

# Adding metadata— and using it

Course material prepared by  
Greenstone Digital Library Project  
University of Waikato, New Zealand

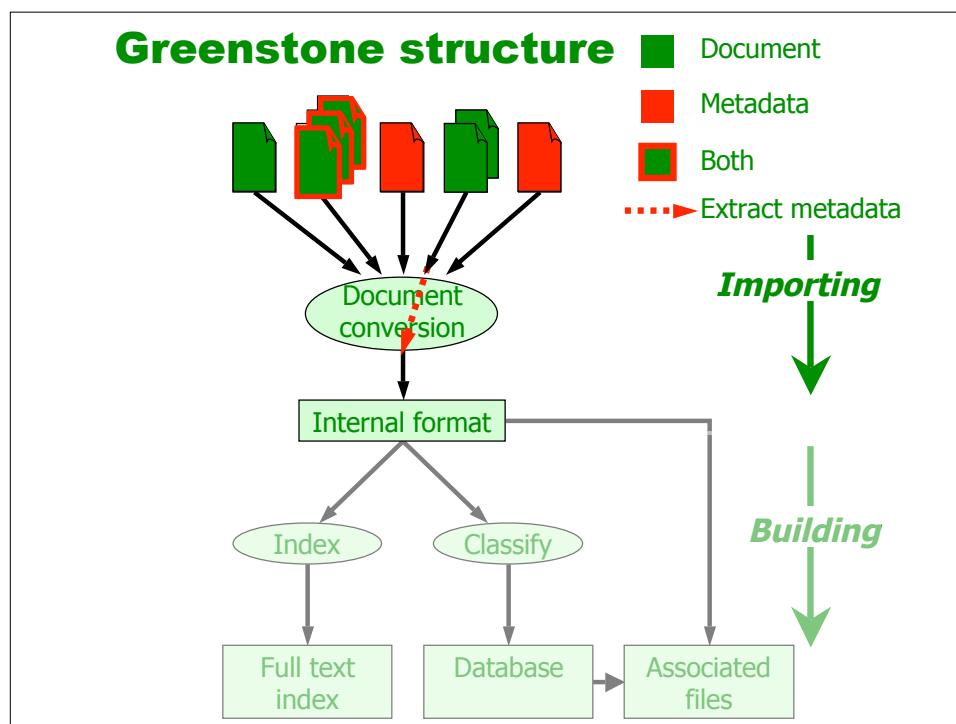


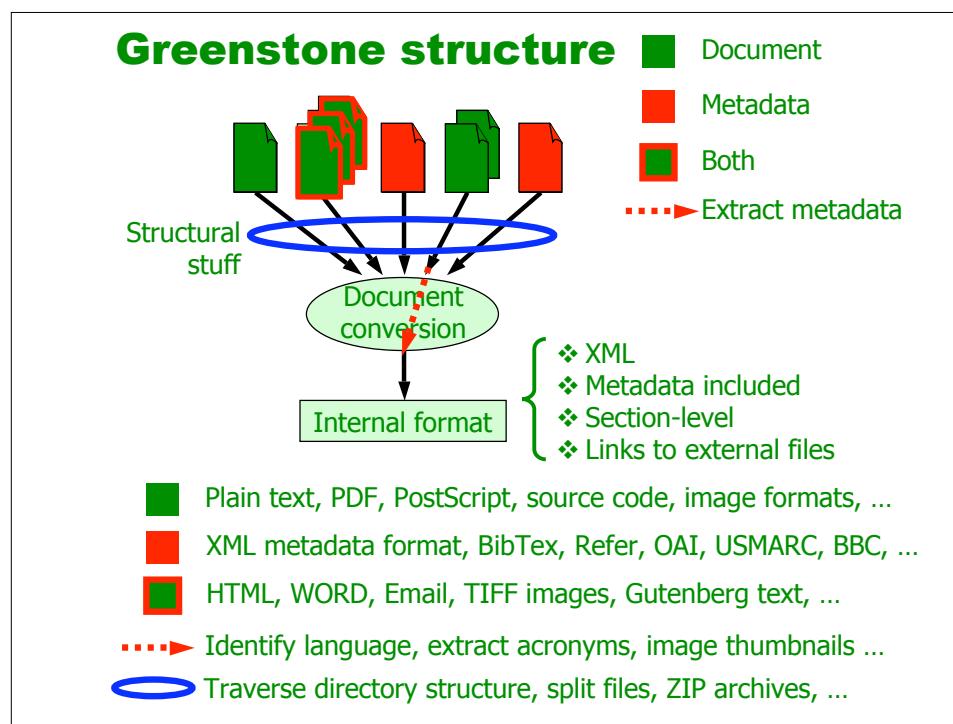
## Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

# **Collection building process**

- ❖ Input: Set of documents + metadata
    - Various formats, e.g. Word, PDF, HTML, images ...
  - ❖ Import: Conversion to common internal (XML) format
    - Uses third party tools
  - ❖ Build: Create indexes, browsing structures, metadata database
  - ❖ Output: Greenstone collection





## Metadata Sets

- ❖ Standard metadata sets
  - Librarian added
  - Good quality, but high cost
  - Many types
    - ❖ Dublin Core (dc)
    - ❖ Development Library Subset (dls)
    - ❖ New Zealand Government Locator Service (nzgls)
- ❖ Greenstone extracted metadata (ex)
  - Automatically extracted
  - Non-editable
  - May be poor quality, but cheap

## Dublin Core metadata

Metadata	Tag	Definition
<b>Title</b>	dc.Title	Name given to the resource
<b>Creator</b>	dc.Creator	Entity primarily responsible for making the content of the resource
<b>Subject and keywords</b>	dc.Subject	Topic of the content of the resource
<b>Description</b>	dc.Description	Account of the content of the resource
<b>Publisher</b>	dc.Publisher	Entity responsible for making the resource available
Contributor	dc.Contributor	Entity responsible for making contributions to the content of the resource
<b>Date</b>	dc.Date	Date associated with an event in the life cycle of the resource
Resource type	dc.Type	Nature or genre of the content of the resource
Format	dc.Format	Physical or digital manifestation of the resource
<b>Resource identifier</b>	dc.Identifier	Unambiguous reference to the resource within a given context: this is the object identifier or OID
<b>Source</b>	dc.Source	Reference to a resource from which the present resource is derived
<b>Language</b>	dc.Language	Language of the intellectual content of the resource
<b>Relation</b>	dc.Relation	Reference to a related resource
<b>Coverage</b>	dc.Coverage	Extent or scope of the content of the resource
Rights management	dc.Rights	Information about rights held in and over the resource

<b>DLS metadata</b>	
<b>Metadata</b>	<b>Definition</b>
dls.Title	Title
dls.Keyword	Keyword (originally "How To")
dls.Organization	Organization that produced the document
dls.Language	Language of the document
dls.Subject and Keywords	Hierarchical subject metadata
dls.AZList	A-Z range into which this document's title falls

### Example of DLS metadata

The screenshot shows the Greenstone Librarian Interface window. The title bar reads "Greenstone Librarian Interface Mode: Library Systems Specialist Collection: Greenstone demo (demo)". The menu bar includes File, Edit, Download, Gather, Enrich, Design, Create, Help, and a toolbar with icons for each. The left pane is titled "Collection" and lists several folder names: b17mie, b1base (which is selected), b20cre, b21wae, b22bue, ec158e, ec159e, ec160e, fb33fe, fb34fe, and wb34te. The right pane displays the DLS metadata for the selected file "b1base". A table shows the following data:

Element	Value
dls.Title	Little Known Asian Animals With a Promising Economic Future
dls.Organization	BOSTID
dls.Subject And Key...	Animal Husbandry and Animal Product Processing Other
dls.Keyword	introduce little-known Asian farm animals with a promising economic future
dls.Language	English
dls.AZList	A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P-Q-R-S-T-U-V-W-Z

At the bottom of the right pane, it says "No Metadata Element Selected". At the bottom left of the interface, there are buttons for "Show Files" and "All Files".

## Example of DLS metadata

The screenshot shows the Greenstone Librarian Interface in 'Library Systems Specialist' mode. The 'Collection' pane on the left lists several collections: b17mie, b1base, b20cre, b21wae, b22hue, b23fe, b24fe, and wb34te. The 'Design' tab is selected. In the main area, a table displays DLS metadata elements and their values:

Element	Value
dls.Title	on Animals With a Promising Economic Future (b1base)
dls.Organization	BOSTID
dls.Subject And Key...	Animal Husbandry and Animal Product Processing O...
dls.Keyword	introduce little-known Asian farm animals with a pro...
dls.Language	English
dls.AZList	A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P-Q-R-S-T-U-V-W...

Below the table, a section titled 'Existing values for dls.Title' lists various titles found in other files:

- Butterfly Farming in Papua New Guinea (b22hue)
- Development in practice: Toward Gender Equality (wb34te)
- Farming snails 1: Learning about snails; Building a pen; Food and shelter plant
- Farming snails 2: Choosing snails; Care and harvesting; Further improvement
- Little Known Asian Animals With a Promising Economic Future (b1base)
- Managing Tropical Animal Resources - Crocodiles as a Resource for the Tropics
- Microlivestock - Little-Known Small Animals with a Promising Economic Future
- The Courier - N°158 - July - August 1996 Dossier Communication and the me
- The Courier - N°159 - Sept- Oct 1996 Dossier Investing in People Country Rej
- The Courier - N°160 - Nov - Dec 1996 - Dossier Habitat - Country reports: Fiji
- The Water Buffalo: New Prospects For An Underutilized Animal (1984) (b21w)

## Example of DLS metadata

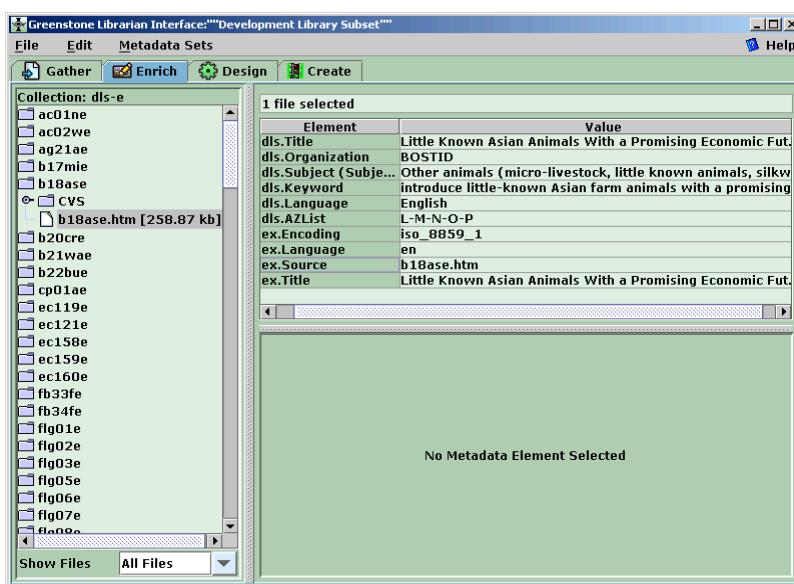
The screenshot shows the Greenstone Librarian Interface in 'Expert' mode. The 'Collection' pane on the left shows a folder structure under b1base, containing files like b1base.htm, b1base.jpg, p00.png, p06.png, p09.png, p101.png, p13.png, p23.png, p26.png, p29.png, p33.png, p40.png, p43.png, p45.png, p48.png, p52.png, p55.png, p58.png, p61.png, p63.png, p66.png, p68.png, p70.png, p73.png, and p77.png. The 'Design' tab is selected. In the main area, a table displays DLS metadata elements and their values:

Element	Value
dls.Title	Little Known Asian Animals With a Promising Econo...
dls.Organization	BOSTID
dls.Subject And Key...	Animal Husbandry and Animal Product Processing O...
dls.Keyword	introduce little-known Asian farm animals with a pro...
dls.Language	English
dls.AZList	A-B-C-D-E-F-G-H-I-J-K-L-M-N-O-P-Q-R-S-T-U-V-W...

Below the table, a note states: "This piece of metadata is inherited from a folder above it, and cannot be edited. You can visit the actual folder containing the metadata by clicking on the folder icon at the start of the row."

<b>Extracted metadata</b>	
<b>Metadata</b>	<b>Definition</b>
ex.Title	Title extracted from HTML, Word, PDF
ex.Source	Name of source file, e.g. Apache.html
ex.Language	Language of document, e.g. en, sp, fr
ex.Encoding	Encoding of document, e.g. iso_8859_1
...	...
ex.Acronym	Acronyms that appear in the document
...	...

## Example of extracted metadata



The screenshot shows the Greenstone Librarian Interface with the title bar "Greenstone Librarian Interface: 'Development Library Subset'". The interface has tabs: Gather, Enrich, Design, and Create. The Gather tab is selected. On the left, there is a tree view of files under a collection named "dls-e", with "b1base.htm [258.87 kb]" selected. On the right, a table displays the extracted metadata for this file:

Element	Value
dls.Title	Little Known Asian Animals With a Promising Economic Future
dls.Organization	BOSTID
dls.Subject (Subject)	Other animals (micro-livestock, little known animals, silkworms, etc.)
dls.Keyword	introduce little-known Asian farm animals with a promising future
dls.Language	English
dls.AZList	L-M-N-O-P
ex.Encoding	iso_8859_1
ex.Language	en
ex.Source	b1base.htm
ex.Title	Little Known Asian Animals With a Promising Economic Future

The status bar at the bottom shows "No Metadata Element Selected".

## Example of extracted metadata

The screenshot shows the Greenstone Librarian Interface window titled "Development Library Subset". The left pane displays a tree view of files under the collection "dis-e", with "b18base.htm" selected. The right pane shows a table of extracted metadata:

Element	Value
dls.Title	Little Known Asian Animals With a Promising Economic Future
dls.Organization	BOSTID
dls.Subject (Subje...)	Other animals (micro-livestock, little known animals, silkw... introduce little-known Asian farm animals with a promising
dls.Keyword	introduce little-known Asian farm animals with a promising
dls.Language	English
dls.AZList	L-M-N-O-P
ex.Encoding	iso_8859_1
ex.Language	en
ex.Source	b18base.htm
ex.Title	Little Known Asian Animals With a Promising Economic Future

A note below the table states: "This element, ex.Title, belongs to the Greenstone metadata set. This set represents metadata that can usually be extracted automatically. To check what ex.Title metadata can be extracted create the collection using the 'Create' tab."

