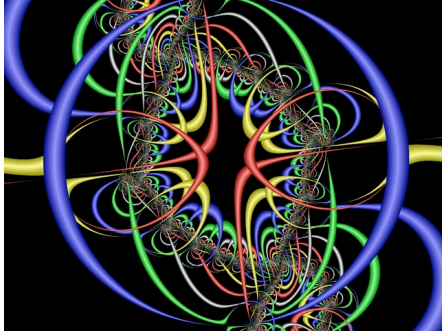


Advanced collection configuration



Course material prepared by

Greenstone Digital Library Project
University of Waikato, New Zealand



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Format Statements: Components

- ❖ HTML
- ❖ [Text] *displays document text*
- ❖ [Title], [Howto] ... *displays metadata*
- ❖ [link] ... [/link] *links to document*
- ❖ [srclink] ... [/srclink] *links to original file*
- ❖ [icon], [srcicon] *page/book/bookshelf source icons*
- ❖ *If and Or statements conditional processing*

Format Statements

```
{If}{test, if true, if false}
```

❖ Test can be:

[metadata]	exists
[metadata] eq 'value'	equals
[metadata] ne 'value'	not equals

❖ Examples

```
{If}{[ex.Source], (<i>[ex.Source]</i>)}
{If}{[numleafdocs], [ex.Title], [dc.Creator]}
{If}{[ex.FileFormat] eq PDF, [srclink] PDF
document [/srclink]}
```

Format Statements

```
{Or}{ [metadata1] , [metadata2] , ...
```

❖ Chooses the first metadata that exists

❖ Last item can be plain text

❖ Examples

```
{Or}{ [dls.Title] , [dc.Title] , [ex.Title] , Untitled }
```

```
{Or}{ [ex.thumbicon] , [ex.srcicon] }
```

HTML page

```
<html>  
  <head>  
    <title>The page title</title>  
  </head>  
  <body formatting-attributes>  
    Page content goes here  
  </body>  
</html>
```

HTML elements

`Bold, <i>italics</i>, <u>underline</u>`

`
` a line break

`<p>a paragraph</p>`

`<table><tr><td>cell content</td></tr></table>`

`link text`

``

Format statements

- ❖ Defaults provided to give reasonable results for most cases
- ❖ Interpreted at run-time – no need to rebuild
- ❖ More information: FAQ

[http://www.greenstone.org/cgi-bin/library?
a=p&p=faqcustomize#customizeformat](http://www.greenstone.org/cgi-bin/library?a=p&p=faqcustomize#customizeformat)

List classifier (Howto)

achieve gender equality
add fish, mollusc and crab culture or duck raising as livelihoods for coastal communities
... their households

```
<br>[link] [Howto] [/link]
```

```
format CL4VList "<br>[link] [Howto] [/link]"
```

```
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
```

apply accounting for small projects and groups
apply biomass energy technologies
apply ethnoveterinary medicine
... the latest sustainable energy technologies
apply these 101 small-scale technologies from the South for the South
assess environmental impacts of waste management
assess health care needs after floods and hurricanes
assess the environmental impacts of agriculture
... a documentation and a library system
build a pond for aquaculture
build climate responsive homes
build economically with local materials

http://www.nzdl.org/tast.cgi-bin/library?e=d-0hdl-00-prompt-1-1011-1-en-50--20-about--1-01-w-0ka=d&cl=CL

Formatting search results

msword and pdf demonstration

HOME HELP PREFERENCES

search

titles a-z

Search for documents that contain of the words

results

Word count: computer: 115
11 documents matched the query.

- biblio_for_dl_scientometrics.do (pdf03.pdf)
- Information Seeking, Retrieving, Reading, and Storing Behaviour of Library-Users. (word02.doc)
- 1992-96 Working Papers (word04.doc)
- 1997-00 Listing of Working Papers (word01.doc)
- Distributing Digital Libraries on the Web, CD-ROMs, and Intranets: Same information, same... (pdf02.pdf)
- Greenstone: A Comprehensive Open-Source Digital Library Software System

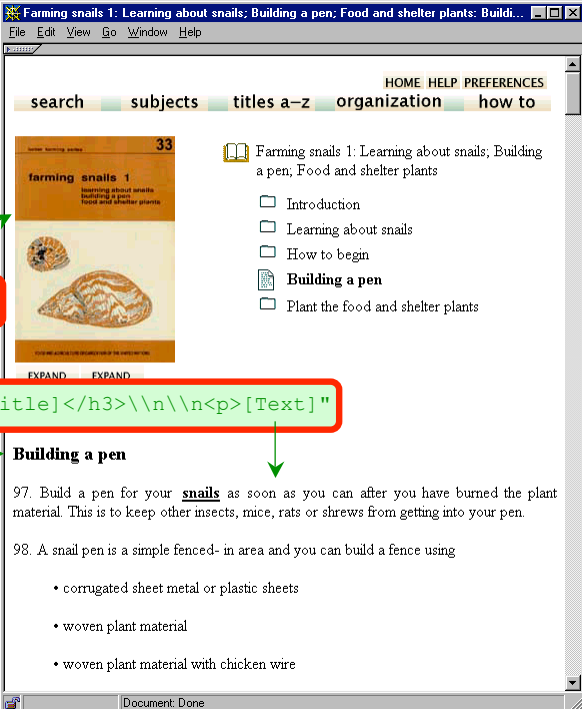
```
[srclink] [srcicon] [/srclink]
```

```
[link] [icon] [/link]
```

```
[ex.Title]
```

```
<br> (<i>[ex.Source]</i>)
```

Formatting document text



`format DocumentImages true`

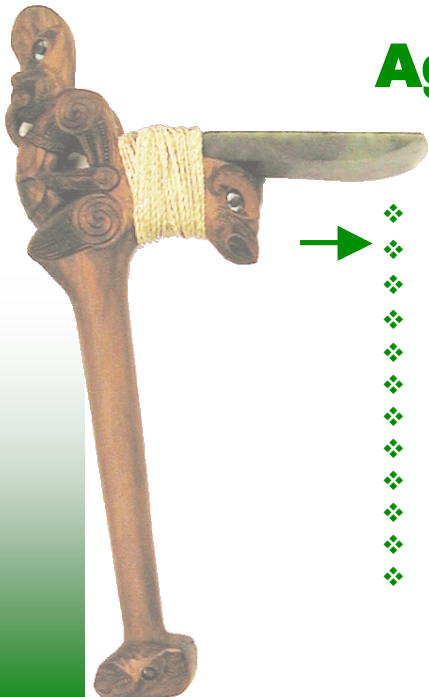
`format DocumentText "<h3>[Title]</h3>\n\n<p>[Text]"`

Building a pen

97. Build a pen for your snails as soon as you can after you have burned the plant material. This is to keep other insects, mice, rats or shrews from getting into your pen.

98. A snail pen is a simple fenced- in area and you can build a fence using

- corrugated sheet metal or plastic sheets
- woven plant material
- woven plant material with chicken wire



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Extended metadata

- ❖ Basic metadata:
 - [Title], [Source] (extracted)
 - [dc.Subject], [dls.Organization] (manually assigned)
- ❖ Extended metadata:
 - [parent:Title] (Title of parent section)
 - [parent(Top):dc.Title] (dc.Title of document, i.e. top section)
 - [sibling:Subject] (all Subjects of current section)
 - [child:Author] (Author of first child section)
 - [child(All):Author] (Author of all children)
- ❖ Formatting:
 - Sibling(All' and '):Subject
(between quotes can specify a separator)

Format statements: Extended metadata

Snail Farming

Subject: Agriculture

Subject: Farming

Subject: Cuisine

1. Introduction

1.1 Snails are good to eat

1.2 What is snail farming?

2. Getting started

2.1 How to prepare the pens

Format statements: Extended metadata

Snail Farming	→	Subject	Agriculture
Subject: Agriculture		sibling:Subject	Agriculture, Farming, Cuisine
Subject: Farming		sibling(last):Subject	Cuisine
Subject: Cuisine		Sibling(All' and `):Subject	Agriculture and Farming and Cuisine
1. Introduction		child:Title	Introduction, Getting Started
1.1 Snails are good to eat		child(2):Title	Getting started
1.2 What is snail farming?			
2. Getting started			
2.1 How to prepare the pens			

Format statements: Extended metadata

Snail Farming		parent:Title	Introduction
Subject: Agriculture		parent(Top):Title	Snail Farming
Subject: Farming		parent(All):Title	Snail Farming, Introduction
Subject: Cuisine		Parent(Top):Subject	Agriculture
1. Introduction		Parent(Top):sibling:Subject	Agriculture, Farming, Cuisine
1.1 Snails are good to eat		Parent(Top):sibling(All' and `):Title	Agriculture and Farming and Cuisine
1.2 What is snail farming?			
2. Getting started			
2.1 How to prepare the pens			

Format statements: Combining If and Or

- ❖ `{If}{test, if true, if false}`
- ❖ `{Or}{[metadata1],[metadata2]...}`
- ❖ `{Or}` can have a conditional as the last element:
- ❖ `{Or}{[Creator],[Editor],
 {If}{[FileFormat] eq "PDF",
 xxx,anonymous}}`

Format statements: Combining If and Or

- ❖ `{If}` can have another conditional at 'true' or 'false' position
- ❖ `{If}{[numleafdocs],[Title],
 [dc.Title]{If}{[Date],: [Date]}
 {If}{[Subject], ([Subject], unclassified)}}`

