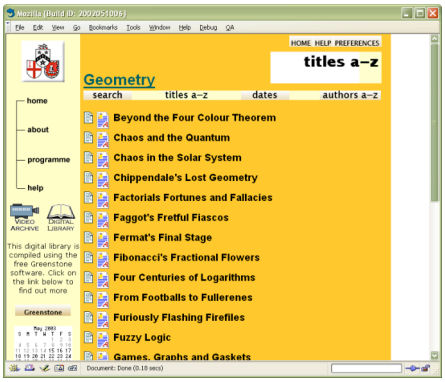
Blended with other technologies

Issued on CD-ROM

Gresham College archive



Gresham College archive



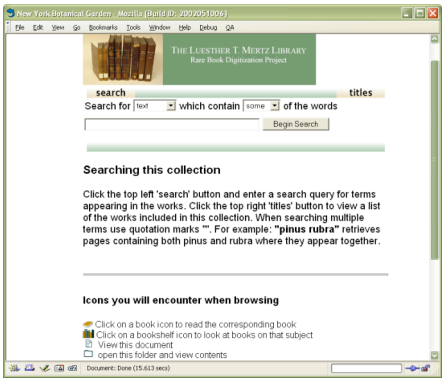
The NY Botanical Gardens

Single collection *is* site

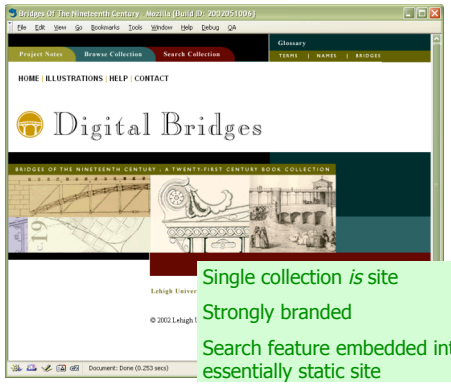
Supplementary information provided through static pages

Redesign of *about collection* page

The NY Botanical Gardens



Lehigh University



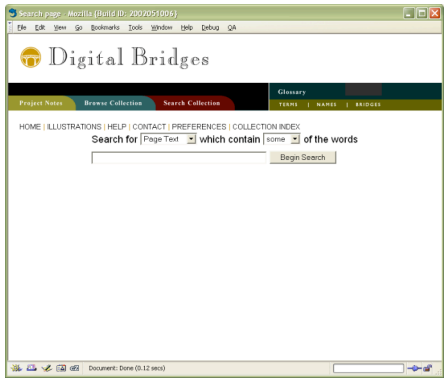
Single collection *is* site

Strongly branded

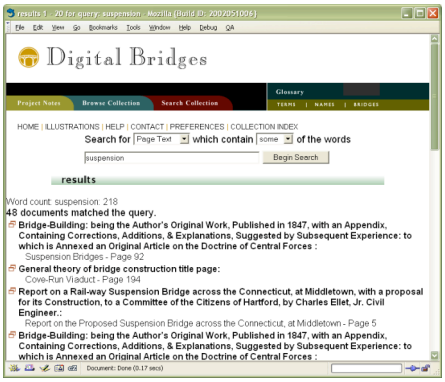
Search feature embedded into essentially static site