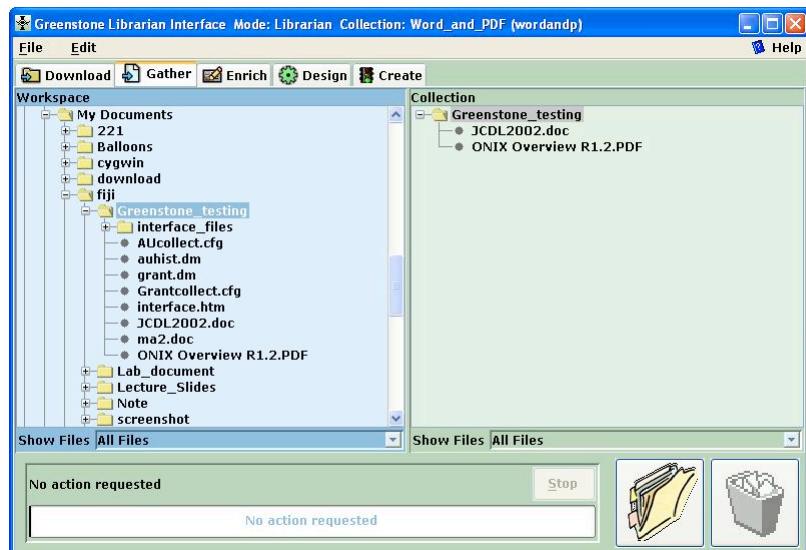
## Agenda

- 
- ❖ Overview of collection building
  - ❖ Metadata sets
  - ❖ Adding metadata in GLI (+demo)
  - ❖ GLI tricks
  - ❖ Review: Searching and browsing
  - ❖ General Options
  - ❖ Plugins
  - ❖ Searching – indexes
  - ❖ Browsing – classifiers
  - ❖ Simple formatting
  - ❖ Form search
  - ❖ Partitioning indexes
  - ❖ PHIND phrase index
  - ❖ CDS/ISIS
  - ❖ GEMS: metadata set editing

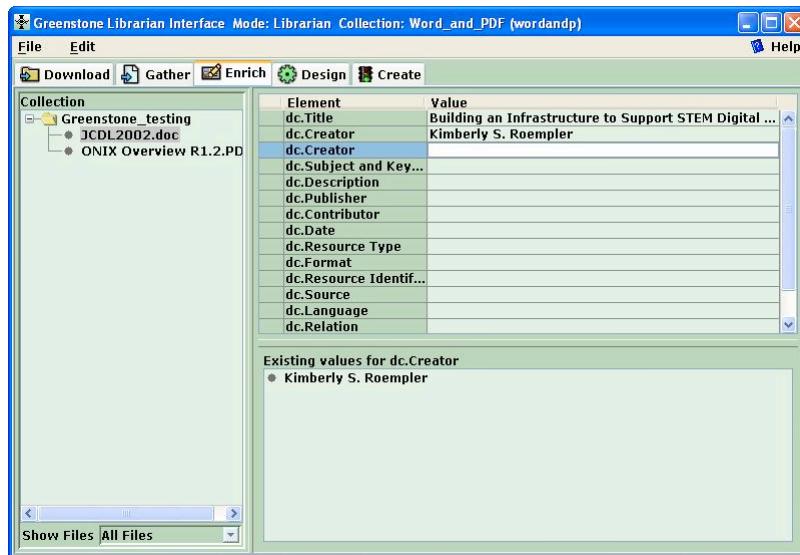
## GLI: Adding metadata

- ❖ Enrich pane
- ❖ Folder level or document level
- ❖ Specified metadata set
- ❖ Can view extracted metadata but not edit

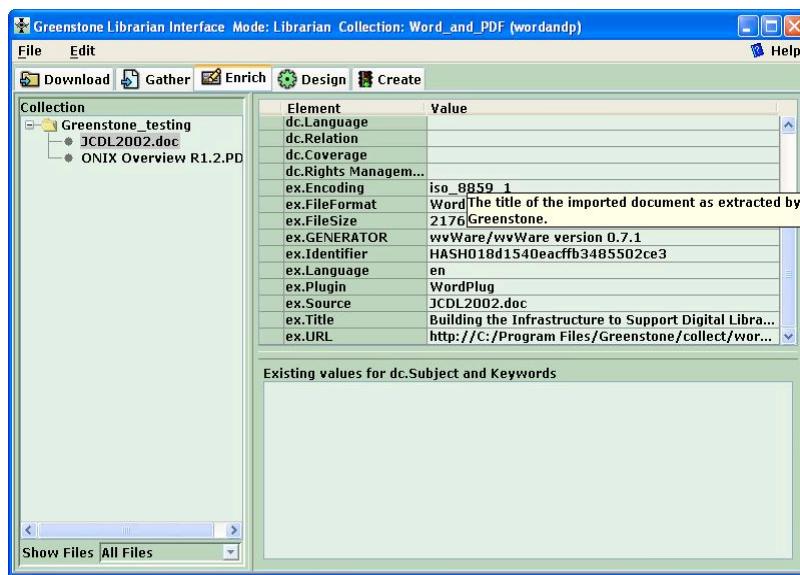
## GLI: adding metadata



## GLI: adding metadata



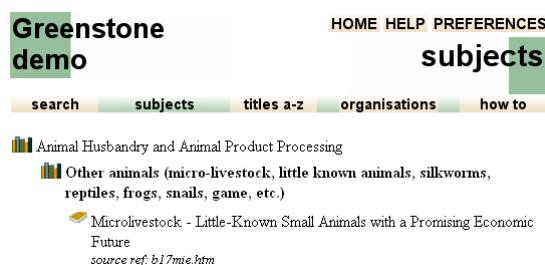
## Extracted Metadata



## Greenstone Librarian Interface demo: adding metadata

## Hierarchical Metadata

❖ Metadata is often hierarchical:



❖ How is this done in Greenstone?

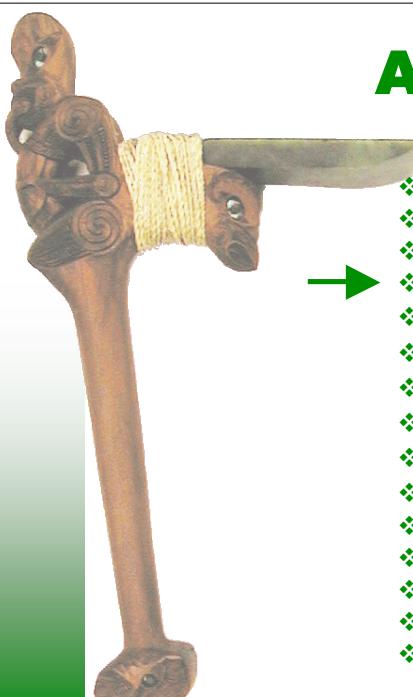
## Hierarchical Metadata

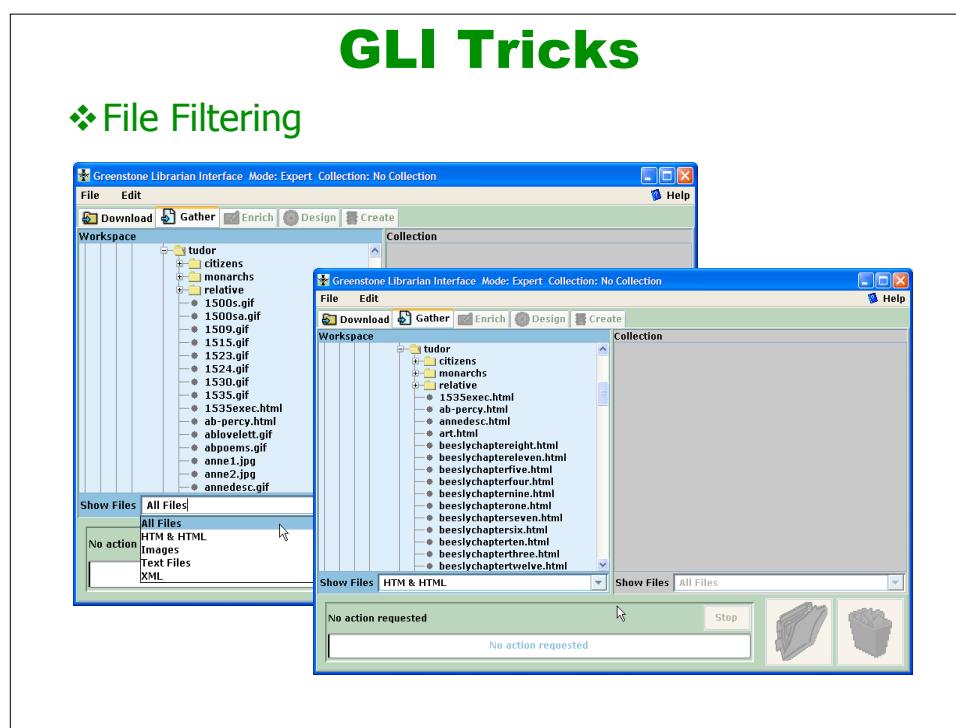
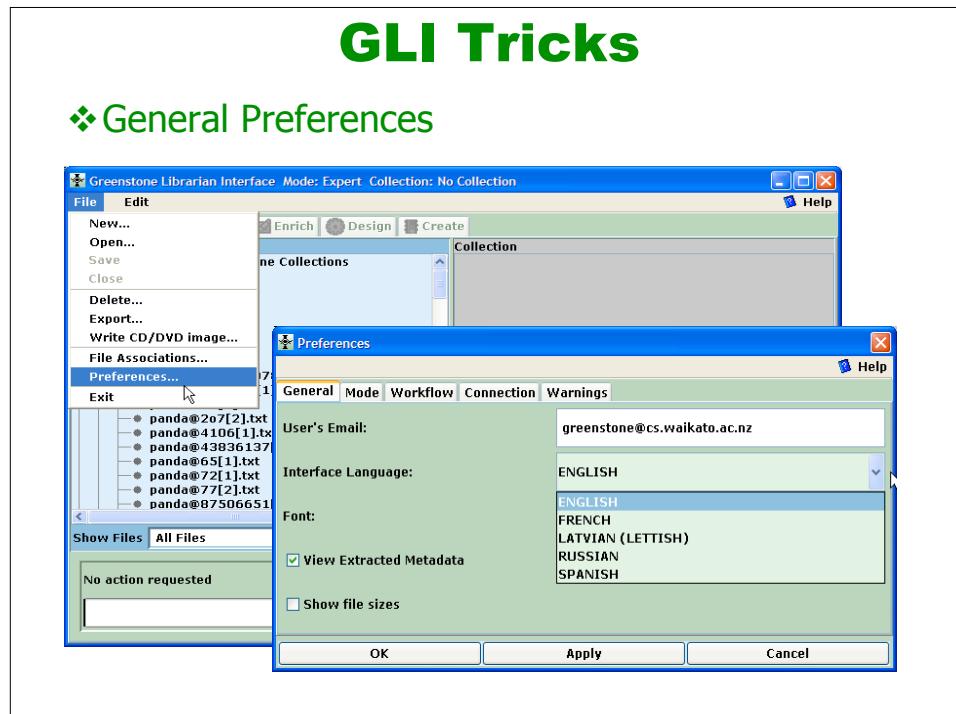
- ❖ In the GLI's Enrich pane, use the pipe character (|) to separate the levels

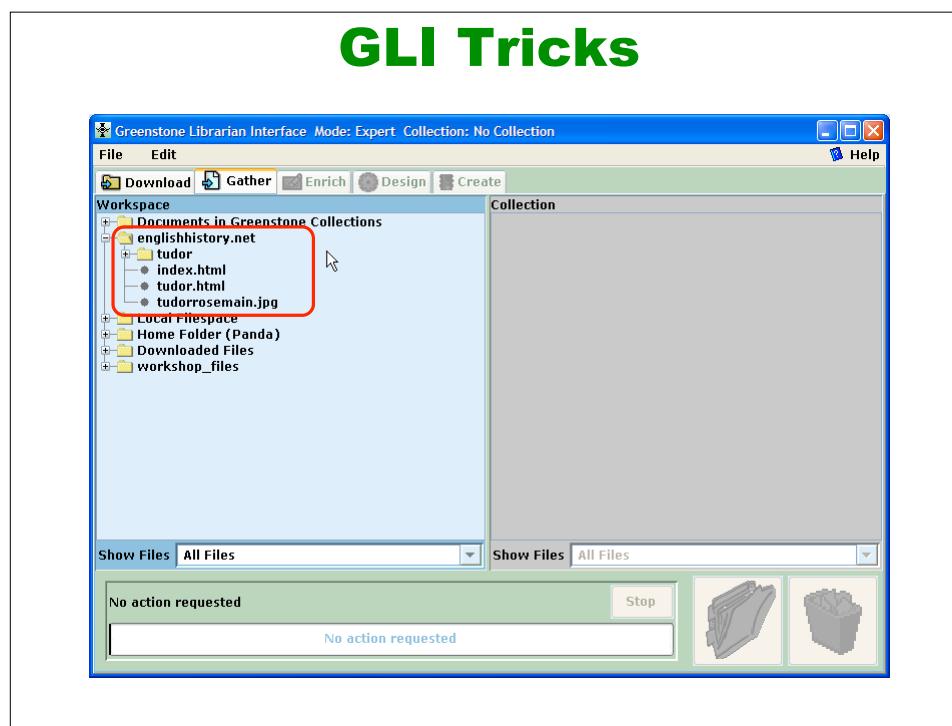
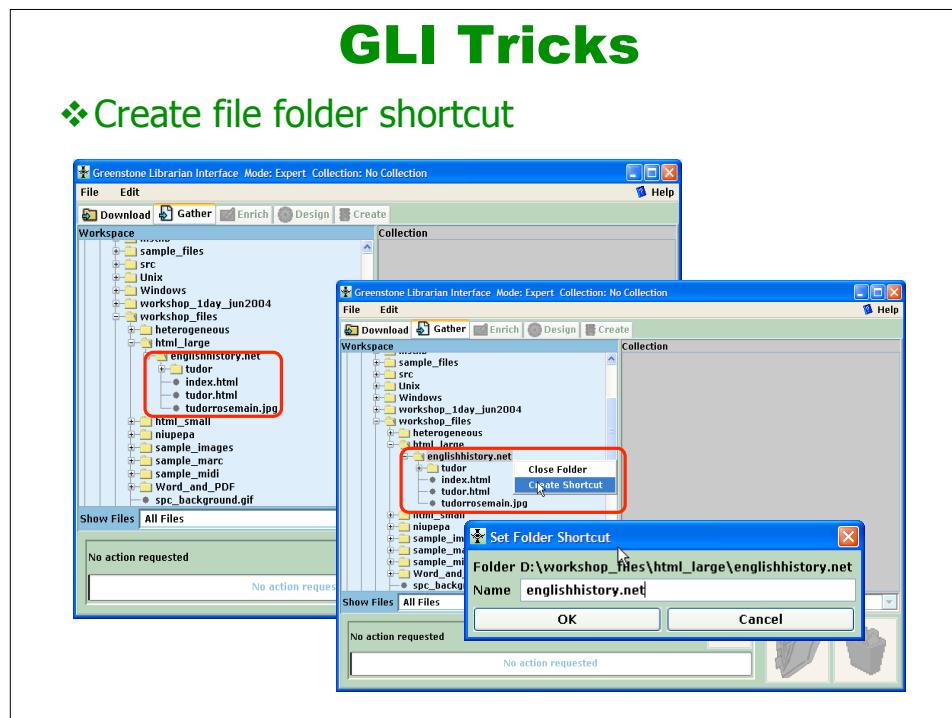


- ❖ Preview the hierarchy in the value tree

## Agenda

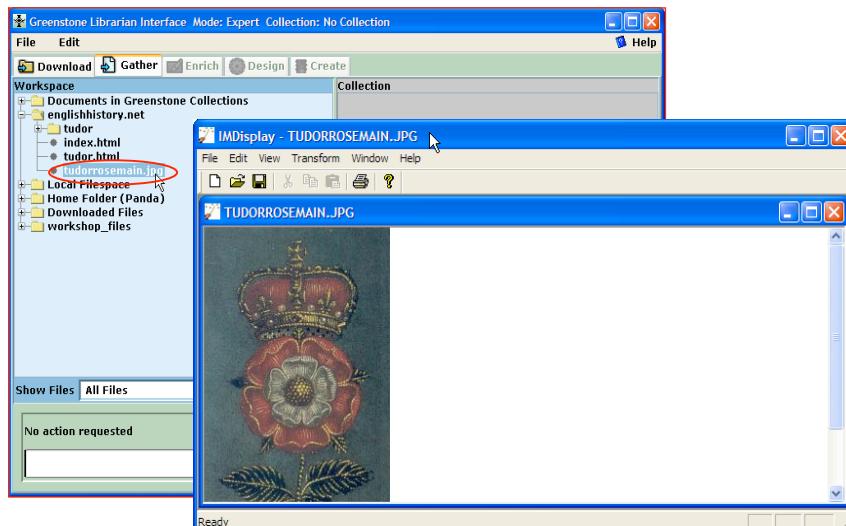
- 
- ❖ Overview of collection building
  - ❖ Metadata sets
  - ❖ Adding metadata in GLI (+demo)
  - ❖ GLI tricks
  - ❖ Review: Searching and browsing
  - ❖ General Options
  - ❖ Plugins
  - ❖ Searching – indexes
  - ❖ Browsing – classifiers
  - ❖ Simple formatting
  - ❖ Form search
  - ❖ Partitioning indexes
  - ❖ PHIND phrase index
  - ❖ CDS/ISIS
  - ❖ GEMS: metadata set editing