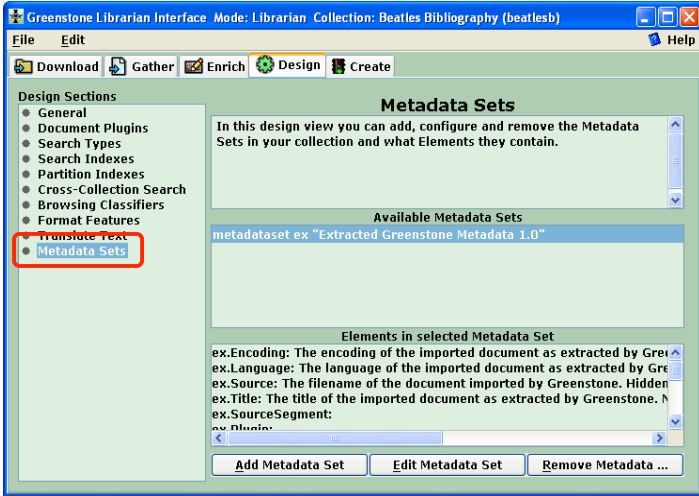
Snail farming: 26 Jun 1998 (Small Animal Farming)



Agenda

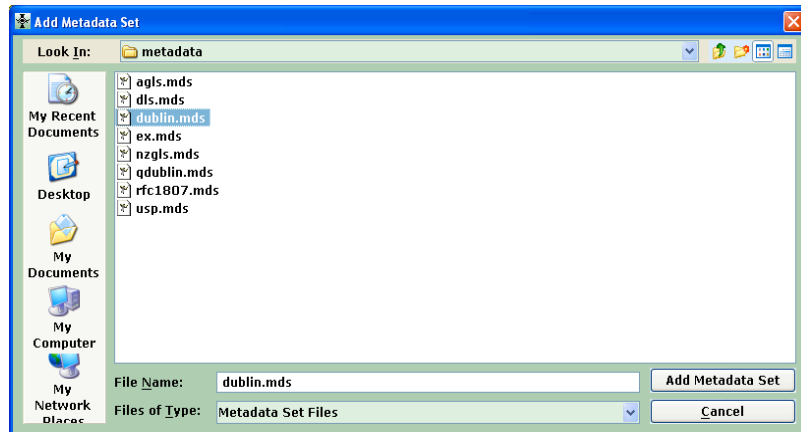
- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Metadata Sets

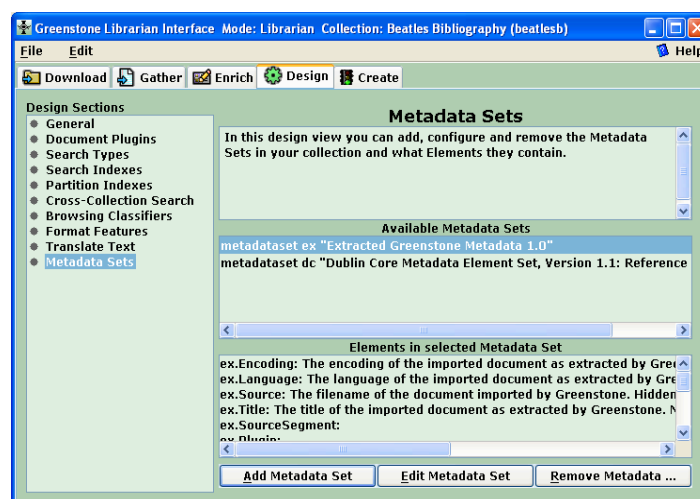


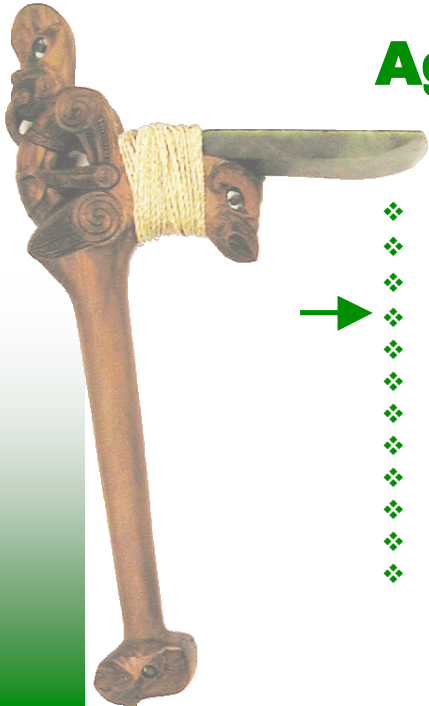
The screenshot shows the Greenstone Librarian Interface (GLI) in Librarian mode for the 'Beatles Bibliography (beatlesb)' collection. The 'Design' menu is active, and the 'Metadata Sets' option is selected in the left-hand 'Design Sections' pane. The main window displays the 'Metadata Sets' configuration screen, which includes instructions on how to add, configure, and remove metadata sets. It lists an available metadata set named 'metadataset ex "Extracted Greenstone Metadata 1.0"' and shows a list of elements in the selected set, such as 'ex.Encoding', 'ex.Language', 'ex.Source', 'ex.Title', and 'ex.SourceSegment'. Buttons for 'Add Metadata Set', 'Edit Metadata Set', and 'Remove Metadata ...' are visible at the bottom.

Add Metadata Set



Metadata Sets





Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ➔ ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