Lehigh University

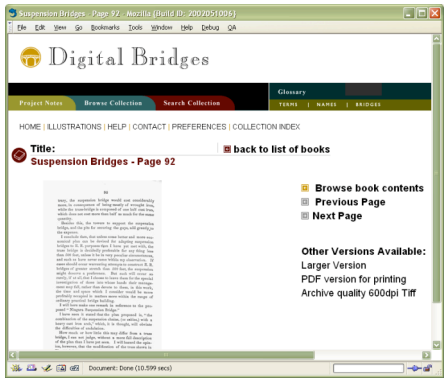
Lehigh University



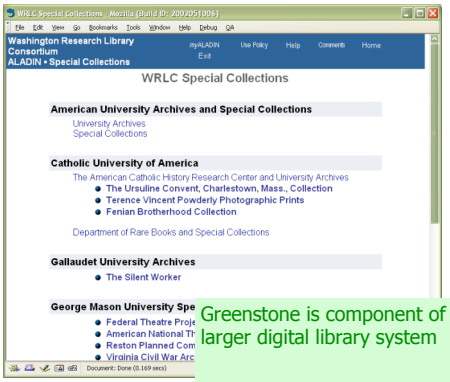
Lehigh University



Lehigh University



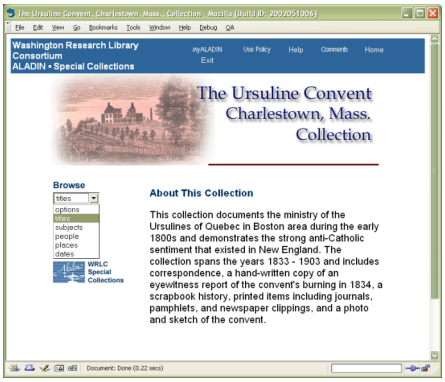
Aladin digital library system



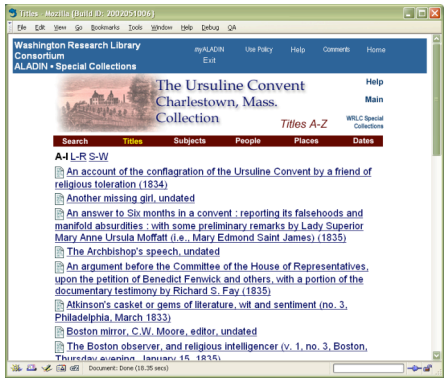
Greenstone is component of larger digital library system

Aladin digital library system

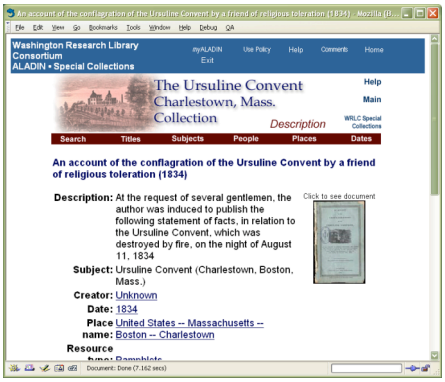
Aladin digital library system



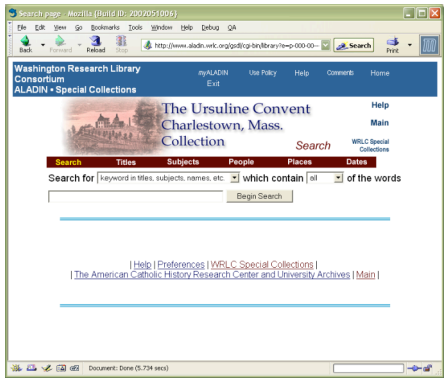
Aladin digital library system



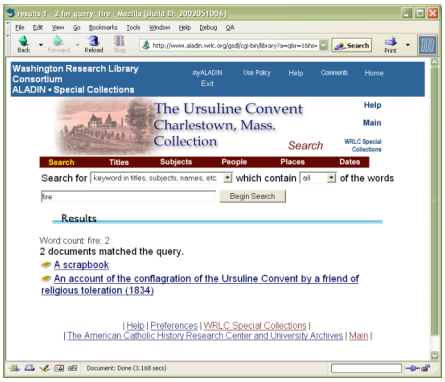
Aladin digital library system



Aladin digital library system



Aladin digital library system



Agenda

Overview
What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections

Librarian interface
Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)

Advanced stuff
Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done

Reaching out
Serving and acquiring OAI
DSpace and METS
Greenstone3

The Greenstone documentation wiki

- go to greenstone.org

The Greenstone OAI server

- Runs as a CGI program called *oaiserver*
  - Greenstone installation

```
http://.../cgi-bin/gsdl
```

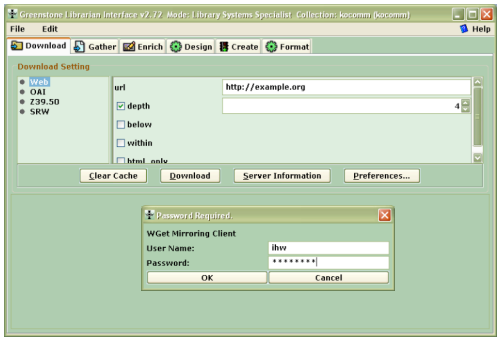
  - OAI server

```
http://.../cgi-bin/oaiserver?verb=Identify
http://.../cgi-bin/oaiserver?verb=ListSets
http://.../cgi-bin/oaiserver?verb=ListIdentifiers&set=xxx
http://.../cgi-bin/oaiserver?verb=ListIdentifiers&set=xxx&metadataPrefix=oai_dc
http://.../cgi-bin/oaiserver?verb=ListRecords&set=xxx&metadataPrefix=oai_dc
http://.../cgi-bin/oaiserver?verb=GetRecord&identifier=xxx&metadataPrefix=oai_dc
```

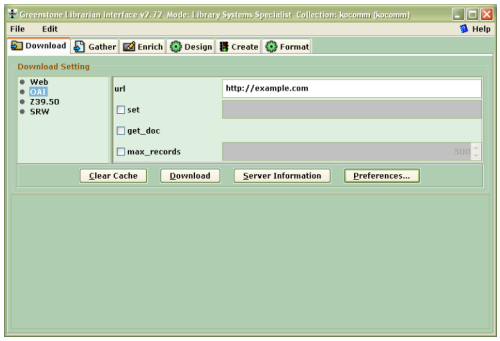
OAI 1.1
OAI 2.0

- Requires a full webserver (not "local library" version)
- Configuration file: `etc/oai.cfg` in the Greenstone filespace
  - repository name and version (OAI 1.1 or 2.0)
  - collections to be made accessible to OAI clients
  - metadata mapping file into DC (server only supports DC)