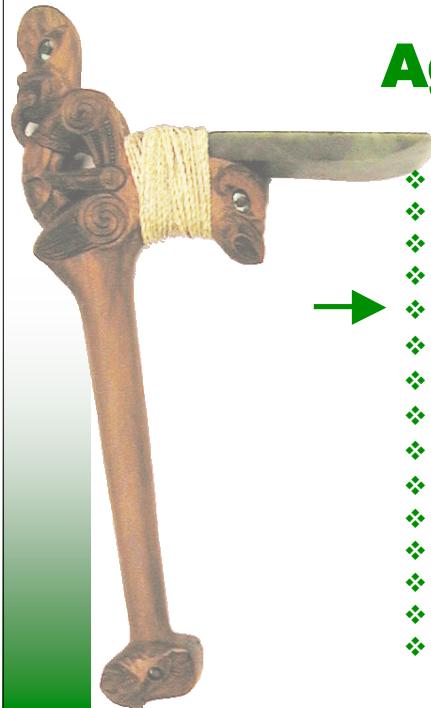


## GLI Tricks

### ❖ Viewing files: Double-clicking



## Agenda



- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Hierarchical document model

❖ Metadata specified at any level

**Title metadata**

Farming snails 2: Choosing snails; Care and harvesting; Further improvement - Netscape

search subjects titles a-z organization how to

34

farming snails 2 choosing snails care and harvesting further improvement

Preface

The snails

- Choosing your snails
- Putting your snails in the pen
- When to take out the fully grown snails

Taking care of your snails

- Wet the food and shelter plants and moisten the ground
- Tend the food and shelter plants**
- Take away the weeds and creeping grass
- Feeding your snails other kinds of food

Building the second pen

- When to build the second pen and how to build it
- Putting snails in your second pen
- Harvesting your snails
- When to harvest your snails

EXPAND TEXT CONTRACT CONTENTS NO HIGHLIGHTING DETACH

## Searching and browsing

❖ Searching

❖ Metadata-based browsing

Farming snails 2: Choosing snails; Care and harvesting; Further improvement - Netscape

File Edit View Go Window Help

Bookmarks Location: snails-1-00-w-0&cl=search&d=HASH1cbcd3220e50101bae361b&gc=1

search subjects titles a-z organization how to

34

farming snails 2 choosing snails care and harvesting further improvement

Preface

HOME HELP PREFERENCES

Subject Title Publisher "HowTo"

Dublin Core

ad hoc

**Multiple search indexes**

The screenshot shows the Microsoft Internet Explorer browser displaying the Greenstone demo homepage. The search interface includes dropdown menus for searching chapters, section titles, or entire documents. A red box highlights the dropdown menu, and green arrows point from the text "text" and "metadata" labels to the dropdown options "chapters" and "entire documents".

**text**

**metadata**

**Greenstone demo**

search subjects titles a-z organisations how to

Search for chapters that contain some of the words

chapters  
section titles  
entire documents

about

**About this collection**

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

**How to find information in the Greenstone demo collection**

There are 5 ways to find information in this collection:

- search for particular words
- access publications by subject
- access publications by title
- access publications by organization
- access publications by "how to" listing

You can search for particular words that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be

**Collection-dependent**

The screenshot shows the Netscape browser displaying a search results page for the query "Honkala". The search interface includes dropdown menus for searching references, titles, authors, or conference/publication titles. A red box highlights the dropdown menu, and green arrows point from the text "metadata" to the dropdown options "titles", "authors", and "conference/publication titles".

**computer science bibliographies**

search browse

Search for references which contain some of the words

Honkala  
titles  
authors  
conference/publication titles  
references  
references and abstracts

Word count: 1 - 1 - 102

4 documents matched the query.

Juha Honkala (1994) "On generalized DTOL systems and their fixed points" *Theoretical Computer Science* 127(2), 269-286. From [Bibliography for the journal "Theoretical Computer Science"](#)

Juha Honkala (1993) "On DOL systems with immigration" *Theoretical Computer Science* 120(2), 229-245. From [Bibliography for the journal "Theoretical Computer Science"](#)

Juha Honkala (1991) "On algebraic generalized zeta functions of formal power series" *Theoretical Computer Science* 79(1), 263-273. From [Bibliography of "Theoretical Computer Science"](#)

Juha Honkala and Werner Kuich (1997) "On Lindenmayerian algebraic power series" *Theoretical Computer Science* 183(1), 113-142. From [Bibliography of "Theoretical Computer Science"](#)

The screenshot displays two search interfaces side-by-side. On the left is a Microsoft Internet Explorer window showing a search page for 'arabic collection'. The search bar contains the query 'السلطان المنصور' (The Sultan Mansour) and the results section shows several Arabic documents. On the right is another Microsoft Internet Explorer window showing a search page for 'titles a-z'. The search bar contains the query '部分 提问题的记录' (Part of the record asking questions) and the results section shows several Chinese documents.

This screenshot shows a complex multi-pane search interface for 'folktales: language extraction demo'. It features several overlapping windows and panes. At the top, a main window has tabs for 'titles a-z', 'language', and 'about'. Below it, several smaller windows and panes show search results for 'titles a-z', 'language', 'phrases', 'authors a-z', and 'organization'. Each pane has its own search bar and results list. The interface is highly integrated, with multiple search functions running simultaneously.

**Browsing using classifiers**

**AZList classifier (Title metadata)**

titles a-z

A B C D E F G H I J L M N O P Q R S T U Y W 0-9 Magazines

- » About the International Institute of Rural Reconstruction
- » Above and beyond
- » Abstracts on sustainable agriculture 1992 Gate- GTZ
- » Access to archival records: A review of current issues: A RAMP study
- » Accounting for the microbusiness a teaching manual - A Handbook for Volunteers
- » African agriculture: The critical choices
- » Agribusiness in India: Lessons and visions (1996)
- » Agricultural development workers training manual - Volume III Crops
- » Agricultural development workers training manual: Volume I Orientation for Trainers
- » Agricultural development workers training manual: Volume II Extension Skills
- » Agricultural development workers training manual: Volume IV Livestock
- » Agricultural Expansion and Pioneer Settlements in the Humid Tropics (1988)
- » Agricultural extension
- » Agricultural Extension: Guidelines for extension workers in rural areas
- » Agricultural policy in India: need for a fresh look (1992)
- » Agriculture and rural development in the 1990's and beyond: redesigning the chemistry between state and institutions of development (1992)
- » Agro-forestry in the African Humid Tropics (1982)

**AZCompactList classifier (Creator metadata)**

authors a-z

A-H I-L P-R Y

- » Applerley, M. D. (7)
- » Bainbridge, D. (29)
- » Bar, M. I. (1)
- » Beale, R. (2)
- » Beasley, C. (1)
- » Bell, T. C. (4)
- » Blandford, A. (4)
- » Boudie, S. J. (10)
- » Bray, B. (1)
- » Bray, Z. (1)
- » Brind, M. (1)
- » Buchanan, G. R. (8)
- » Ch., Y. (1)
- » Ch., Y. C. (1)
- » Choi, C. (1)
- » Cockburn, A. (1)
- » Connaway, L. S. (1)

**AZCompactList classifier (*Creator* metadata)**

nzdl publications

titles a-z

authors a-z

dates

phrases

A-H I-P R-Z  
Bainbridge, D. (29)

- Assembling and enriching digital library collections
- Bainbridge, D., Thompson, J., Witten, I. H. (2003) C. C. Marshall, G. Henry and L. Delcambre (eds), Proc Third ACM/IEEE Joint Conference on Digital Libraries (JCDL'03), Houston, Texas, 323-334. IEEE Computer Society, Los Alamitos, California.
- Bulk processing of optically scanned music
- Bainbridge, D., Wijaya, K. (1999) 7th International Conference on Image Processing and its Applications, Manchester, UK, pp 474-478.
- Dynamic digital library construction and configuration
- Bainbridge, D., Don, K. J., Buchanan, G. R., Witten, I. H., Jones, S. R., Jones, M., Barr, M. I. (2004) R. Heery and L. Lyon (eds), Proc Eighth European Conference on Research and Advanced Technology for Digital Libraries (ECDL'04), LNCS 3232, Bath, UK, 1-13. Springer-Verlag, Berlin.
- Forming a corpus of voice queries for music information retrieval: a pilot study
- Bainbridge, D., McPherson, J. R., Cunningham, S. J. (2002) Proc Third International Conference on Music Information Retrieval, 289-290.
- Greenstone: a comprehensive open-source digital library software

**GenericList classifier (*Title* metadata)**

nzdl publications

titles a-z

authors a-z

titles a-z

- ლაპარ არა-ქვეყნიური  
(lands ara kveknuri.doc)
- ლეილმაჯნუნიანი  
(leimajnuniani.doc)
- ლენინი  
(lenini.doc)
- ლიტერატურული საღამო  
(literaturuli saghamo.doc)
- ლოდის დამკორბელი  
(lodis damgorebeli.doc)

**DateList classifier  
(*Date* metadata)**

niupepa:  
māori newspapers

search series dates

1842-1843 1844-1846 1848 1855-1856 1857 1858 1859 1860 1861 1862  
1863-1882 1883 1884 1885-1886 1887-1897 1903-1904 1905-1906 1907-1908  
1909-1910 1911-1912 1913-1914 1915-1916 1917-1918 1919-1921

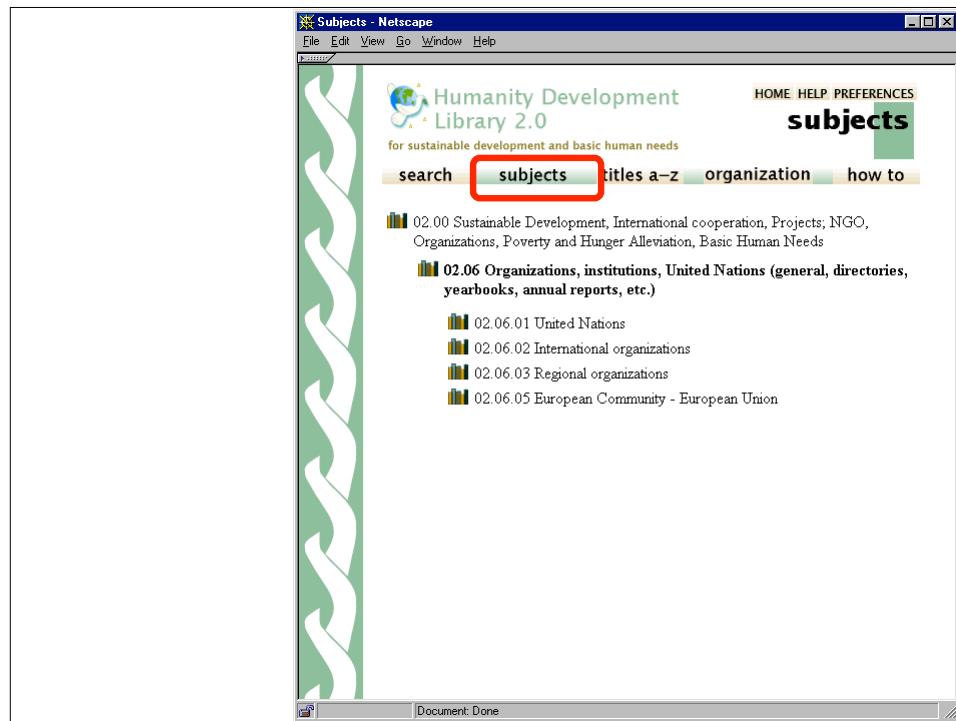
Month	Result Description
January	Te Karere o Nui Tirihi: Volume 1: January 1, 1842 Number 1
February	Te Karere o Nui Tirihi: Volume 1: February 1, 1842 Number 2
March	Te Karere o Nui Tirihi: Volume 1: March 1, 1842 Number 3
April	Te Karere o Nui Tirihi: Volume 1: April 1, 1842 Number 4
May	Te Karere o Nui Tirihi: Volume 1: May 2, 1842 Number 5
June	Te Karere o Nui Tirihi: Volume 1: June 1, 1842 Number 6
July	Te Karere o Nui Tirihi: Volume 1: July 1, 1842 Number 7
August	Te Karere o Nui Tirihi: Volume 1: August 1, 1842 Number 8
September	Te Karere o Nui Tirihi: Volume 1: September 1, 1842 Number 9
	Te Karere o Nui Tirihi: Volume 1: September 10, 1842

**Hierarchy classifier  
(*Subject* metadata)**

Humanity Development Library 2.0  
for sustainable development and basic human needs

search subjects titles a-z organization how to

- 02.00 Sustainable Development, International cooperation, Projects; NGO, Organizations, Poverty and Hunger Alleviation, Basic Human Needs
  - EARTH SUMMIT REPORT
  - 02.02 Development, national planning, national plans
  - 02.03 Project planning and evaluation (incl. project management and dissemination strategies)
  - 02.04 Regional development and planning incl. regional profiles
  - 02.05 Nongovernmental organizations (NGOs) in general, self-help organizations (their role in development). For specific organizations such as women's or peasants' see respective class within 06.
  - 02.06 Organizations, institutions, United Nations (general, directories, yearbooks, annual reports, etc.)
  - 02.07 Sustainable Development, Development models and examples; Best practices (general)
  - 02.08 Basic Human Needs
  - 02.09 Hunger and Poverty Alleviation

A slide titled "Agenda" featuring a decorative illustration of a Maori-style staff or hilt on the left. A green arrow points from the staff towards the agenda list on the right. The title "Agenda" is in large green font. The agenda list is a vertical column of green bullet points.

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

# General Options

The design section of the Librarian Interface allows you to control many aspects of your collection's appearance. The design is split up into several sections. This section contains general options and settings. To choose a different section, click on its name in the list to the left.

**Creator's email:** greenstone@cs.waikato.ac.nz  
**Maintainer's email:** greenstone@cs.waikato.ac.nz  
**Collection title:** Greenstone demo  
**Collection folder:** demo  
**URL to 'about page' icon:** \_httprefix\_/\_collect/demo/images/en/de [Browse...]  
**URL to 'home page' icon:** \_httprefix\_/\_collect/demo/images/en/de [Browse...]  
 **This collection should be publicly accessible**  
**Collection description:**  
 This is a demonstration collection for the Greenstone digital library software. It contains a small subset (11 documents) of the Humanitarian and Development Libraries.