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"

```

Documentation and Help

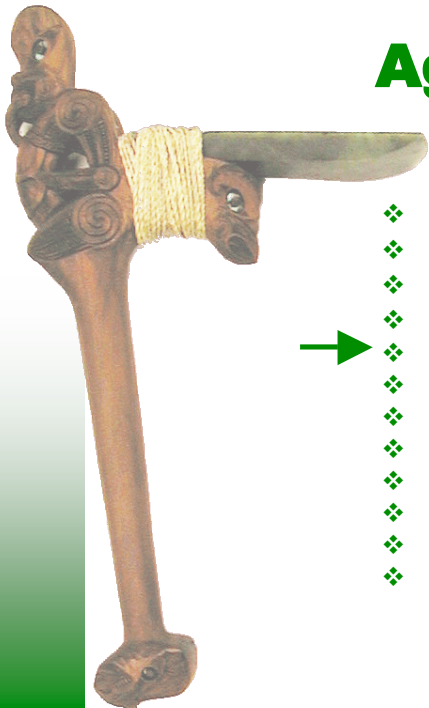
- ❖ *User's Guide* (user.pdf)

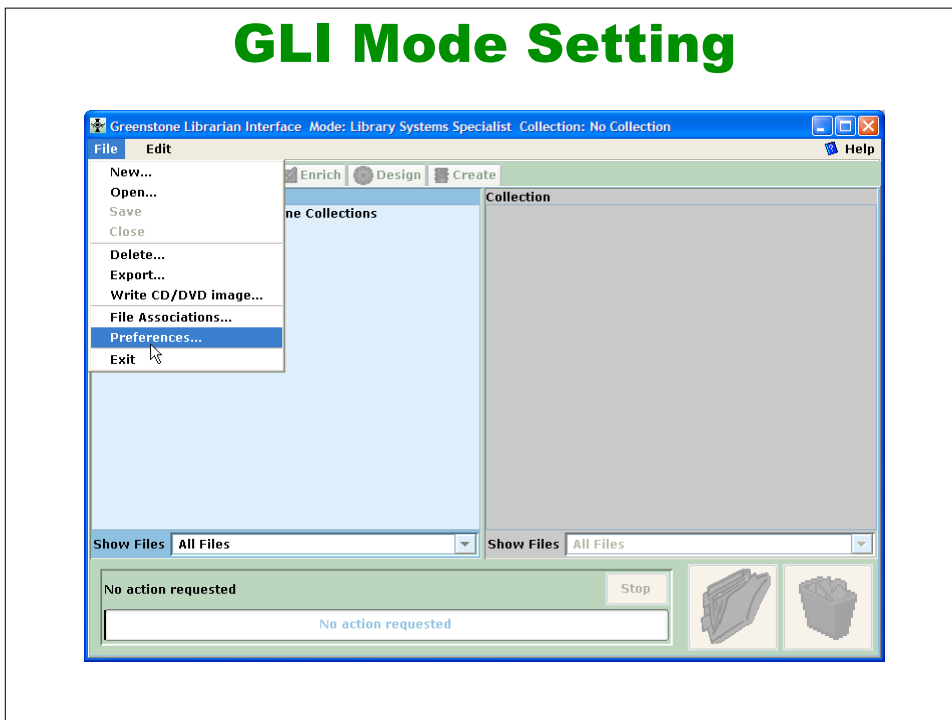
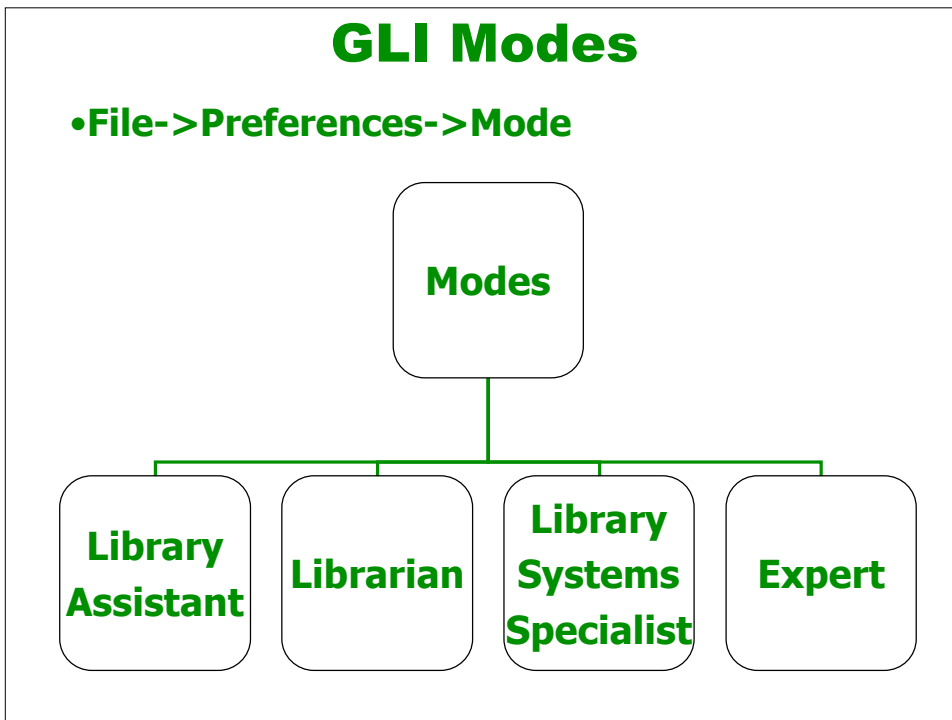
Includes substantial sections on the GLI
(Sections 3.1 and 3.2, 36 pp)

- ❖ Tooltips

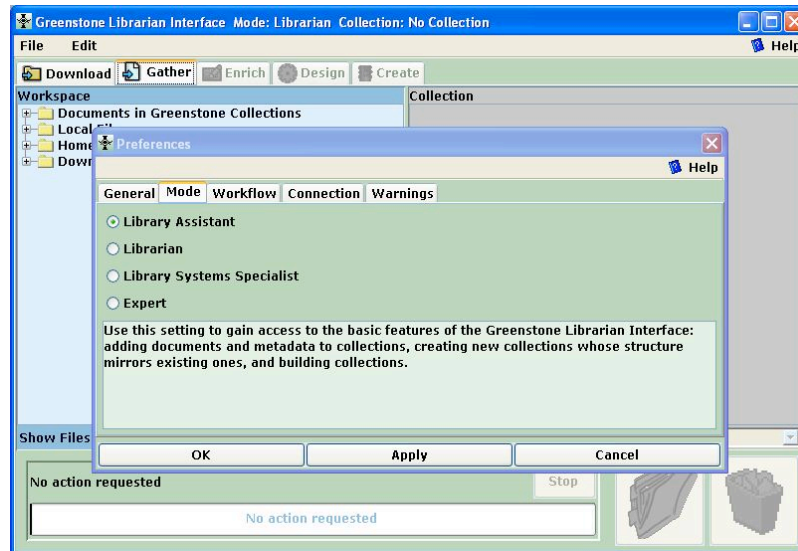
- ❖ On-line help

Agenda

- 
- ❖ Formatting
 - ❖ Formatting extended metadata
 - ❖ Changing metadata sets
 - ❖ Collection configuration file
 - ❖ GLI modes
 - ❖ PDF documents
 - ❖ PPT documents
 - ❖ Exploding metadata databases
 - ❖ Cross collection searching
 - ❖ Translate text
 - ❖ Full-text tagging
 - ❖ Creating a CD-ROM



GLI Mode Setting



GLI Modes

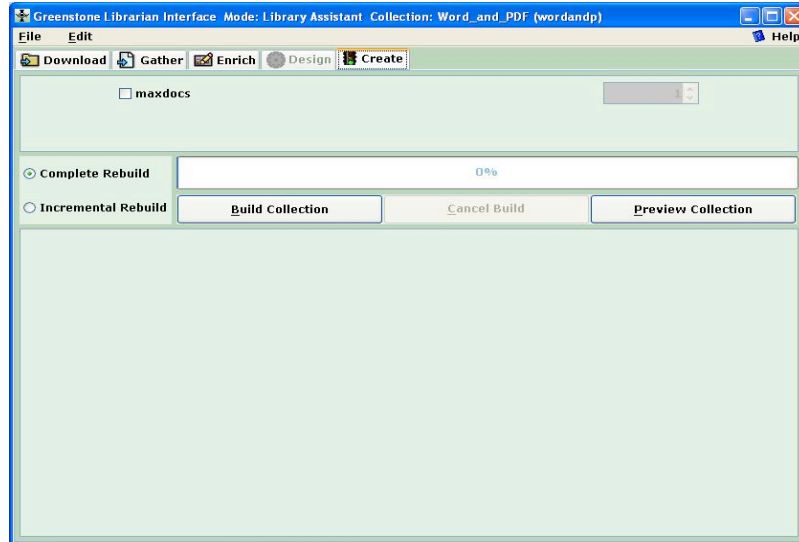
❖ Library Assistant

- Access to basic features: creating new collections; adding documents and metadata; building collections
- No design functions are available

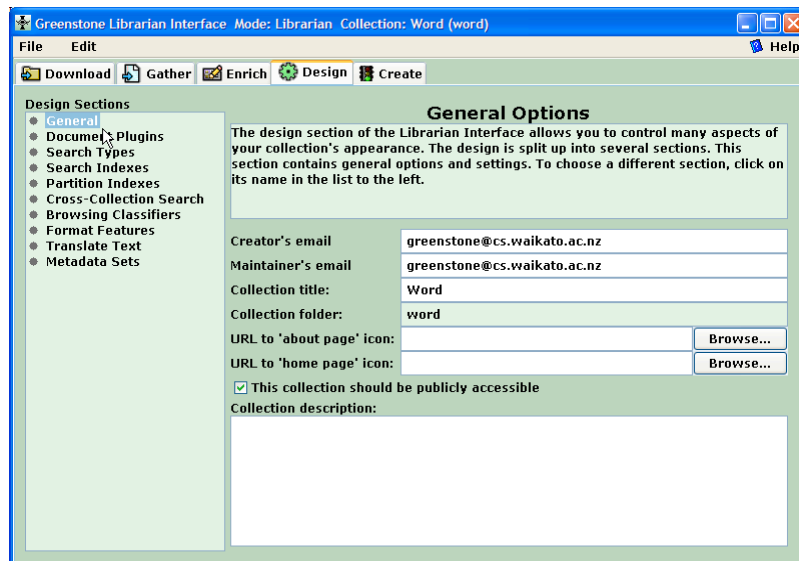
❖ Librarian

- Basic features + Design

Library Assistant Mode



Librarian Mode



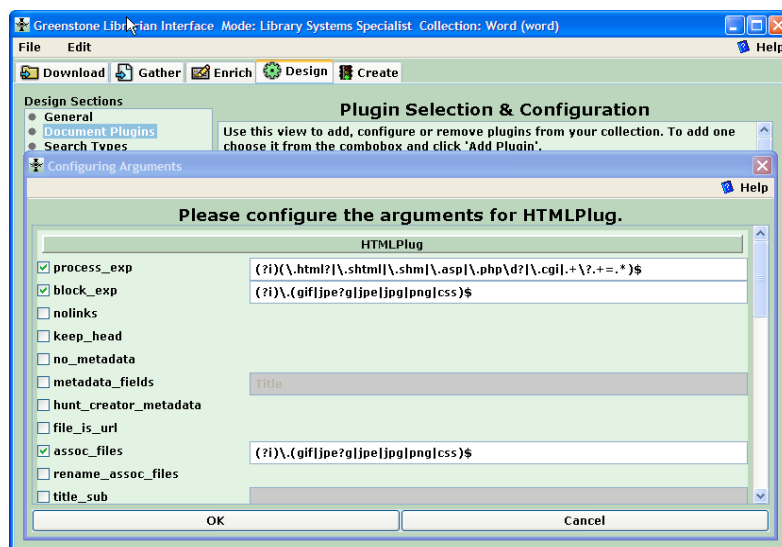
GLI Modes

❖ Library Systems Specialist

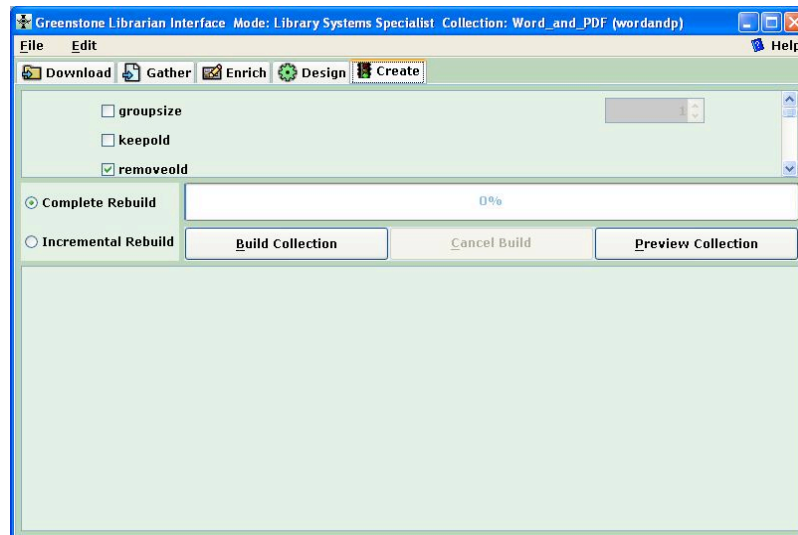
- Full use of GLI
- Formulate regular expressions to make use of formatting features
 - ❖ For example: HTML block expression