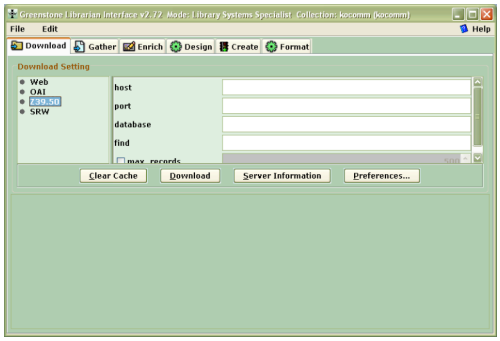
Acquiring web documents



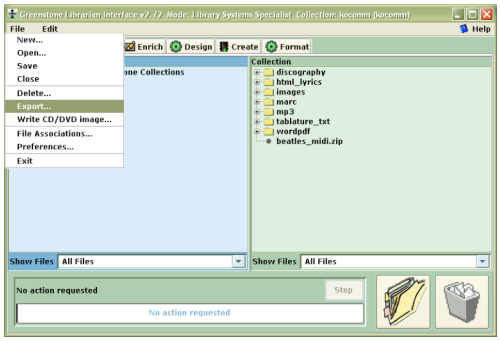
Acquiring OAI metadata + docs



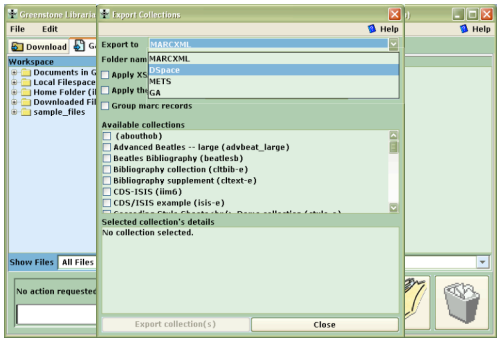
Acquiring Z39.50/SRW metadata



Exporting a collection



Exporting as DSpace, METS ...

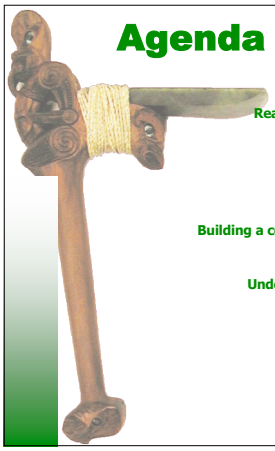


MODS: XSLT mapping

MODS mapping into Greenstone using XSLT
(specify as argument to OAIPlug):

```
...
<xsl:template match="mods:topic">
  <dc:Subject><xsl:value-of select="." /></dc:Subject>
</xsl:template>

<xsl:template match="mods:titleInfo">
  <dc:Title>
    <xsl:value-of select="mods:title" />
    <xsl:if test="mods:subTitle">
      <xsl:text>: </xsl:text>
      <xsl:value-of select="mods:subTitle" />
    </xsl:if>
  </dc:Title>
</xsl:template>
...
```

Agenda



Overview
What does Greenstone do?
Greenstone facts; standards
Reader's Interface: examples of collections

Librarian interface
Build a collection in 30 sec (Hobbits)
Build a multimedia collection (Beatles)
Adding and using metadata
Browsing classifiers, search indexes
Building a collection manually (for masochists only)

Advanced stuff
Under the hood: collection configuration file
Customizing with macros
Personalizing your home page
Different interface languages
Examples of what others have done

Reaching out
Serving and acquiring OAI
DSpace and METS
Greenstone3

Greenstone 3

Greenstone 2 was designed about 7 years ago
Since then, new software and library standards have emerged ...

- **Easier to modify interface: generates data (XML) rather than presentation (HTML)**
  - XML can be transformed using XSLT, rather than HTML via macro files
- **More dynamic: more can be done at runtime**
  - e.g. sort search results on the fly, generate new classifiers at runtime
- **METS foundation: multiple hierarchies possible**
  - e.g. chapter-section-subsection hierarchy can coexist with page header-contents-footer hierarchy
- **Distributed modules, using SOAP**
  - interface can be on different computer to collection, can talk to several different Greenstone sites
- **Java basis: platform independent**
  - Greenstone2 operates on different platforms, but at a high cost to implementors
- **Better software technology**
  - easier for people (e.g. our students!) to modify the software to make it do radically different things