**about**

**demo collection** ①

**demo collection** ②

**Help text generated automatically**

③

**about this collection**

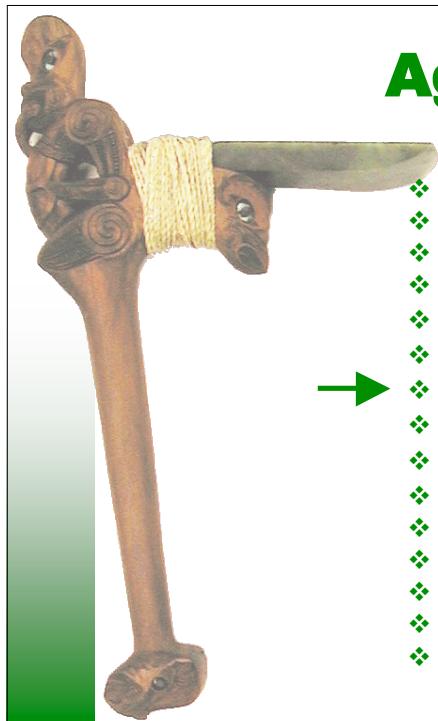
This is a demonstration collection for the Greenstone digital library software. It contains a small subset (11 books) of the Humanity Development Library.

**How to find information**

There are 5 ways to find information in this collection:

- search for particular words
- access publications by subject
- access publications by title
- access publications by organization
- access publications by how to listing

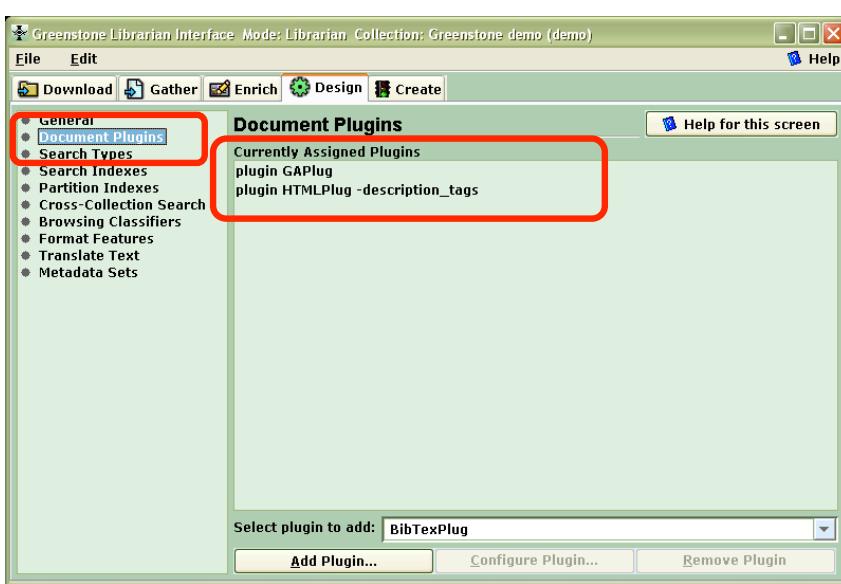
You can *search for particular words* that appear in the text from the "search" page. This is the first page that comes up when you begin, and can be reached from other pages.



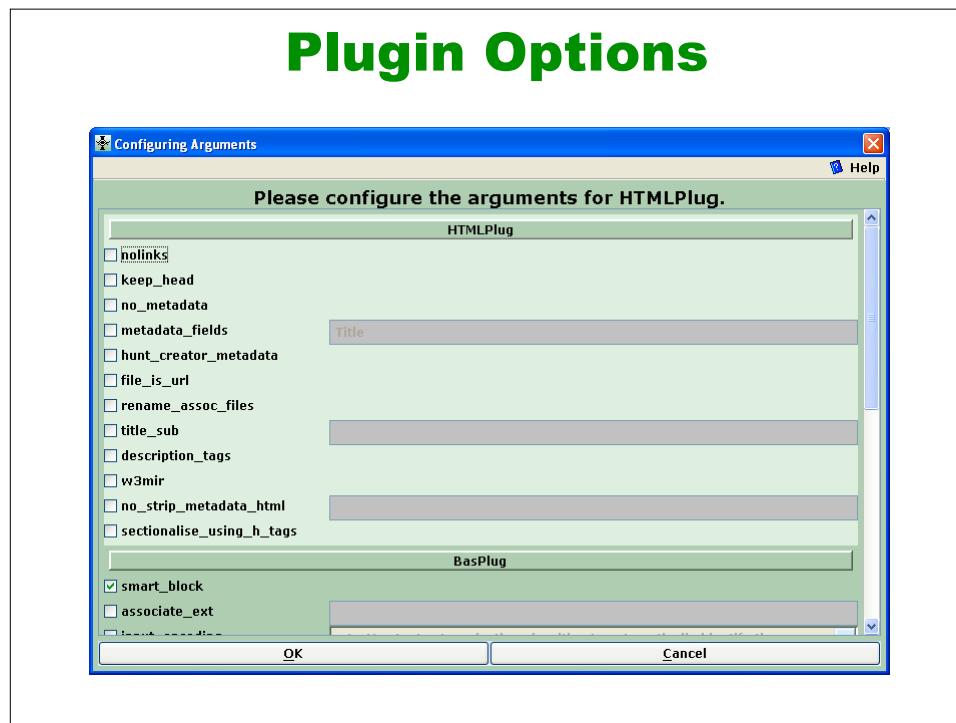
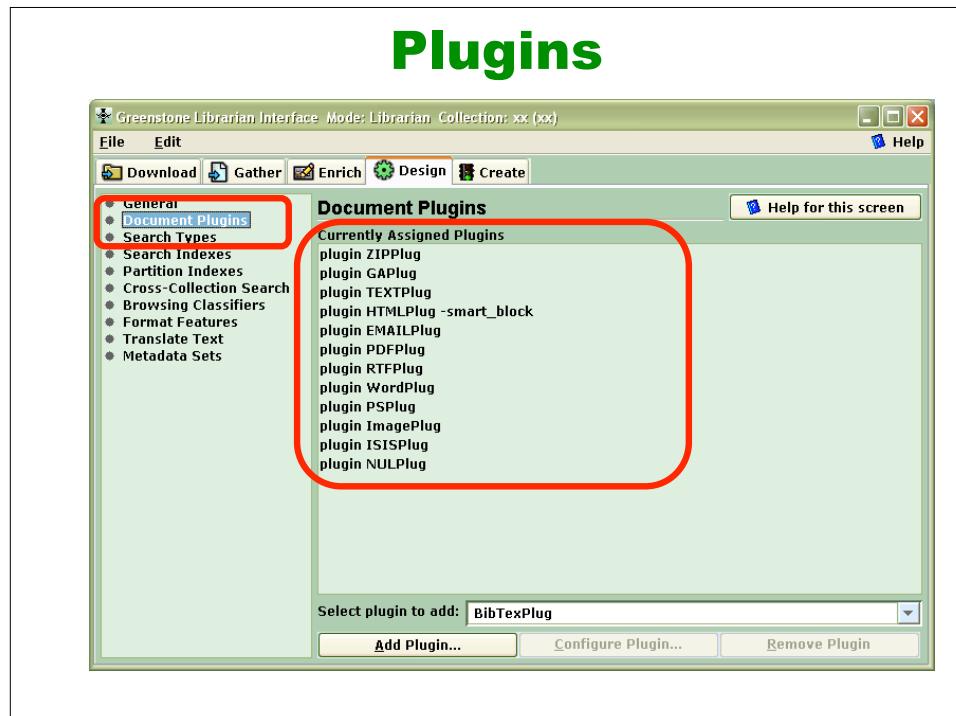
# Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Plugins



The screenshot shows the 'Document Plugins' configuration screen in the Greenstone Librarian Interface. The left sidebar lists various plugin categories, with 'Document Plugins' selected and highlighted by a red box. The main panel displays the 'Currently Assigned Plugins' list, which includes 'plugin GAPPlug' and 'plugin HTMLPlug -description\_tags', also highlighted by a red box. At the bottom, there's a dropdown menu for selecting a plugin to add, and buttons for 'Add Plugin...', 'Configure Plugin...', and 'Remove Plugin'.



## Plugins

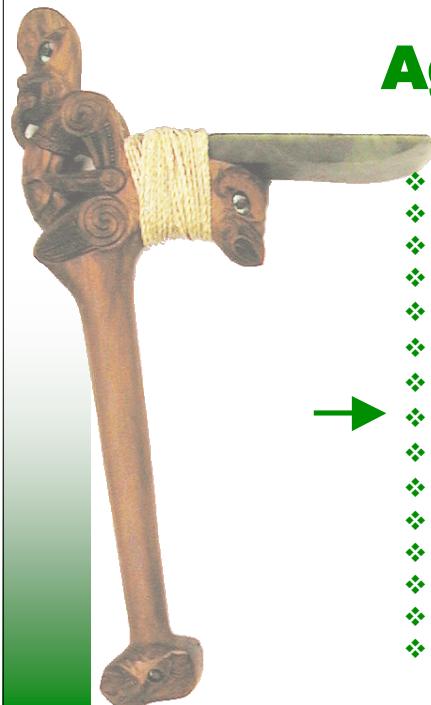
Used by collection-building software to accomplish format-specific parsing of source documents

Plugin pipeline: files are passed to each plugin in turn until one is found that can process it

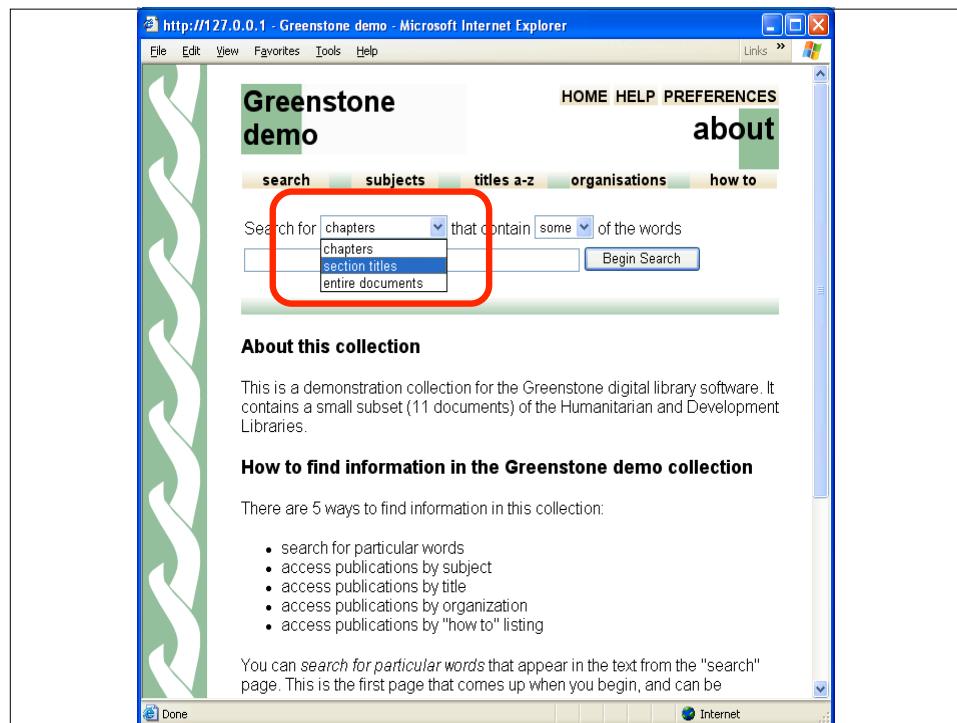
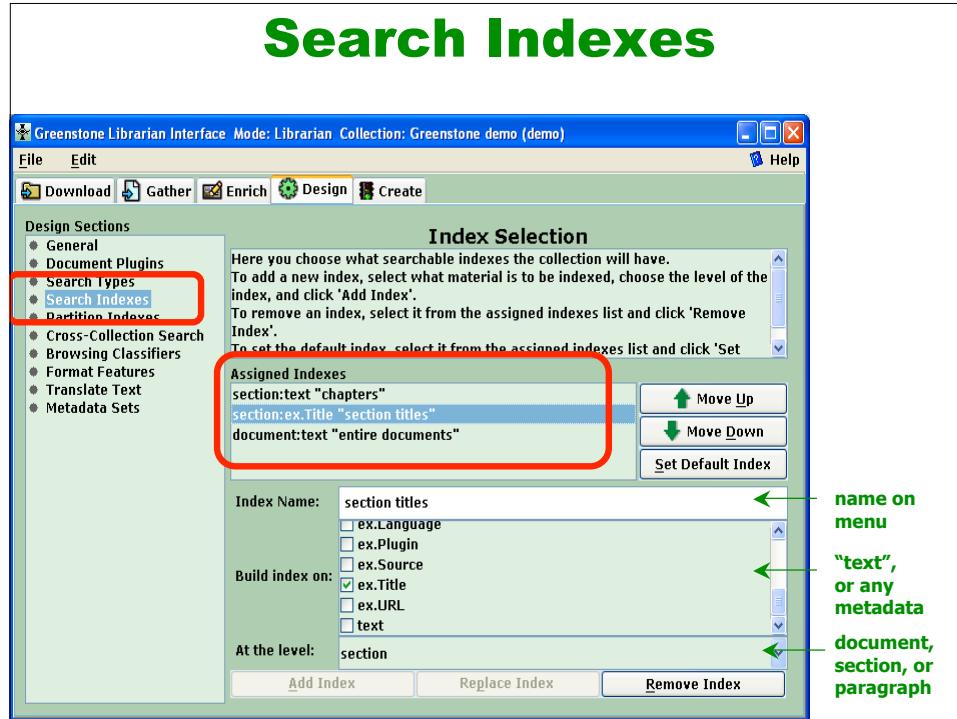
**TEXTPlug**  
**HTMLPlug**  
**EMAILPlug**  
**WORDPlug**  
**RTFPlug**  
**PDFPlug**  
**PSPlug**  
**ImagePlug**  
**PPTPlug**  
**ISISPlug**  
**TCCPlug ...**

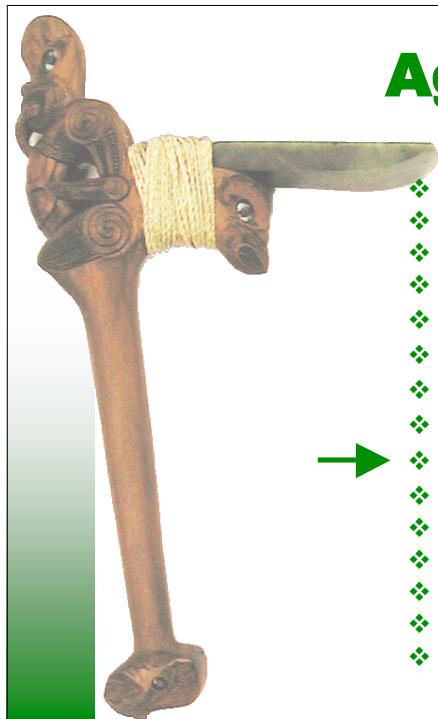
- ❖ GAPPlug processes doc.xml files generated during *import*
- ❖ ArcPlug processes filelist in *archives.inf*
- ❖ RecPlug recurses through a directory structure

## Agenda



- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

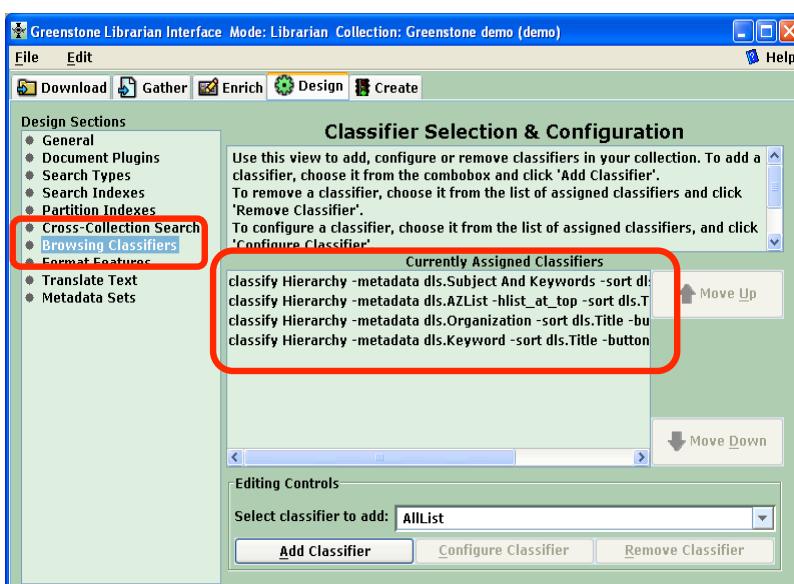




# Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Browsing Classifiers



The screenshot shows the Greenstone Librarian Interface with the title bar "Greenstone Librarian Interface Mode: Librarian Collection: Greenstone demo {demo}". The menu bar includes File, Edit, Help, Download, Gather, Enrich, Design, Create, and a toolbar icon. The main window has a sidebar titled "Design Sections" with options like General, Document Plugins, Search Types, Search Indexes, Partition Indexes, Cross-Collection Search, Browsing Classifiers (which is highlighted with a red box), Format Features, Translate Text, and Metadata Sets. The central area is titled "Classifier Selection & Configuration" with instructions for adding, configuring, or removing classifiers. A list titled "Currently Assigned Classifiers" contains several classifier names, each with "Move Up" and "Move Down" buttons. At the bottom, there are "Editing Controls" for selecting a classifier to add (set to "AllList") and buttons for "Add Classifier", "Configure Classifier", and "Remove Classifier".