```
q^(?i)\.(gif|jpe?g|jpe|png|css)$^
```
- Partial options for import and building

Library Systems Specialist Mode



Library Systems Specialist Mode

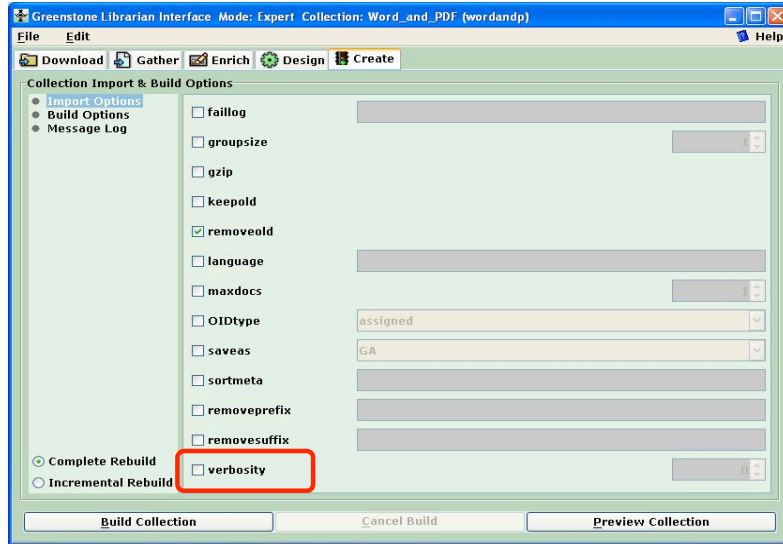


GLI Modes

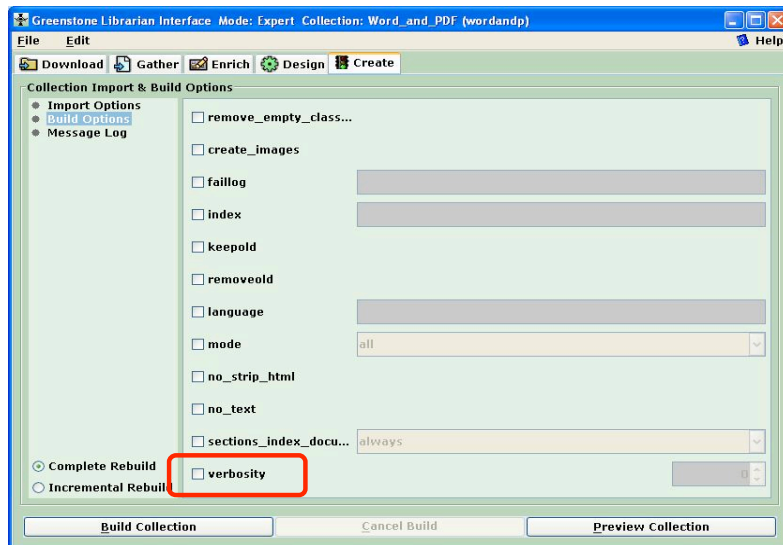
❖ Expert

- All features are enabled
- Recommended for experienced users
- Perform troubleshooting tasks
- Options to control the import and build processes
- Shows the output from the processes

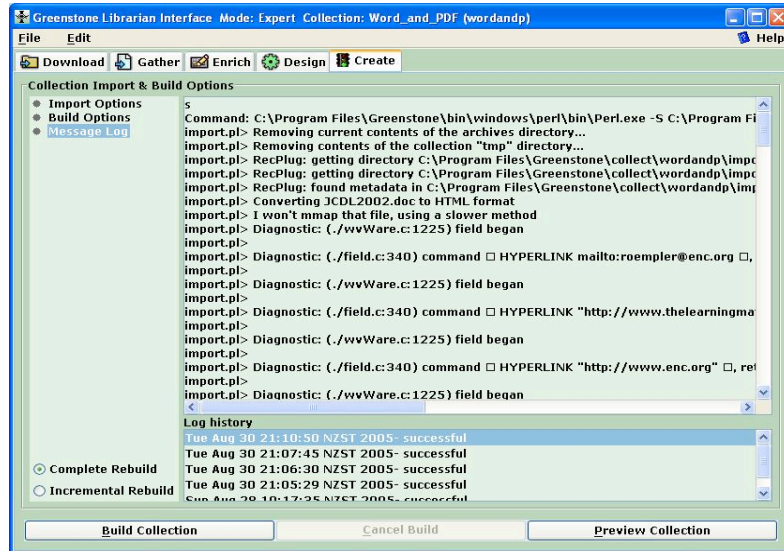
Expert Mode—Import Options



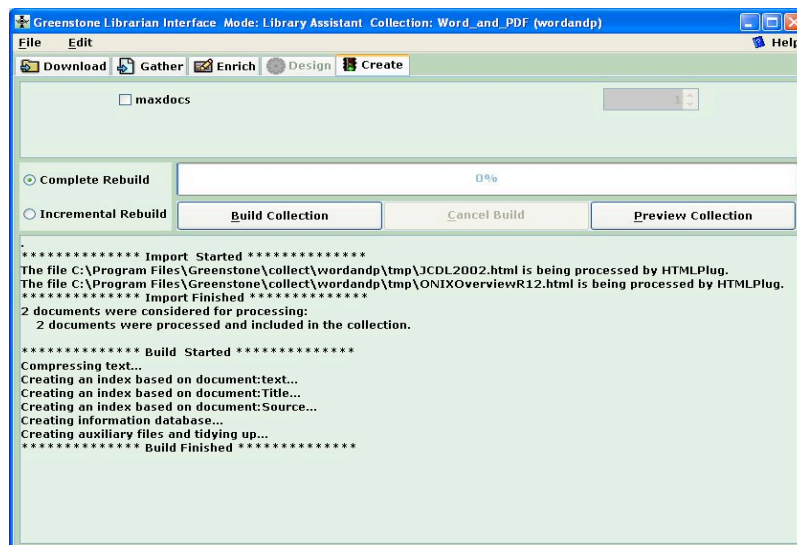
Expert Mode—Build Options

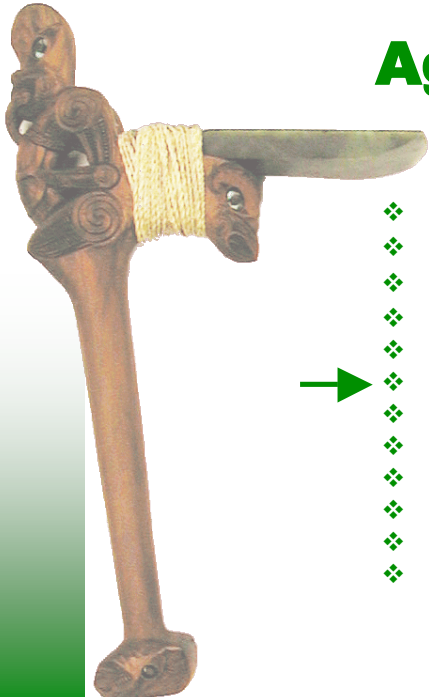


Message Log—Expert Mode



Building Message—Others





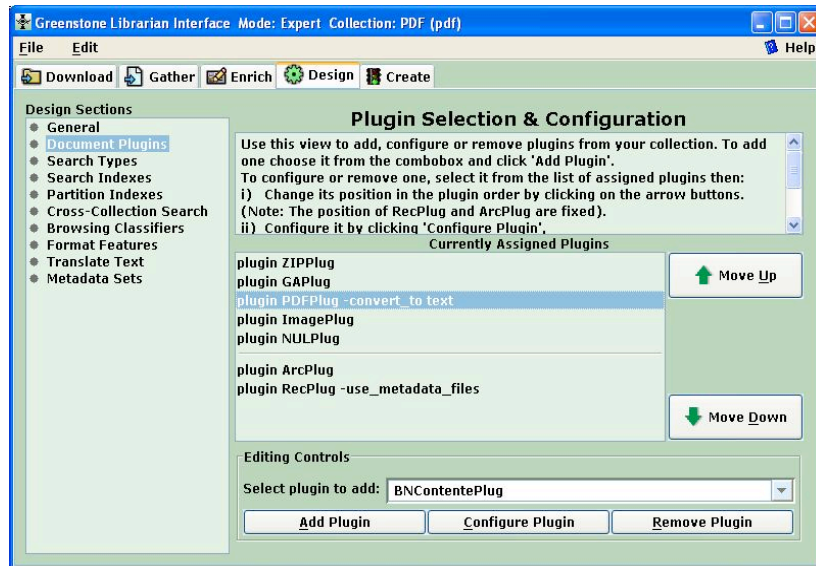
Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ➔ ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

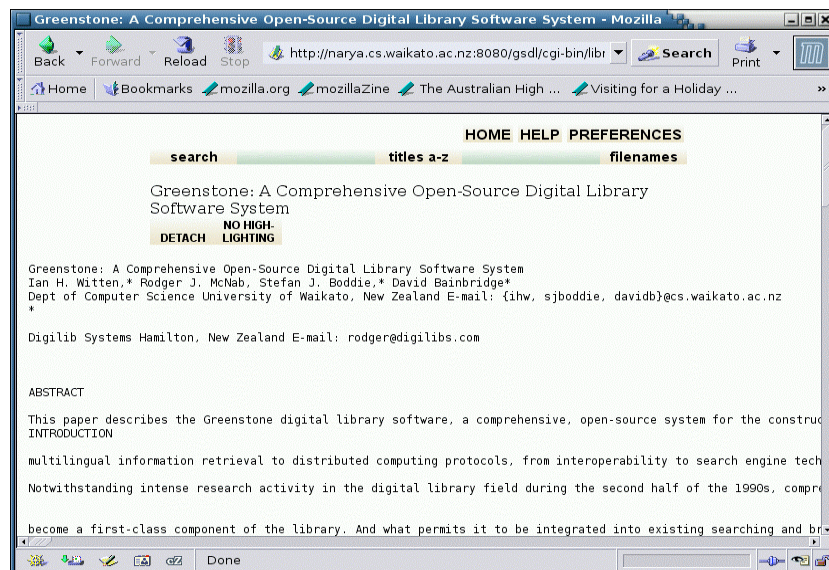
PDF Document

- ❖ PDF conversions in Greenstone
 1. Text only for Unix system
 2. HTML
 - ❖ use_sections option
 - ❖ complex option
 3. Image
 - ❖ ImageMagick needs to be installed
 - ❖ For advanced conversions, GhostScript must be installed
 - ❖ Use of convert utility
 - ❖ Convert_to
 - pagedimg_jpg
 - pagedimg_gif
 - pagedimg_png

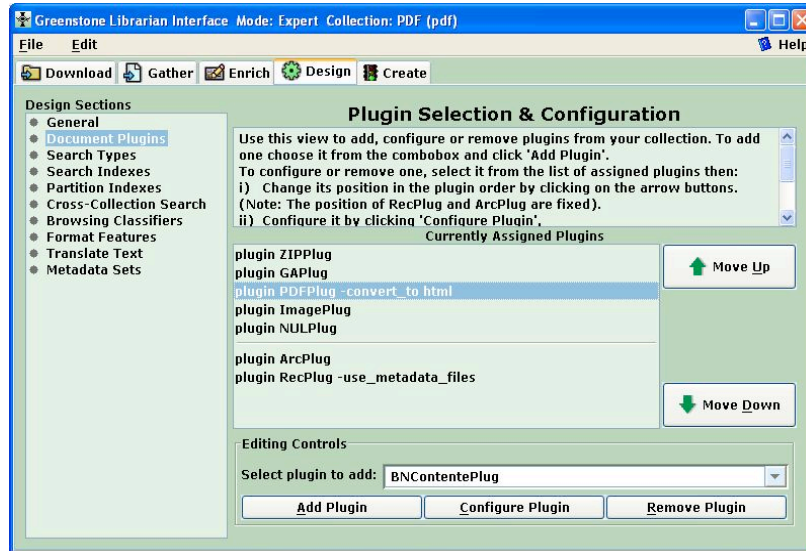
PDF -> Text



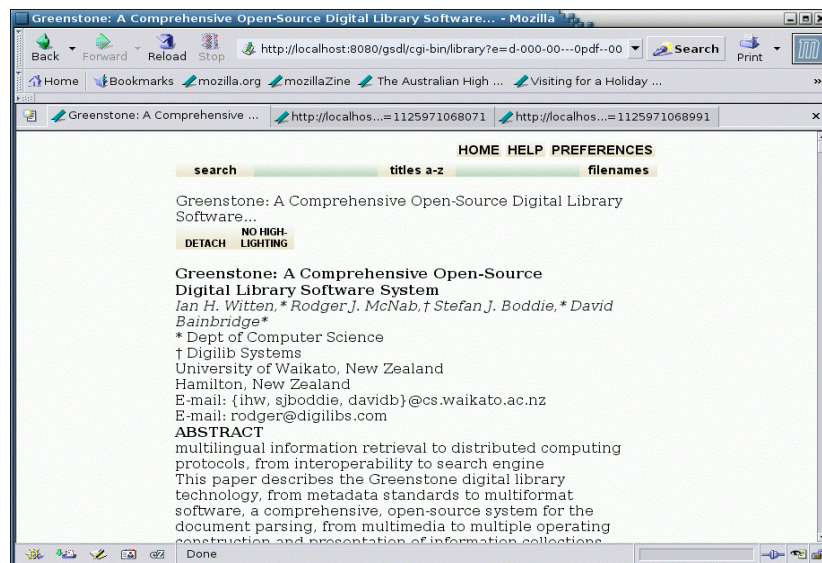
PDF: Text Document Display



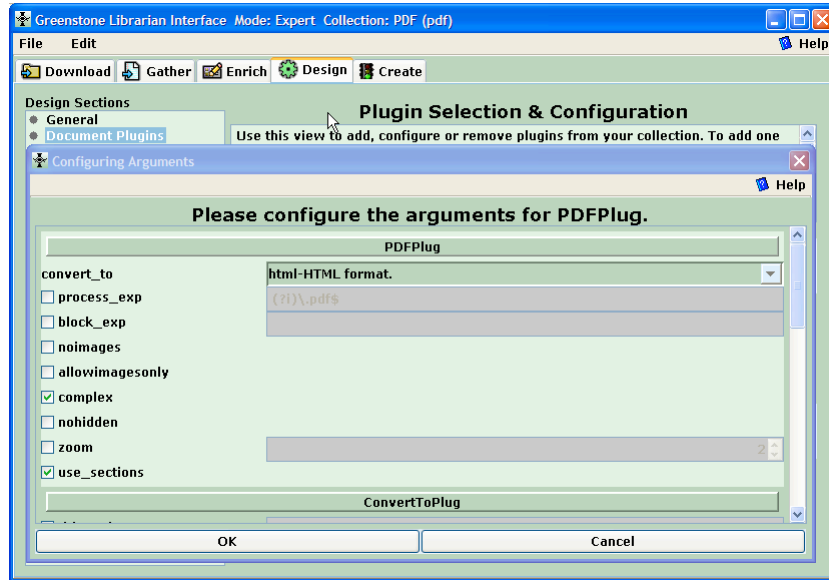
PDF -> HTML



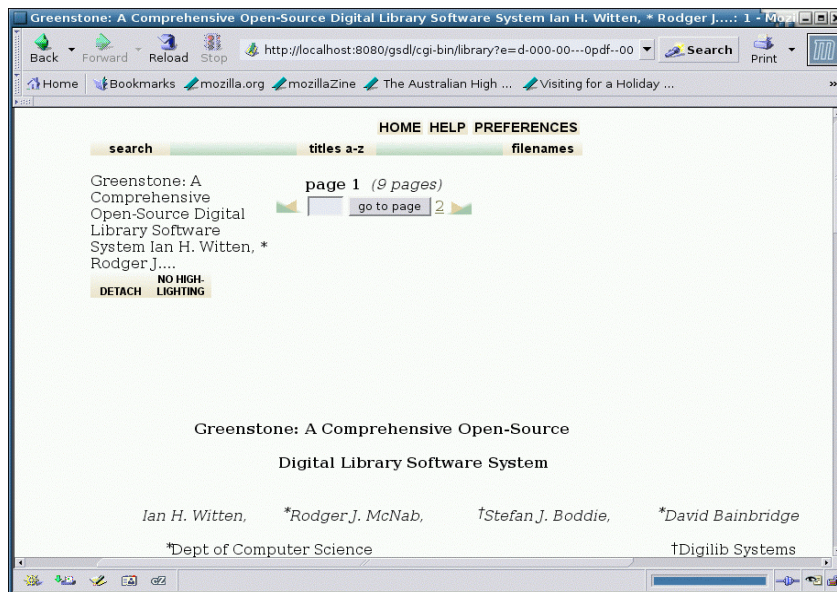
PDF: HTML Document Display 1



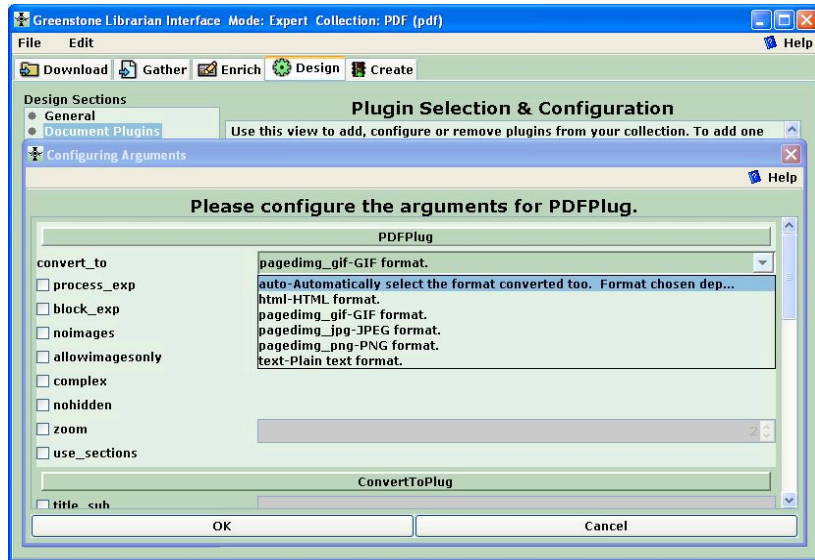
PDF: use_sections



PDF: HTML Document Display 2

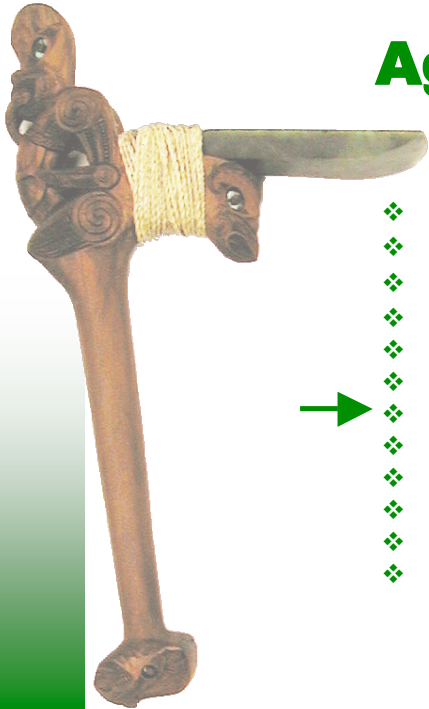


PDF -> Image



PDF: Image Document Display





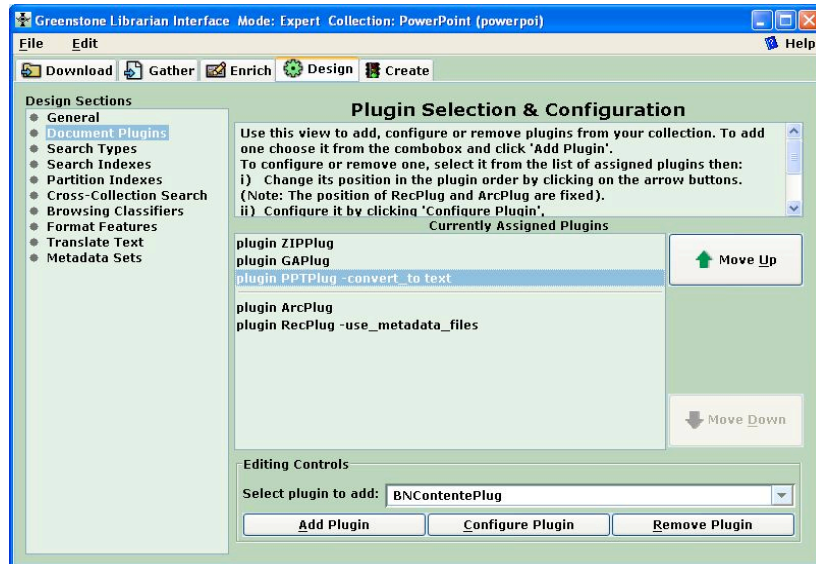
Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ➔ ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

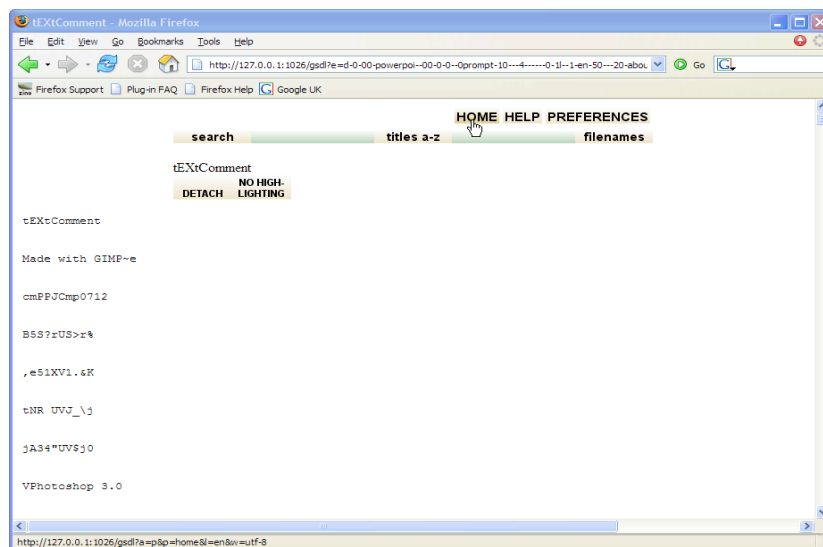
PowerPoint Document

- ❖ PPT conversions in Greenstone
 1. Text
 - ❖ use_strings option
 2. HTML
 3. Image (JPEG, GIF, PNG)
 - ❖ windows_scripting option
 - ❖ convert_to
 - pagedimg_jpg
 - pagedimg_gif
 - pagedimg_png

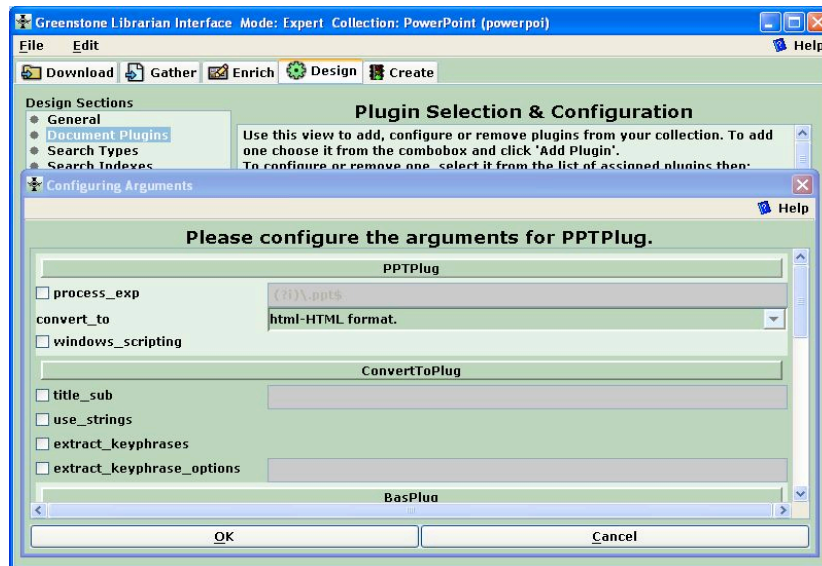