The screenshot shows a Microsoft Internet Explorer window displaying the 'Greenstone demo' collection. The title bar reads 'http://127.0.0.1 - Greenstone demo - Microsoft Internet Explorer'. The page has a green header with the title 'Greenstone demo' and navigation links for 'HOME', 'HELP', 'PREFERENCES', and 'about'. A red box highlights the 'subjects' link in the menu. Below the menu is a search bar with dropdowns for 'chapters' and 'some' words, and a 'Begin Search' button. The main content area includes sections for 'About this collection' and 'How to find information in the Greenstone demo collection'. It lists five ways to find information: search for particular words, access publications by subject, access publications by title, access publications by organization, and access publications by "how to" listing. A note at the bottom says you can search for particular words from the 'search' page.

The screenshot shows a Netscape browser window displaying the 'AZList classifier (Title)' interface. The title bar reads 'Titles A-Z - Netscape'. The page has a green header with the title 'Humanity Development Library 2.0' and navigation links for 'HOME', 'HELP', 'PREFERENCES', and 'titles a-z'. A red box highlights the 'titles a-z' link in the menu. Below the menu is a search bar with dropdowns for 'A B C D E F G H I J L M N O P Q R S T U Y W 0-9 Magazines'. The main content area lists various titles under the 'titles a-z' section, such as 'About the international institute of rural reconstruction', 'Above and beyond', 'Abstracts on sustainable agriculture 1992 Gate- GTZ', and a large list of agricultural development titles. A red box highlights the command 'classify AZList -metadata Title' in the left sidebar.

**AZList classifier (Title) for the demo collection**

**demo collection**

**titles a-z**

search subjects titles a-z organization how to

Butterfly Farming in Papua New Guinea  
 The Courier - N°159 - Sept- Oct 1996 Dossier Investing in People Country Reports: Mali , Western Samoa  
 The Courier - N°160 - Nov - Dec 1996 - Dossier Habitat - Country reports: Fiji ,

classify AZList -metadata Title

\*158-July- August 1996 Dossier Communication and the media - Country report Cape Verde  
 Development in practice: Toward Gender Equality  
 Farming snails 1: Learning about snails; Building a pen; Food and shelter plants  
 Farming snails 2: Choosing snails; Care and harvesting; Further improvement  
 Little Known Asian Animals With a Promising Economic Future  
 Managing Tropical Animal Resources - Crocodiles as a Resource for the Tropics  
 Microlivestock - Little-Known Small Animals with a Promising Economic Future  
 The Water Buffalo: New Prospects For An Underutilized Animal (1984)

**AZCompactList classifier (Creator)**

**nzdl publications**

**authors a-z**

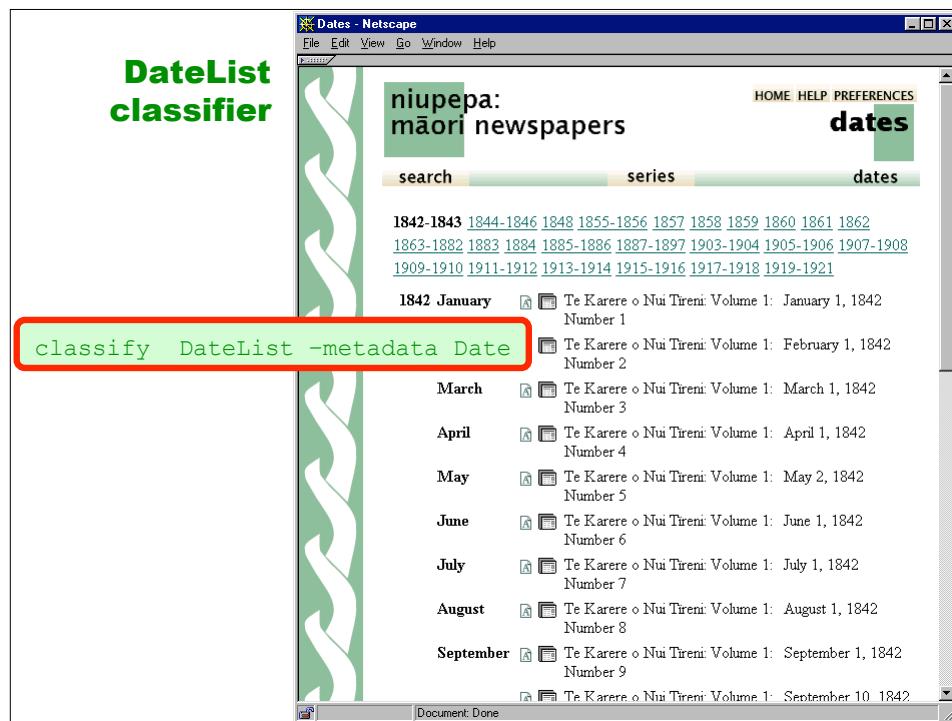
search titles a-z authors a-z dates phrases

A-H I-L P-R Y

Apperley, M. D. (7)  
 Bainbridge, D. (29)

classify AZCompactList -metadata dc.Creator -mingroup 1

Belisle, C. (1)  
 Bell, T. C. (4)  
 Blandford, A. (4)  
 Boddie, S. J. (10)  
 Bray, B. (1)  
 Bray, Z. (1)  
 Britland, M. (1)  
 Buchanan, G. R. (8)  
 Chu, Y. (1)  
 Chu, Y. C. (1)  
 Chui, C. (1)  
 Cockburn, A. (1)  
 Connaway, L. S. (1)



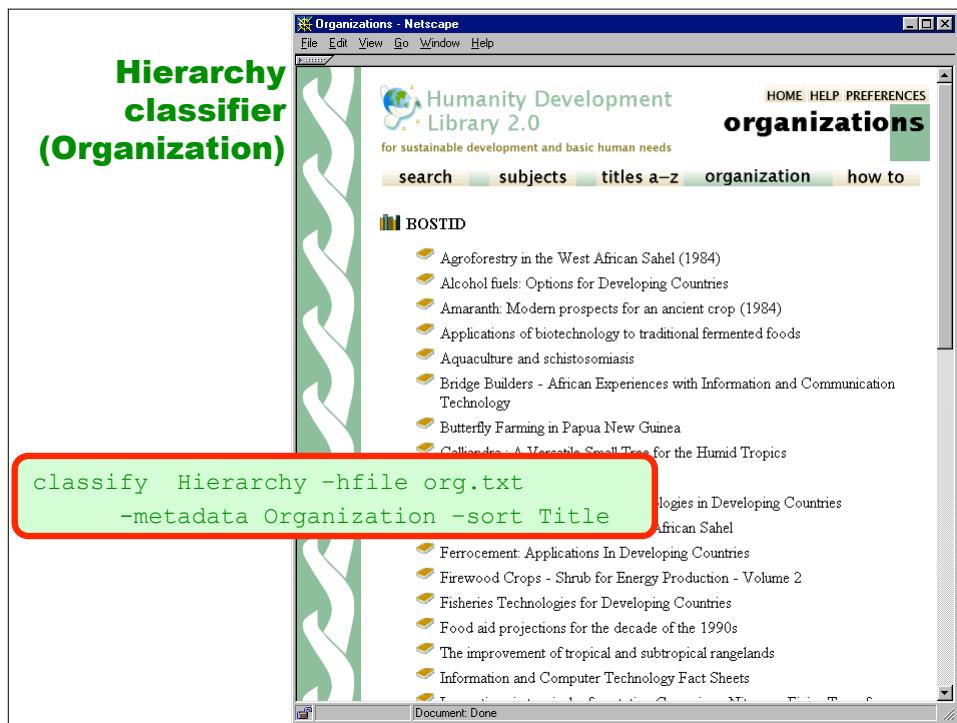
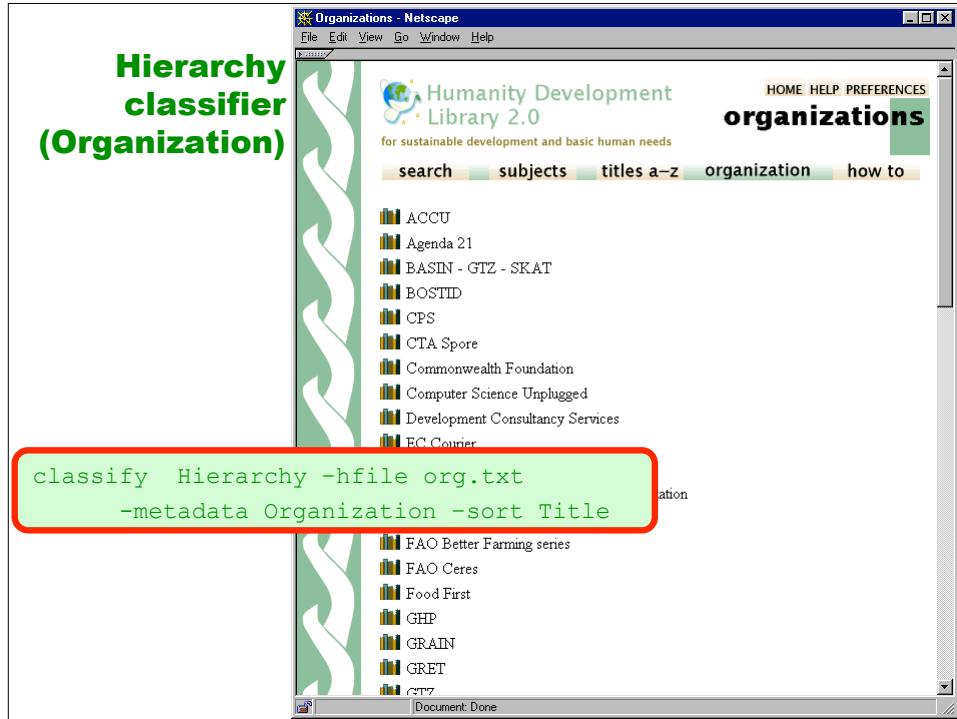
search subjects titles a-z organization how to

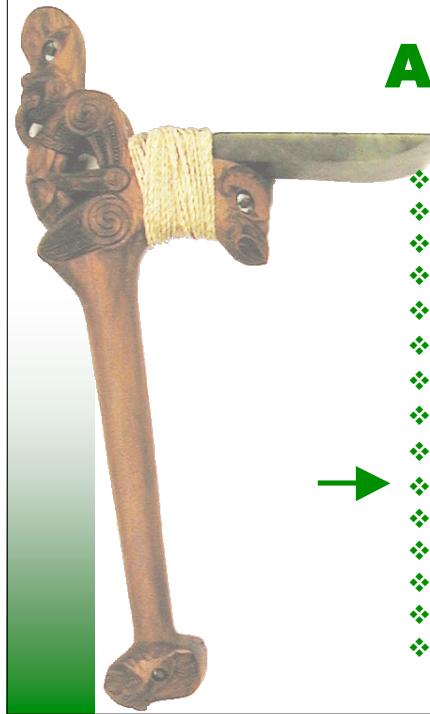
classify Hierarchy -hfile org.txt  
-metadata Organization -sort Title

**etc/org.txt**

identifier (matches *Organization* metadata value)  
position in the hierarchy that the browser implements  
title to be displayed in hierarchy browser

accu	1	ACCU
ag21	2	"Agenda 21"
bgz	3	"BASIN - GTZ - SKAT"
bostid	4	BOSTID
cps	5	CPS
cpas	6	"CTA Spore"
cf	7	"Commonwealth Foundation"
csu	8	"Computer Science Unplugged"
dcs	9	"Development Consultancy Services"
ecc	11	"EC Courier"
ecdg8	12	"EC DG8"
echo	13	"Educational Concerns for Hunger Organization"
fao	14	FAO
faobfs	15	"FAO Better Farming series"
ceres	16	"FAO Ceres"
ff	17	"Food First"

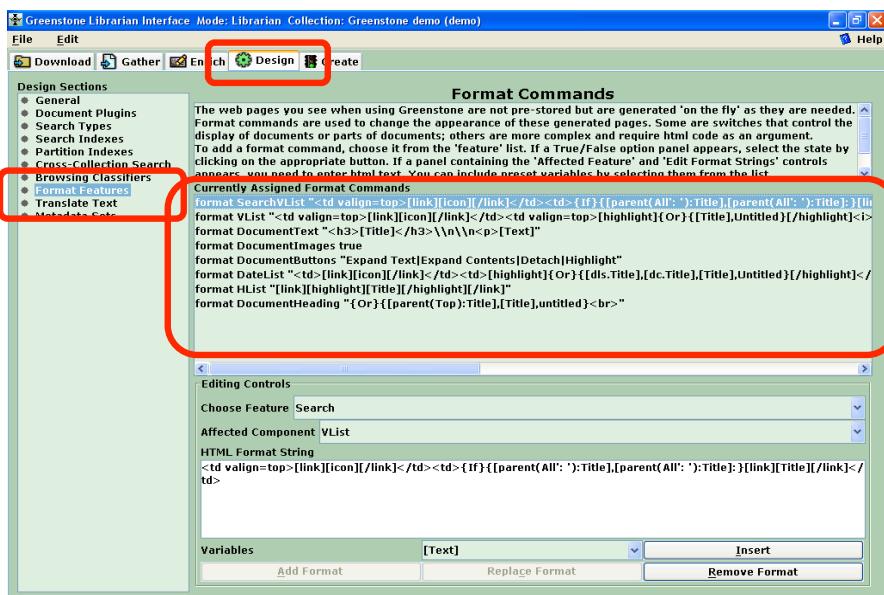




# Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Format Features



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

**Format Commands**

```

format SearchVList "<td valign=top><[link][icon]>/link</td><td>{if}{[parent>All:'':Title],[parent>All:'':Title]}{[if]}<br>
format VList "<td valign=top><[link][icon]>/link</td><td>{highlight}{Or}{[Title],Untitled}</highlight><br>
format DocumentText "<h3>[Title]</h3><br>\n<p>[Text]</p>"
```

```

format DocumentImages true
format DocumentText true
format DocumentHeading true
format DateList "Expand Text|Expand Contents|Detach|Highlight"
format HList "[link][highlight][Title]/highlight[/link]"
format DocumentHeading "{Or}{[parent(Top):Title],[Title],untitled}<br>"
```

**Editing Controls**

Choose Feature: Search  
Affected Component: VList  
HTML Format String:  
<td valign=top><[link][icon]>/link</td><td>{if}{[parent>All:'':Title],[parent>All:'':Title]}{[link][Title]}</link><br>

Variables: [Text] Insert  
Add Format Replace Format Remove Format

## Format Statements

- ❖ **VList** – vertical lists: search, classifiers, document table of contents
- ❖ **HList** – horizontal lists: classifiers
- ❖ **SearchVList, CL1VList** – individual vertical lists
- ❖ **DocumentImages/DocumentHeading/ DocumentContents** – document header display
- ❖ **DocumentText** – document display

Search VList

The screenshot shows a Microsoft Internet Explorer window with the URL <http://www.nzdl.org - results 1 - 20 for query: food - Microsoft Internet Explorer>. The page title is "development library subset". On the left, there's a sidebar with "HOME HELP PREFERENCES" and a "search" button highlighted with a red box. Below it are links for "subjects", "titles a-z", "organisations", and "how to". The main content area has a search bar with "Search for chapters that contain some of the words" dropdown set to "food" and a "Begin Search" button. A red box highlights the search results section titled "results". The results list includes:

- Word count: food: 4571
- More than 50 documents matched the query.
- Ideas for action: Save, recycle and do not pollute: Pesticides: [Food safety practices](#)
- Recording and using indigenous knowledge: A manual: Part 5 - Question guides: [Food and nutrition](#)
- Homemaking handbook for village workers in many countries: Section II - What you will teach: Food and nutrition: [Food needs vary](#)
- Homemaking handbook for village workers in many countries: Section II - What you will teach: Food and nutrition: [Your government is concerned](#)
- Homemaking handbook for village workers in many countries: Section II - What you will teach: Food and nutrition: [What is nutrition?](#)
- Basic Concepts in Environment, Agriculture and Natural Resources Management: An information kit: Ecological basics: [The food chain](#)
- Traditional field crops: [Introduction](#)

CL1  
VList

The screenshot shows a Microsoft Internet Explorer window with the URL <http://www.nzdl.org - Subjects - Microsoft Internet Explorer>. The page title is "subjects". A red box highlights the "subjects" tab in the navigation bar. Below the navigation bar is a vertical sidebar with a wavy pattern. The main content area displays a list of subjects, each preceded by a small icon. A large red circle highlights the entire list of subjects. The subjects listed are:

- Industry, Manufacture and Services
- Energy , Renewable energies, Household energy
- Water, Sanitation and Sewage Disposal
- Ecology, Biodiversity, Natural resources, Environment, Waste Disposal
- Plant and Crop Production, Protection, and Processing
- Fishery and Aquaculture
- Politics, Public administration, Law and Legislation, Peace, Human Rights
- Health, Nutrition, Medicine
- Sustainable Development, International cooperation, Projects; NGO, Organizations, Poverty and Hunger Alleviation, Basic Human Needs
- Society, Culture, Community, Woman, Youth, Population
- Education and Training
- Communication, Information and Documentation
- Agriculture and Food Processing
- Animal Husbandry and Animal Product Processing
- Settlements, Housing, Building - Infrastructure Construction (Roads etc)
- Development Periodicals and Magazines
- Business, Micro-enterprises, Management, Marketing, Finance, Microfinance

CL2  
HList

CL2  
VList

The screenshot shows a Microsoft Internet Explorer window with the URL <http://www.nzdl.org - Titles A-Z - Microsoft Internet Explorer>. The page title is "titles a-z". A red box highlights the "titles a-z" tab in the navigation bar. Below the navigation bar is a vertical sidebar with a wavy pattern. The main content area displays a list of titles, each preceded by a small icon. A large red circle highlights the entire list of titles. The titles listed are:

- A-B C-D-E F-G-H-I-J-K L-M-N-O-P Q-R-S T-U-V-W-X-Y-Z Periodicals
- 1st Session of the Commission on the Status of Women
- source ref: [ng04se.htm](#)
- Above and beyond
- source ref: [m0052e.htm](#)
- Accounting for the microbusiness a teaching manual - A Handbook for Volunteers
- source ref: [r0023e.htm](#)
- Agroforestry in-service training: A training aid for Asia & the Pacific Islands
- source ref: [t0016e.htm](#)
- Alcoholism: prevention and cure
- source ref: [cp01ae.htm](#)
- Alternative techniques - For teaching about HIV/AIDS in the classroom
- source ref: [r0086e.htm](#)
- Animal Disease control and treatment
- source ref: [ii15ae.htm](#)
- Animal traction
- source ref: [m0012e.htm](#)
- Animation skills
- source ref: [m0055e.htm](#)

Document Images

Document Buttons

Document Text

**Alcoholism: prevention and cure**

source ref: cp01ae.htm

- Introduction
- Part 1: General Knowledge of Alcoholism
- Part 2: The flip-chart and slide show "alcoholism" and how to use it
- Part 3: Suggested didactic material
- Part 6: Slideshow: The damaging effects of alcohol
- Part 7: Slideshow: Preventing alcoholism
- Part 8: List of the pictures
- Part 9: Anecdotes, simple health education talks

by Courtejoie J., Pierre B.

Bureau of Study and research for the promotion of Health, PO Box 1800 Kangu -Mayombe, Republic of Zaire or preferably / C/o Avenue des Hortesias, B- 1950 Kraainem - BELGIUM

copyright No. 124 in the Republic of Zaire - 1983

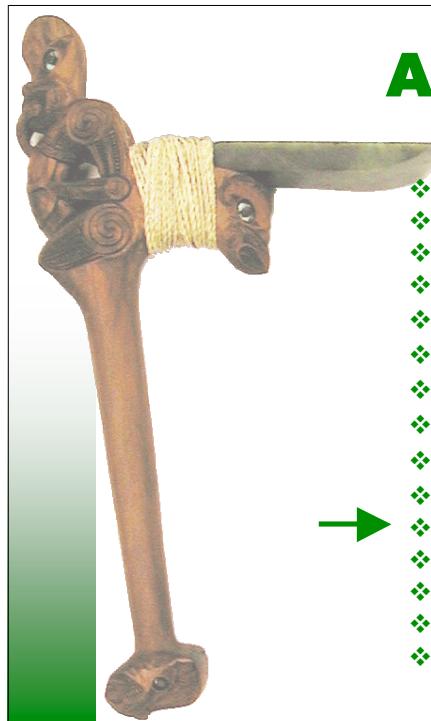
Document Heading

Document Buttons

Document Text

Be a wolf! / by Brad Strickland.

LDR 01497cam 2200385 a 4500  
 001 1184190  
 005 19981015160255.6  
 008 970627s1997 txua c 000 1 eng d  
 035 \_9 (DLC) 97072875  
 906 \_a 7  
 \_b cbc  
 \_c copycat  
 \_d 2  
 \_e opcn  
 \_f 19  
 \_g y-genca1tg  
 955 \_a pb04 02-19-98 to hlcd; lb15 04-15-98; lb14 05-18-98; lb15 05-19-98;  
 \_a pn09 06-27-97; OCLC import pb04 02-19-98  
 010 \_a 97072875  
 020 \_a 1570641951 (pbk.) :  
 \_c \$3.99  
 035 \_a (OCoLC)37015253  
 040 \_a IHS  
 \_c IHS  
 \_d DLC  
 \_a lccopycat  
 \_a lcac  
 050 00 \_a P27.S9166

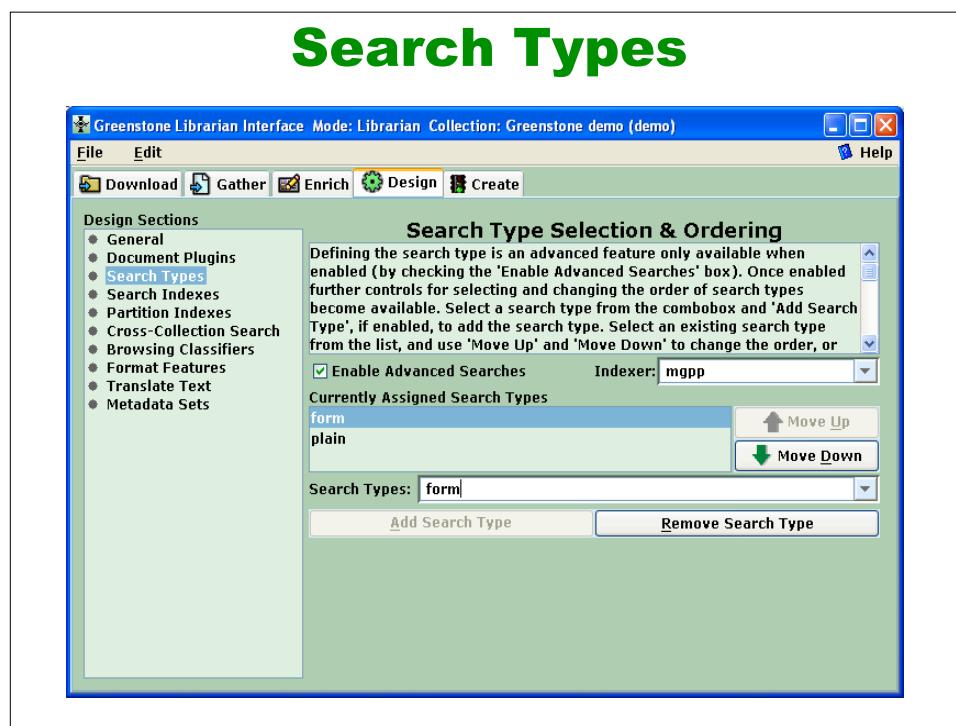
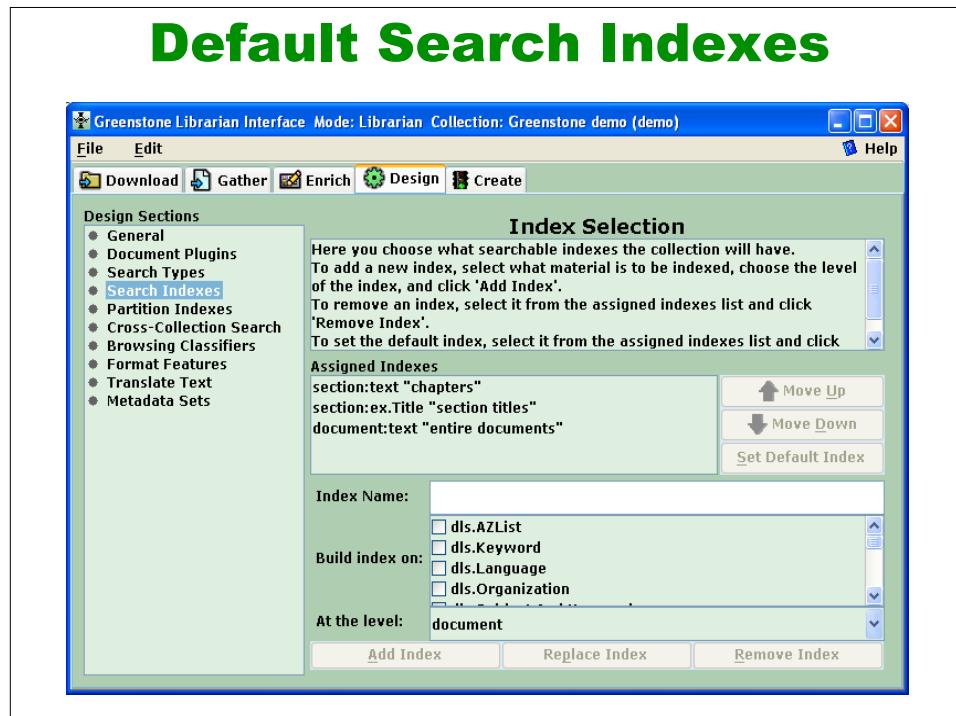


## Agenda

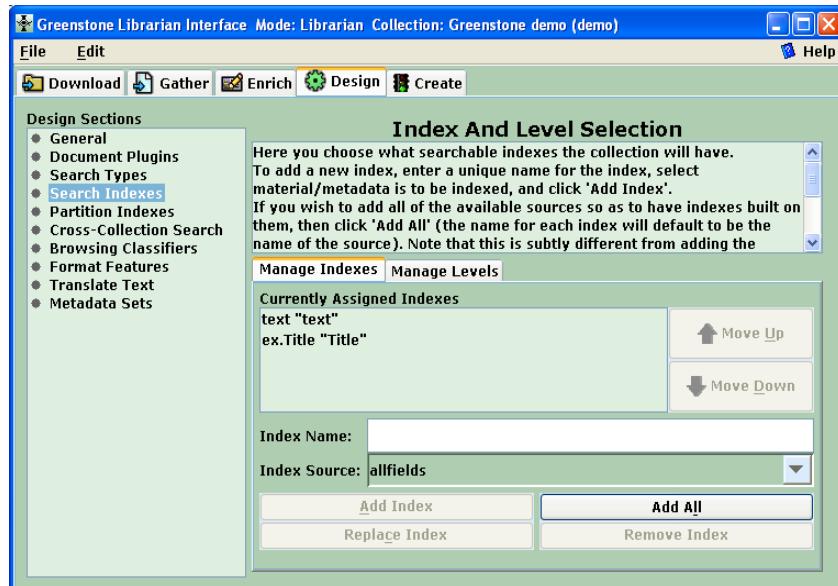
- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Search Types

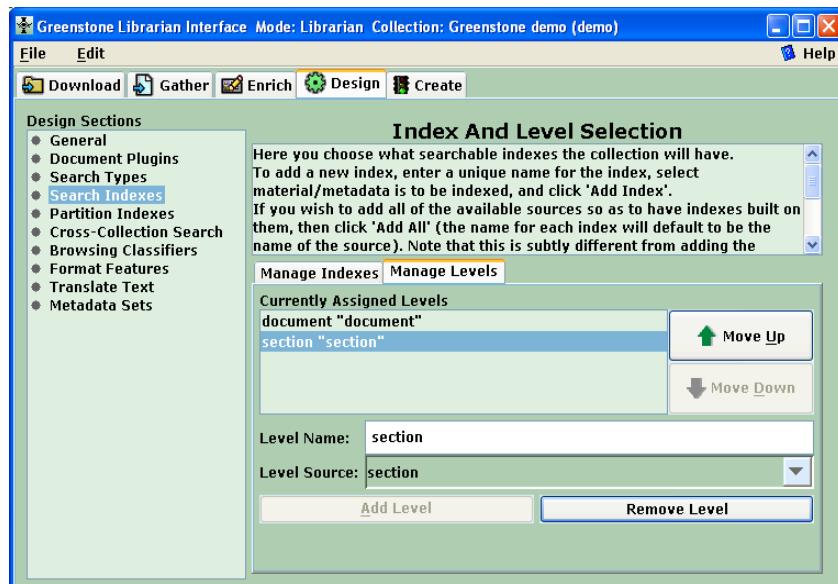
- ❖ “Enable search types” – allows fielded searching
- ❖ MGPP/Lucene
- ❖ Form/plain – can switch on preferences
- ❖ Different index specification



## Form Search Indexes: Indexes



## Form Search Indexes: Levels



## Plain search

The screenshot shows the 'Search preferences' section of the COLT Bibliography Preferences dialog. The 'Plain search' radio button is selected. Under 'Query mode', the 'simple query mode' radio button is selected, which is described as 'allows boolean searching using !, &, & l, and parentheses'. Under 'Query box size', the 'large query box' radio button is selected. At the bottom, it says 'Form type: simple'.