PPT -> Text



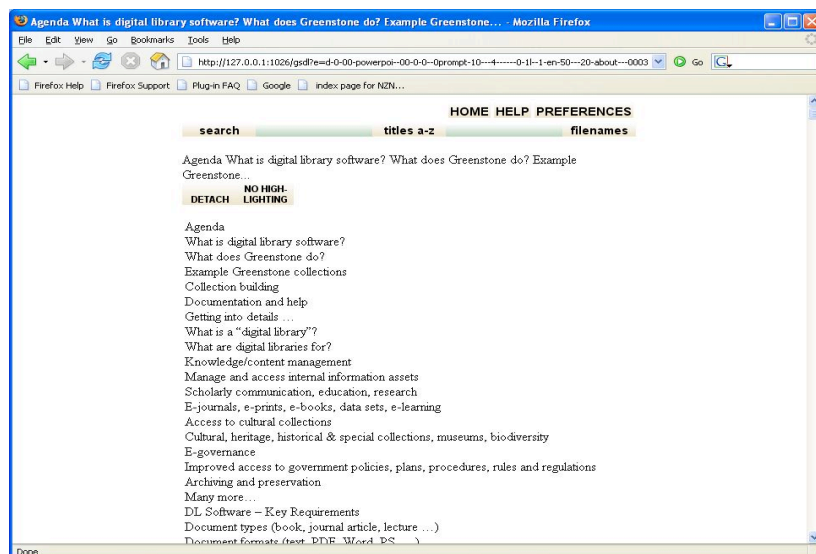
PPT: Text Document Display



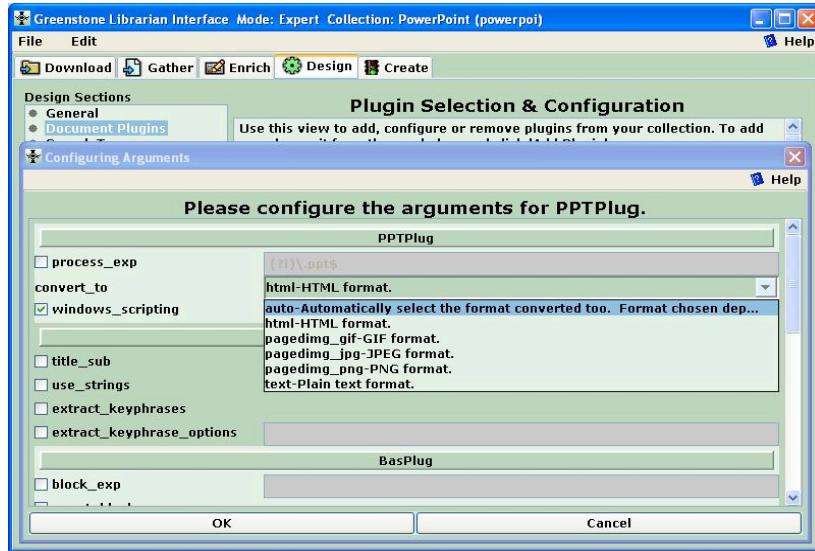
PPT -> HTML



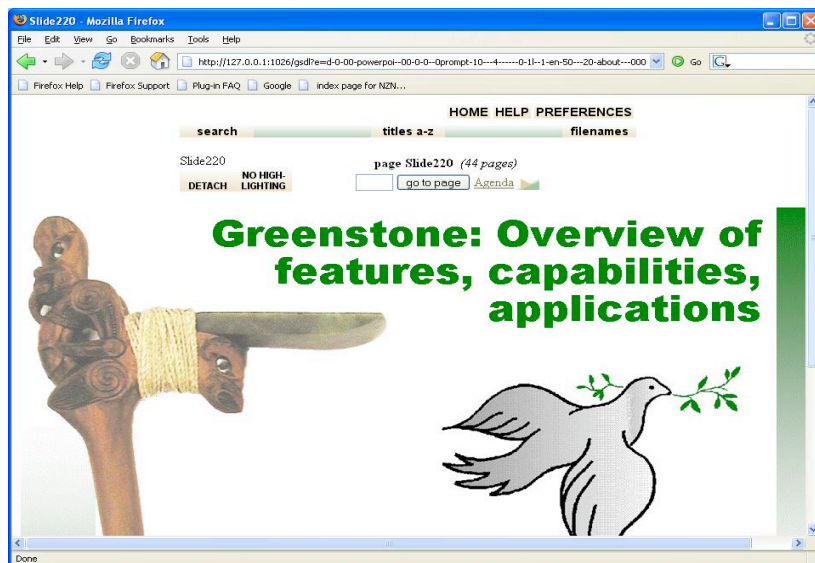
PPT: HTML Document Display



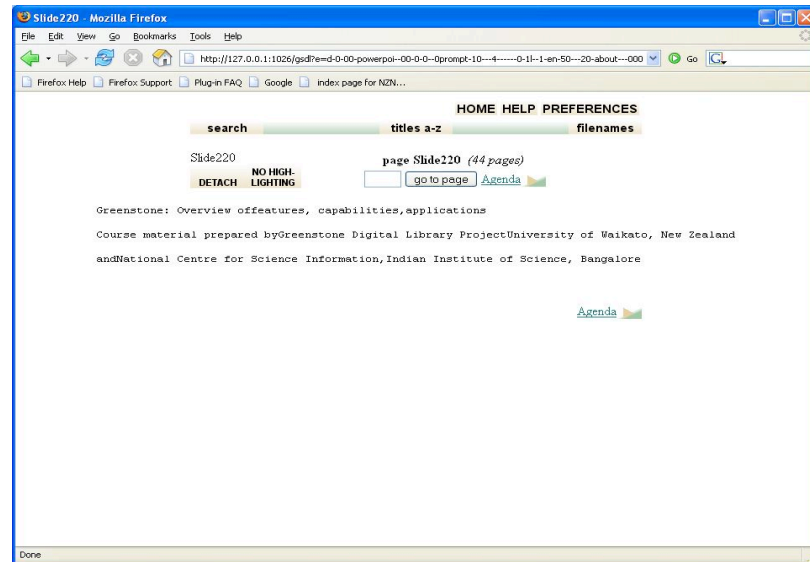
PPT -> Image



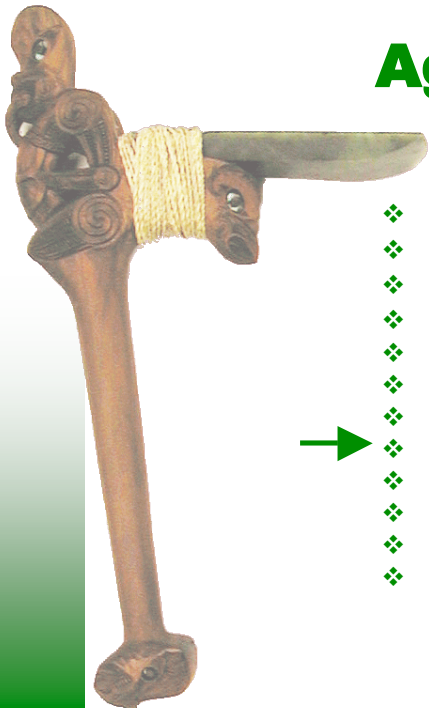
PPT Image: Image View



PPT Image: Text View



Agenda

- 
- ❖ Formatting
 - ❖ Formatting extended metadata
 - ❖ Changing metadata sets
 - ❖ Collection configuration file
 - ❖ GLI modes
 - ❖ PDF documents
 - ❖ PPT documents
 - ➔ ❖ Exploding metadata databases
 - ❖ Cross collection searching
 - ❖ Translate text
 - ❖ Full-text tagging
 - ❖ Creating a CD-ROM

Exploding metadata databases

- ❖ The GLI does not show metadata extracted from bibliographic files: CDS/ISIS, MARC etc.

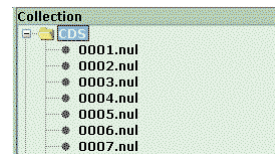
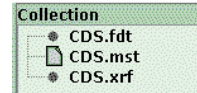


Why not?

- ❖ The GLI treats these files just like any other: it shows extracted and assigned metadata, but not the file's contents
 - Double-click the file to open it in the normal editor for this file type eg. WinISIS for CDS/ISIS files
- ❖ In rare cases you might want to convert these files into Greenstone format, allowing the metadata records to be visible and editable from the GLI
- ❖ This is irreversible: there's no going back!
 - Keep a backup of the original file

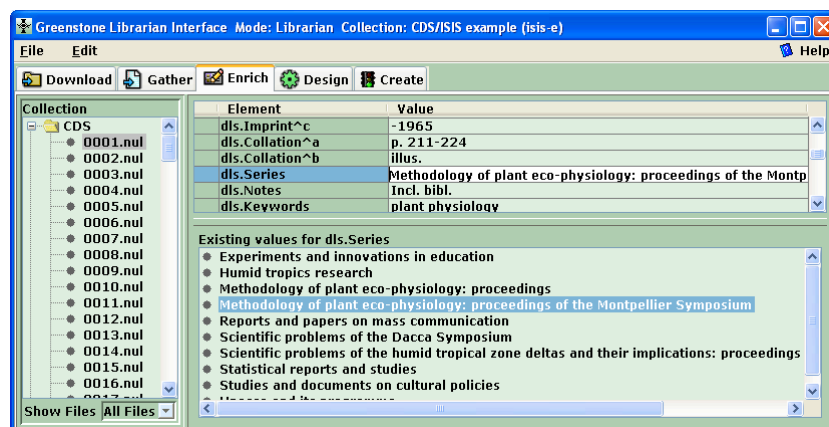
Exploding Metadata Databases

- ❖ Files that can be exploded are shown with a green icon in the GLI
- ❖ Right-click on the file and choose "Explode Metadata Database"
- ❖ You will be prompted to map the metadata into a metadata set
- ❖ When complete, the file will be replaced with a folder containing a file for each record



Exploding Metadata Databases

- ❖ The metadata can now be viewed and edited:

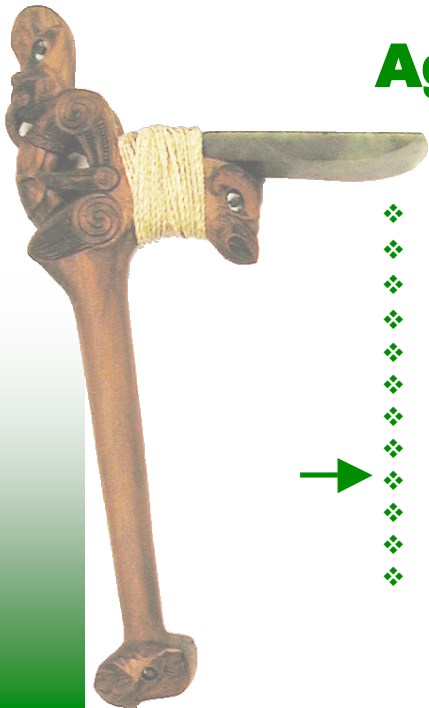


- ❖ Note: the GLI is *not* a real database system
 - Only feasible with a small number of records

Exploding Metadata Databases

- ❖ Change classifiers, index specifications and format statements to use the namespaced metadata elements