The screenshot shows the search page of the COLT Bibliography collection. It features a search bar with dropdown menus for 'titles a-z', 'authors a-z', 'dates', and 'phrases'. Below the search bar is a text input field with placeholder text 'Search for [full records] which contain [some] of the words'. A 'Begin Search' button is at the bottom.

## Simple form search

The screenshot shows the 'Search preferences' section of the COLT Bibliography Preferences dialog. The 'Form search' radio button is selected. Under 'Query mode', the 'simple query mode' radio button is selected. Under 'Query box size', the 'large query box' radio button is selected. At the bottom, it says 'Form type: simple'.

The screenshot shows the search page of the COLT Bibliography collection. It features a search bar with dropdown menus for 'titles a-z', 'authors a-z', 'dates', and 'phrases'. Below the search bar is a text input field with placeholder text 'Search for [some] of Word or phrase'. Below this is a '... in field' dropdown menu containing 'full records', 'Creator', 'BookConfOnly', and 'Source'. A 'Clear Form' button is on the left and a 'Begin Search' button is on the right.

## Advanced form search

**COLT Bibliography: Preferences**

**Search preferences**

Plain search:

- Query mode: simple query mode (selected)
- advanced query mode (allows boolean searching using !, &, |, and parentheses)

Query box size:

- regular query box (selected)
- large query box

**Form search:**

Form type: simple (radio button)

advanced (radio button) **(selected)**

with 4 fields

**Search page**

**bibliography collection**

**search** **titles a-z** **authors a-z** **dates** **phrases**

Search and display results in ranked order

Word or phrase (fold, stem) ... in field

and Creator  
and Title  
and Date  
and Source

Clear Form Begin Search

Or enter a query directly:  Run Query

## Advanced form search

results 1 – 1 for query: [gori#i] :CR & [prob...

**bibliography collection**

**search** **titles a-z** **authors a-z** **dates** **phrases**

Search and display results in ranked order

Word or phrase (fold, stem) ... in field

and gori problem  
and 1992

Clear Form Begin Search

Or enter a query directly:  **[gori#i] :CR & [problem#i] :TI & [1992] :DA** Run Query

results

Word count: gori: 2, problem: 35, 1992: 181  
1 document matched the query.

On the Problem of Local Minima in Backpropagation - Marco Gori and Alberto Tesi - 1992

### ❖ Fielded searching

[terms]:Field  
Specifies searching within a particular field

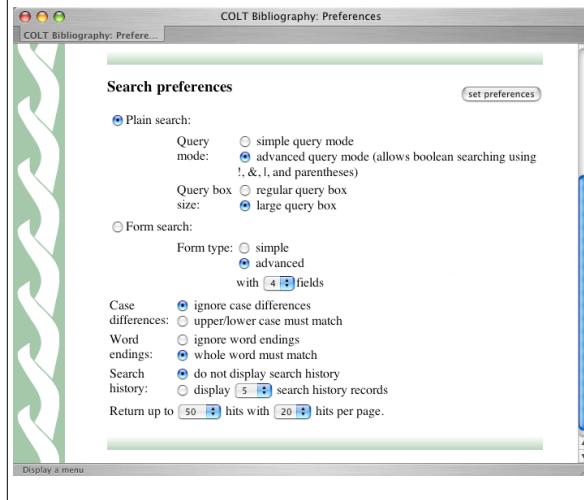
### ❖ Term modifiers

#iclus /x  
Stemming, casefolding and weighting

### ❖ Fields

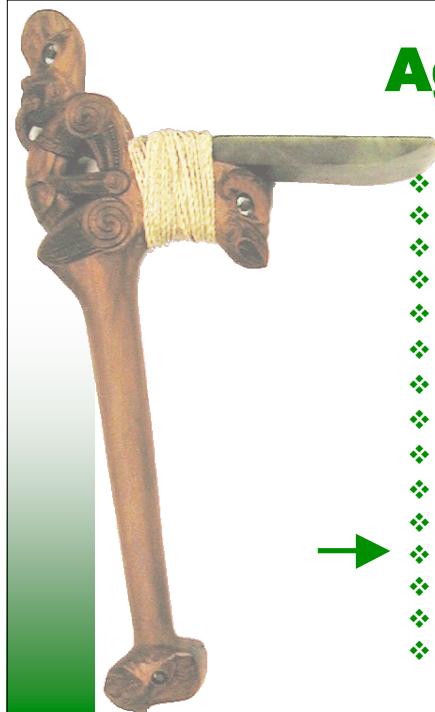
TI title  
CR creator  
SU subject  
TX text  
...

## Advanced plain search



- ❖ *Regular vs large query box*
- ❖ *Boolean operators*  
 & AND  
 | OR  
 ! NOT
- ❖ *Phrase searching ("...")*
- ❖ *Case folding*
- ❖ *Stemming (word endings)*
- ❖ *Search history*
- ❖ *Number of returned results*

## Agenda

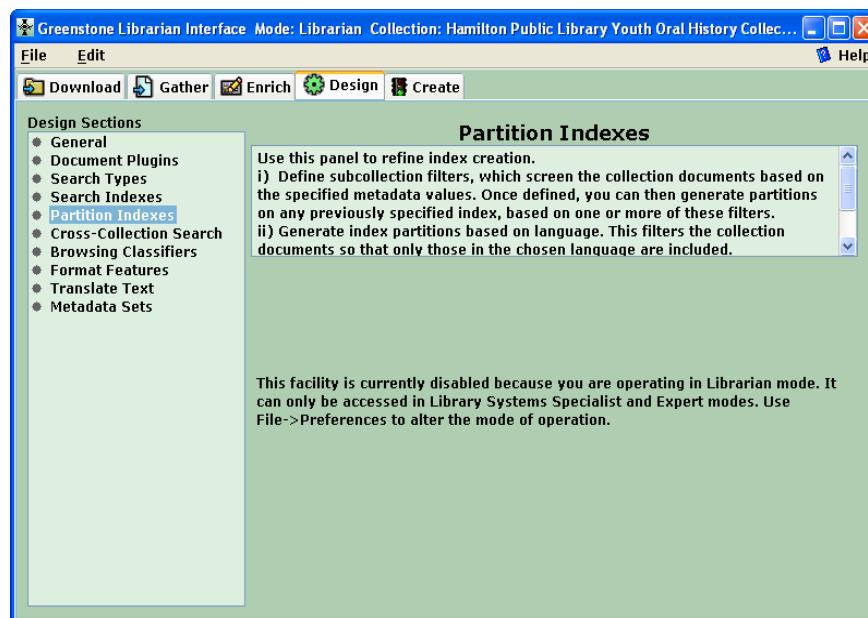


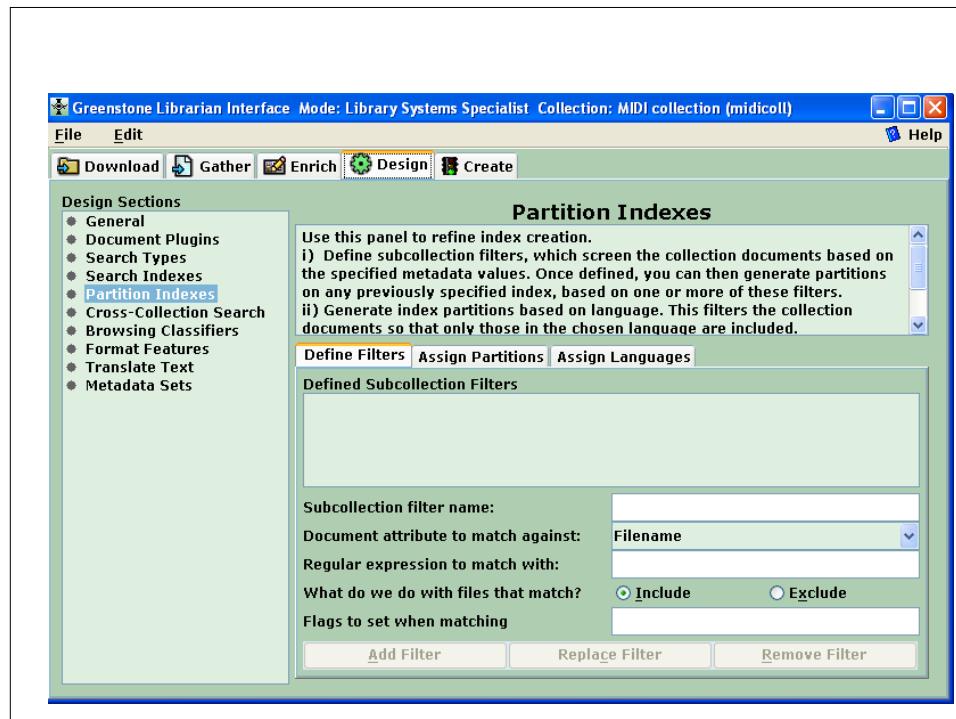
- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Partition Indexes

- ❖ Searching within a single collection
- ❖ The indexes are split based on subcollection definitions
  - E.g. language, file types, metadata
- ❖ Browsing classifiers contain all the documents
- ❖ Library Systems Specialist mode

## Partition Indexes





## Language partitions

The screenshot shows the 'HOME HELP PREFERENCES' page of the Greenstone Documentation Collection. At the top, there are links for 'greenstone documentation', 'about', 'titles a-z', and 'browse'. Below this is a search bar with dropdown menus for 'Search for' (set to 'documents') and 'in' (set to 'all'), followed by a dropdown for 'language that contain' (set to 'some') and a text input field. A 'Begin Search' button is next to the search fields. A dropdown menu for 'About this collection' lists 'english', 'french', 'spanish', 'russian', 'vietnamese', and 'kazakh', with 'english' currently selected. Below this, a paragraph describes the collection's contents. A section titled 'How to find information in the Greenstone Documentation Collection collection' lists three ways to find information: searching for words, accessing publications by title, and browsing publications. A note at the bottom says 'You can search for additional words that appear in the text from the Research page'.

## Language metadata

The screenshot shows the 'Language metadata' section of the Greenstone Librarian Interface. On the left, a tree view of the collection structure shows a 'fr' folder under the 'english' folder. The right panel displays a table of metadata elements:

Element	Value
gs.Title	
gs.Category	Manuals French
gs.TitleExtra	
gs.Language	fr

A red box highlights the 'gs.Language' row. Below the table, a message says 'No Metadata Element Selected'.

## Language partitions

The screenshot shows the 'Partition Indexes' section of the Greenstone Librarian Interface. On the left, a sidebar lists 'Design Sections' with 'Partition Indexes' selected. The main panel has a heading 'Partition Indexes' and contains the following text:

Use this panel to refine index creation.  
 i) Define subcollection filters, which screen the collection documents based on the specified metadata values. Once defined, you can then generate partitions on any previously specified index, based on one or more of these filters.  
 ii) Generate index partitions based on language. This filters the collection documents so that only those in the chosen language are included.

Below this is a table titled 'Assigned Languages' with a red box around it. The table lists:

ENGLISH
FRENCH
SPANISH
RUSSIAN
VIETNAMESE

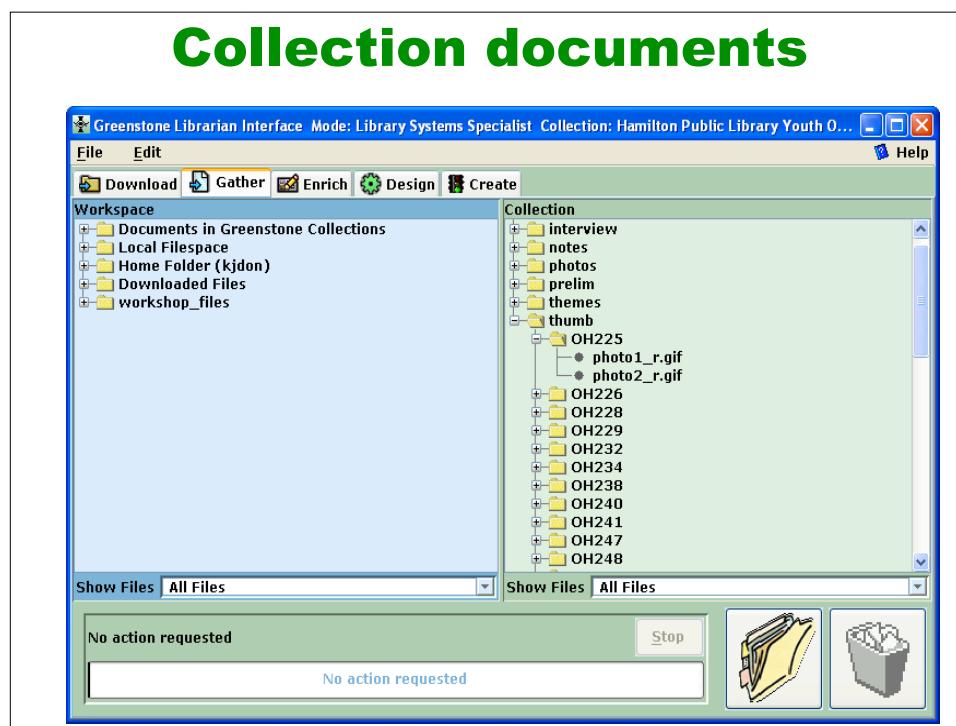
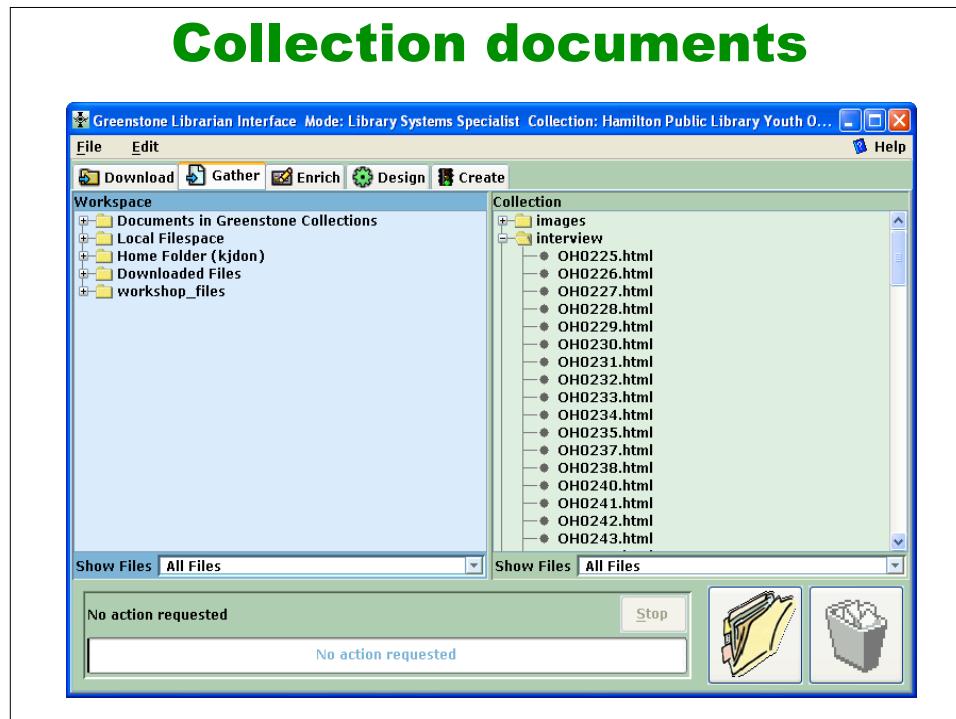
At the bottom, there is a dropdown menu 'Language to add: ABKHAZIAN' and buttons for 'Add Language' and 'Remove Language'.