- ❖ When importing the collection, the files will now be processed by NULPlug rather than the plugin for the original file (eg. ISISPlug)

Agenda

- 
- ❖ Formatting
 - ❖ Formatting extended metadata
 - ❖ Changing metadata sets
 - ❖ Collection configuration file
 - ❖ GLI modes
 - ❖ PDF documents
 - ❖ PPT documents
 - ❖ Exploding metadata databases
 - ➔ ❖ Cross collection searching
 - ❖ Translate text
 - ❖ Full-text tagging
 - ❖ Creating a CD-ROM

Cross Collection Searching

- ❖ Select a list of collections
- ❖ Collections need the same indexes
- ❖ User can select which collections to search on Preferences page
- ❖ Format statements applied from original collection

Bibliography collections

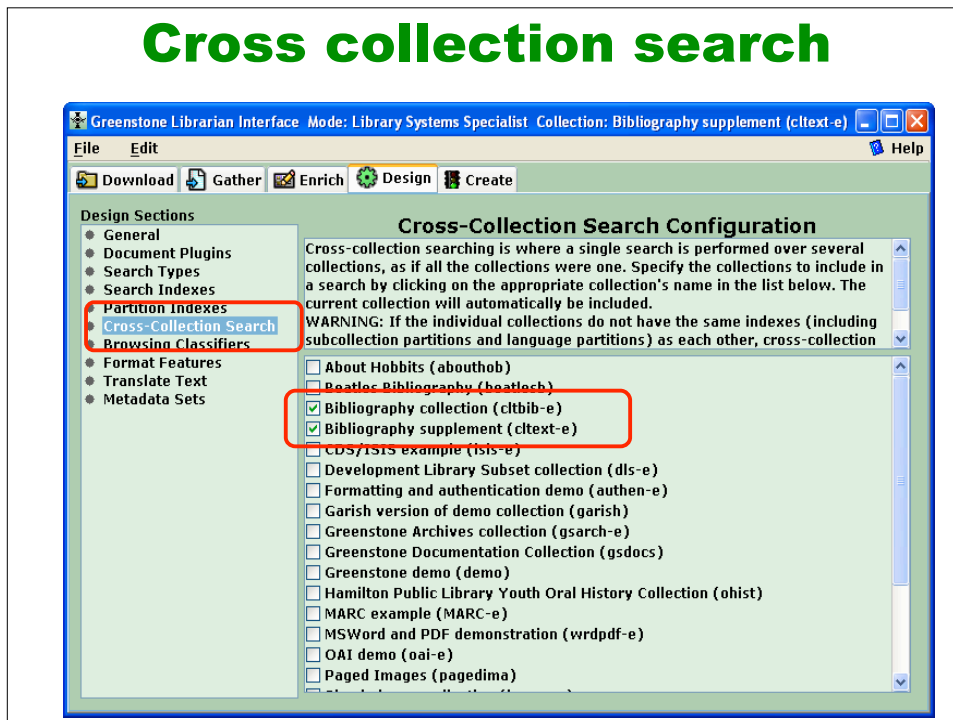
bibliography collection

With about 4,000 bibliography entries, this collection incorporates a form-based search interface that allows fielded searching. It is fairly complex.

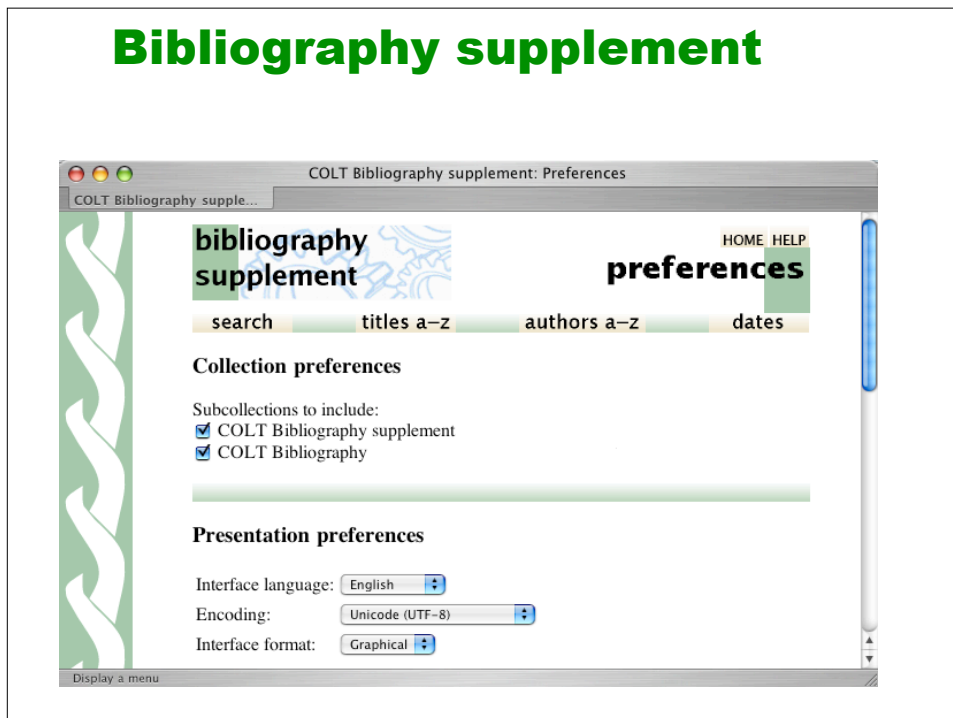
bibliography supplement

This tiny collection of 10 bibliography entries illustrates the "supercollection" facility which searches several collections together, seamlessly. It operates together with the [Bibliography collection](#), and its configuration file is almost the same.

Cross collection search



Bibliography supplement



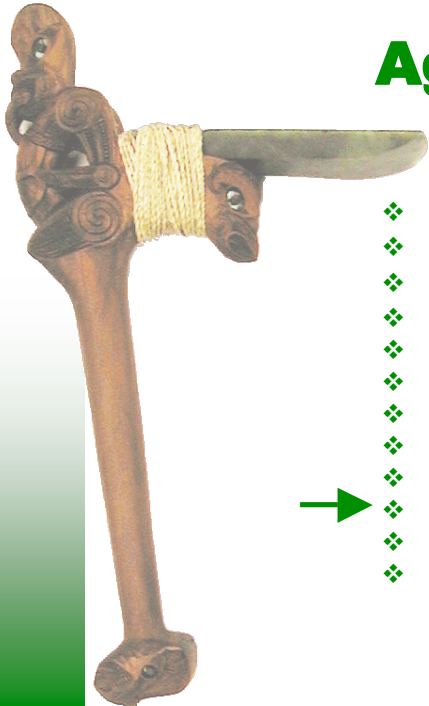
Cross-collection searching ...

Both collections

Supplement

Main collection

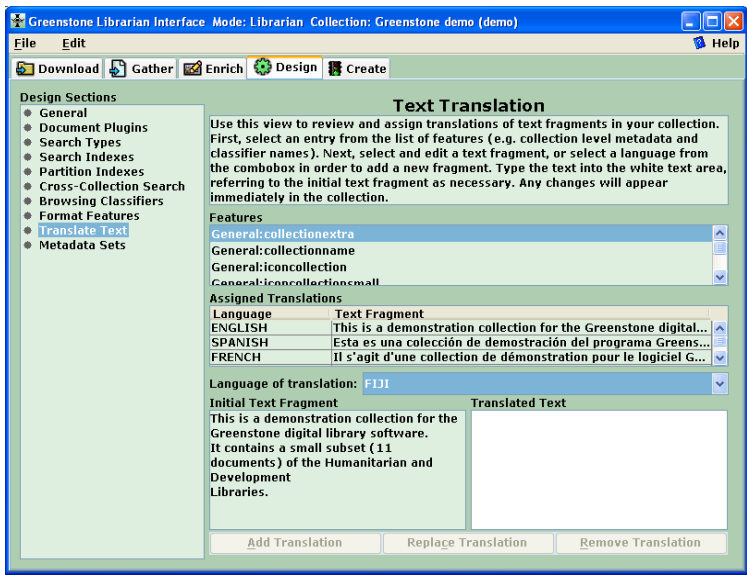
... but not browsing



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ➔ ❖ Translate text
- ❖ Full-text tagging
- ❖ Creating a CD-ROM

Translate Text



The screenshot shows the Greenstone Librarian Interface (GLI) window titled "Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo (demo)". The interface includes a menu bar (File, Edit, Help) and a toolbar (Download, Gather, Enrich, Design, Create). The main window is titled "Text Translation" and contains the following content:

Use this view to review and assign translations of text fragments in your collection. First, select an entry from the list of features (e.g. collection level metadata and classifier names). Next, select and edit a text fragment, or select a language from the combobox in order to add a new fragment. Type the text into the white text area, referring to the initial text fragment as necessary. Any changes will appear immediately in the collection.

Features

- General:collectionextra
- General:collectionname
- General:iconcollection
- General:iconcollectionsmall

Assigned Translations


Language	Text Fragment
ENGLISH	This is a demonstration collection for the Greenstone digital...
SPANISH	Esta es una colección de demostración del programa Greens...
FRENCH	Il s'agit d'une collection de démonstration pour le logiciel G...

Language of translation: **FDJ**

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

Translated Text

Buttons: Add Translation, Replace Translation, Remove Translation



Agenda

- ❖ Formatting
- ❖ Formatting extended metadata
- ❖ Changing metadata sets
- ❖ Collection configuration file
- ❖ GLI modes
- ❖ PDF documents
- ❖ PPT documents
- ❖ Exploding metadata databases
- ❖ Cross collection searching
- ❖ Translate text
- ➔ ❖ Full-text tagging
- ❖ Creating a CD-ROM

Full Text Tagging

- ❖ While creating large digital collections:
 - the collection must be organized
 - the larger the collection the greater the need for organization
 - the larger the documents the greater the need for sections/subsections
- ❖ Greenstone lets you tag the full text of documents
- ❖ Then you can read them hierarchically ...
- ❖ ... and search them by section

HASHa72X → Farming snails 2: Choosing snails; Care and harvesting; Further improvement
 HASHa72X.1 → Preface
 HASHa72X.2 → The snails
 HASHa72X.2.1 → Choosing your snails
 HASHa72X.2.2 → Putting your snails in the pen
 HASHa72X.2.3 → When to take out the fully grown snails
 HASHa72X.3 → Taking care of your snails

EXPAND CONTRACT
 TEXT CONTENTS
 DETACH NO HIGH-LIGHTING

search subjects titles a-z organization hc
 Search for chapters which contain some of the words
 chapters section titles entire books Begin S

Full Text Tagging...

To show the hierarchical structure, tag the source files like this:

```

<!--
<Section>
<Description>
<Metadata name="Title">Realizing human rights for
  poor people: Strategies for achieving the
  international development targets</Metadata>
</Description>
-->
      (text of section goes here)
<!--
</Section>
-->
  
```


Full Text Tagging...

- ❖ Section tags define a hierarchical structure
- ❖ Sections can be nested within other sections
- ❖ All sections must be nested within a single enclosing section that encompasses the entire document
- ❖ In the collection configuration file, put

```
HTMLPlug -description_tags
```

- ❖ Mainly for HTML, but can be used in Word and PDF documents.

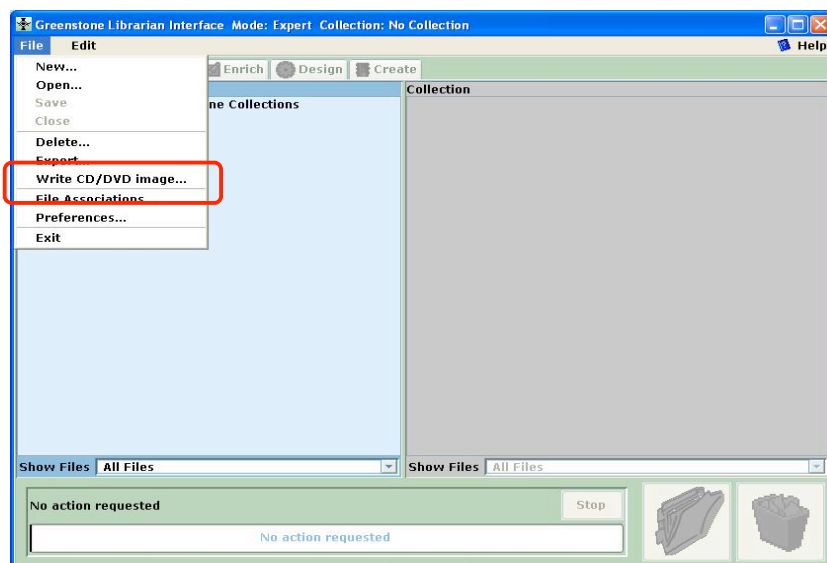
Agenda

- 
- ❖ Formatting
 - ❖ Formatting extended metadata
 - ❖ Changing metadata sets
 - ❖ Collection configuration file
 - ❖ GLI modes
 - ❖ PDF documents
 - ❖ PPT documents
 - ❖ Exploding metadata databases
 - ❖ Cross collection searching
 - ❖ Translate text
 - ❖ Full-text tagging
 - ❖ Creating a CD-ROM

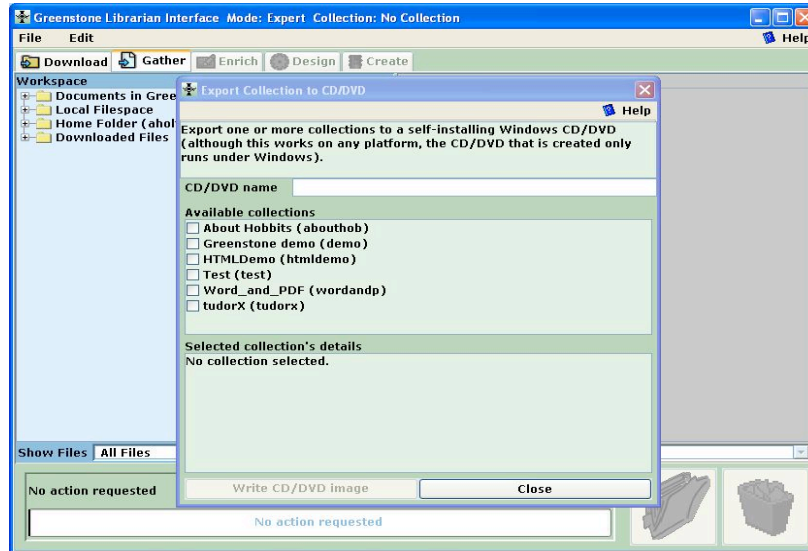
Creating a CD-ROM

- ❖ Export any Greenstone collection as a CD-ROM
- ❖ Self-installing
- ❖ Windows only (sorry!)
- ❖ Full Installation of Greenstone
- ❖ In the Librarian
 - File -> Write CD/DVD image
- ❖ C:\Program Files\Greenstone\tmp\exported_xxx

Exporting to CD-ROM



Exporting to CD-ROM



Note

- ❖ CD-ROM's created this way have not been tested extensively under different Windows configurations
- ❖ But they should work on all Windows platforms ...
- ❖ ... except 3.1/3.11 (is this a problem?)

Creating a CD-ROM:
demo