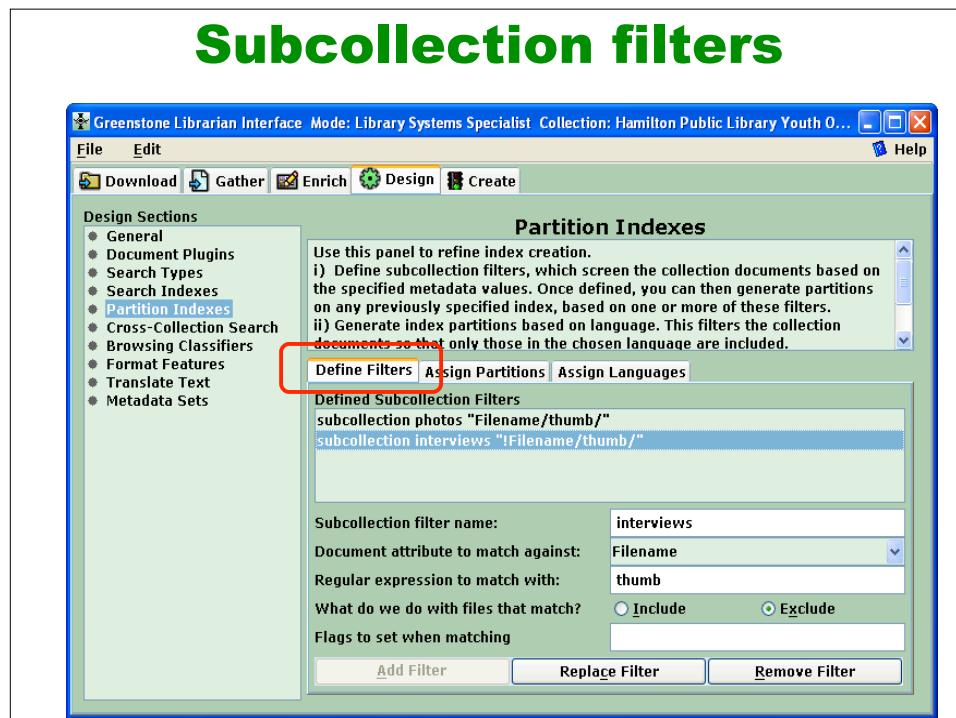
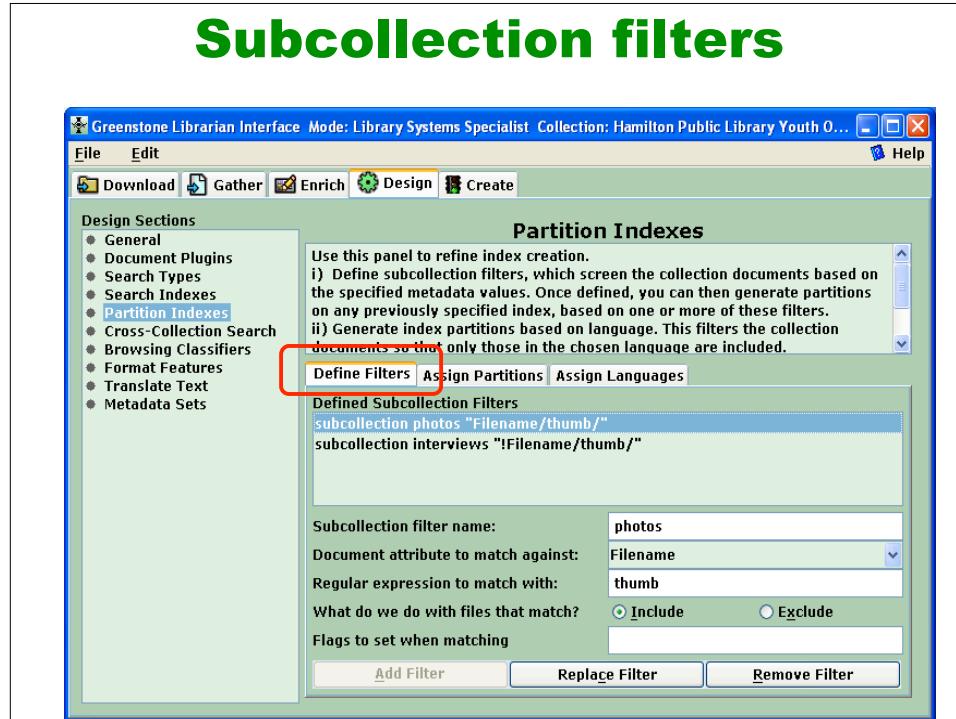
## Document type partitions

The screenshot shows a Microsoft Internet Explorer window displaying the Hamilton Public Library Youth Oral History Collection. The title bar reads "Hamilton Public Library Youth Oral History Collection - Microsoft Internet Explorer". The page features a green decorative border on the left. At the top, there is a logo of two people dancing, followed by the text "Hamilton Public Library" and "Youth Oral History". A navigation menu includes links for "HOME HELP", "PREFERENCE", "about", "search", "titles a-z", and "authors a-z". Below the menu is a search form with dropdown menus for "Search for" (set to "documents") and "of" (set to "all documents"), a dropdown for "that contain" (set to "some"), and a text input field containing "titles", "documents", and "document sections". A "Begin Search" button is present. The main content area contains sections titled "About this collection" and "How to find information in the Hamilton Public Library Youth Oral History Collection collection". The "About this collection" section includes a paragraph about the interview project and a link to the "Hamilton Public Library". The "How to find information" section describes three ways to search: titles, authors, and documents.

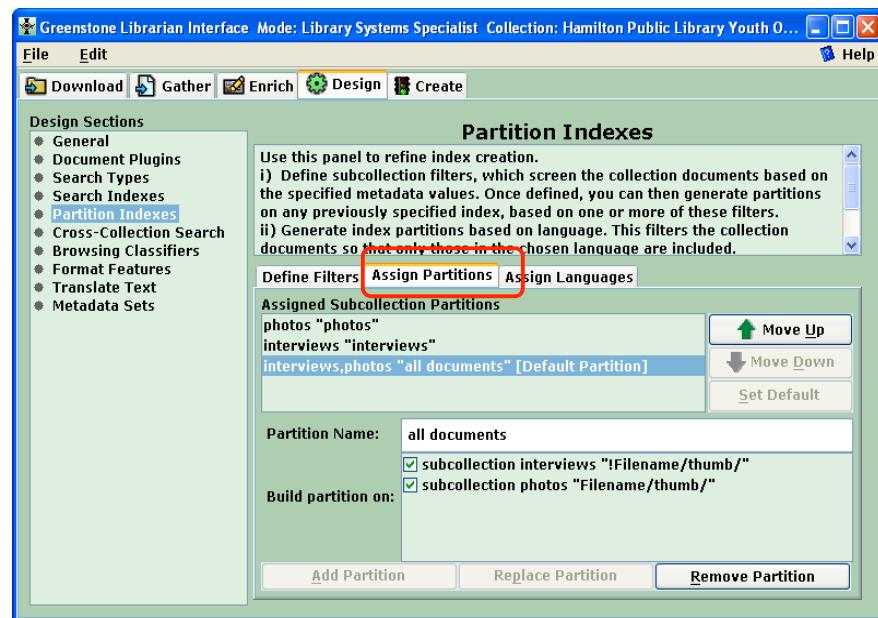
## Document type partitions

This screenshot is identical to the one above, showing the Hamilton Public Library Youth Oral History Collection website in Microsoft Internet Explorer. However, a red rectangular box highlights the dropdown menu in the search form. The dropdown menu lists "titles", "photos", "interviews", and "all documents", with "all documents" currently selected. The rest of the page content, including the "About this collection" and "How to find information" sections, remains the same.

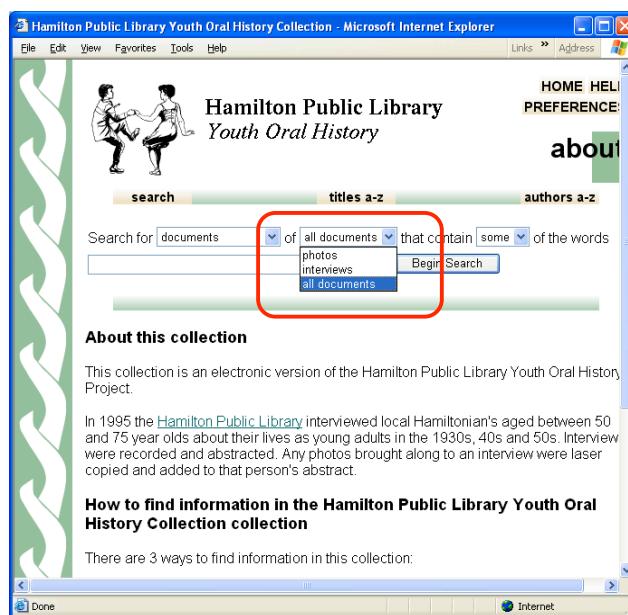




## Assign partitions



## Document type partitions

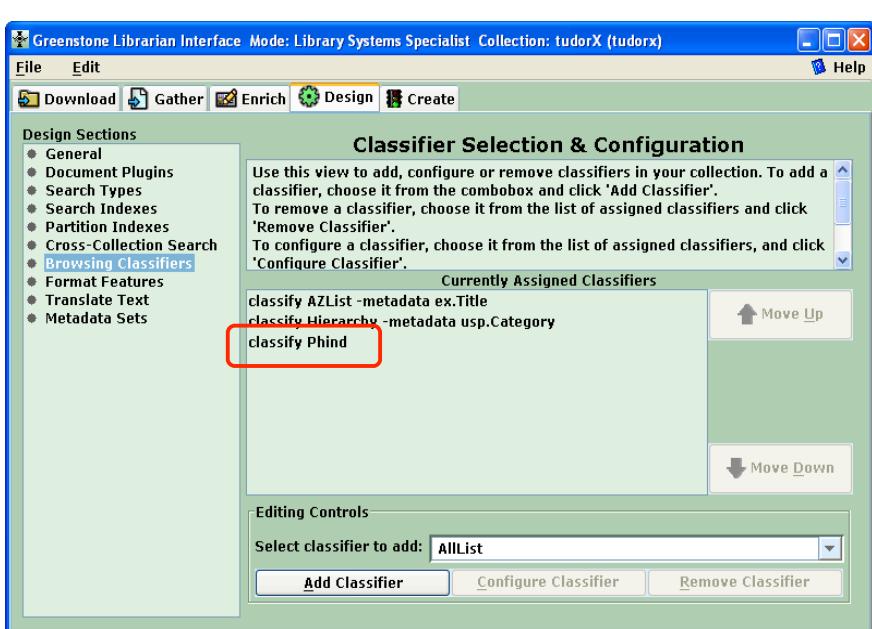




# Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## Phind Phrase Index



The screenshot shows the 'Classifier Selection & Configuration' window in the Greenstone Librarian Interface. The window has a title bar 'Greenstone Librarian Interface Mode: Library Systems Specialist Collection: tudorX (tudorx)' and a menu bar with File, Edit, Help, Download, Gather, Enrich, Design, Create.

The left sidebar lists 'Design Sections' including General, Document Plugins, Search Types, Search Indexes, Partition Indexes, Cross-Collection Search, Browsing Classifiers (which is selected), Format Features, Translate Text, and Metadata Sets.

The main area is titled 'Classifier Selection & Configuration'. It contains instructions for adding, configuring, or removing classifiers. A list of 'Currently Assigned Classifiers' includes 'classify AZList -metadata ex.Title', 'classify Hierarchy -metadata usp.Category', and 'classify Phind' (which is highlighted with a red box).

At the bottom, there are 'Editing Controls' with a dropdown 'Select classifier to add: AllList' and buttons for 'Add Classifier', 'Configure Classifier', and 'Remove Classifier'.

PHIND

### Hierarchical index of phrases

- ❖ **What's in this collection?**
- ❖ **Is it any good?**
- ❖ **What coverage for topic X?**
- ❖ **My query returned too much/little, what now?**

food and agriculture organization  
HOME HELP PREFERENCES  
phrases

search titles a-z language phrases

Search for locust Previous Next

Locust (first 10 of 157 phrases, first 10 of 312 documents)	docs	freq
DesertLocust	225	672
Locust numbers	45	137
Locust control	48	103
Locust situation	78	96
MigratoryLocust	35	87
Locust information	42	45
Locust activity	28	42
Locust operations	39	40
Locust survey	26	36
Locust training	27	29

Get more phrases

FAO on the Internet (1998) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

food and agriculture organization

phrases

search titles a-z language phrases

Search for locust Previous Next

Locust (first 10 of 157 phrases, first 10 of 312 documents) docs freq

	docs	freq
Desert Locust	225	672
Locust numbers	45	137
Locust control	48	103
Locust situation	78	96
Migratory Locust	35	67
Locust information	42	45
Locust activity	28	42
Locust operations	39	40
Locust survey	26	36
Locust training	27	29

Get more phrases

Desert Locust (first 10 of 101 phrases, first 10 of 225 documents) docs freq

	docs	freq
Desert Locust Bulletin	51	105
FAO Desert Locust	46	73
Desert Locust situation	48	50
Desert Locust information	40	41
Desert Locust control	27	39
Desert Locust activity	26	26
Desert Locust survey	15	23
Desert Locust breeding	22	22
Desert Locust controlling	17	20

controlling the Desert Locust 10 20

Get more phrases

FAO on the Internet (1998) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

food and agriculture organization

phrases

search titles a-z language phrases

Search for locust Previous Next

Locust (first 10 of 157 phrases, first 10 of 312 documents) docs freq

	docs	freq
Desert Locust	225	672
Locust numbers	45	137
Locust control	48	103
Locust situation	78	96
Migratory Locust	35	67
Locust information	42	45
Locust activity	28	42
Locust operations	39	40
Locust survey	26	36
Locust training	27	29

Get more phrases

Desert Locust (first 10 of 101 phrases, first 10 of 225 documents) docs freq

	docs	freq
Desert Locust breeding	17	20
controlling the Desert Locust	10	20

Get more phrases

Desert Locust Information Service of FAO: Locust FAQs 37

LOCUST CONTROL SITUATION: MEASURES TAKEN AND FURTHER ACTION 26

AGREEMENT FOR CONTROLLING THE DESERT LOCUST IN THE EASTERN REGION 16

AGREEMENT FOR THE ESTABLISHMENT OF A COMMISSION FOR 14

untitled body 14

body 11

Desert Locust Information Service of FAO: Locust books 10

body 9

Desert Locust Information Service of FAO: Locust FAQs - Microsoft Internet Explorer

File Edit View Favorites Tools Help

HOME HELP PREFERENCES  
search titles a-z language phrases

DESERT LOCUST INFORMATION SERVICE  
of the MIGRATORY PESTS GROUP

LOCUST

Locust Mapper News Archives Plagues Activities Books FAQ

Frequently Asked Questions (FAQs) about Desert Locusts

- [What is the difference between locusts and grasshoppers?](#)
- [What is a Desert Locust?](#)
- [What countries are affected by the Desert Locust?](#)
- [Do Desert Locust plagues occur with any regularity?](#)
- [How long does a Desert Locust live?](#)
- [How many eggs does a Desert Locust female produce?](#)
- [How far and how fast can Desert Locusts migrate?](#)
- [How big are swarms and how many locusts are there in a swarm?](#)
- [What percentage of the Desert Locust's exoskeleton is chitin?](#)
- [How much food can a Desert Locust eat?](#)
- [What is the relationship between locusts and ecology?](#)
- [Why do locusts change their behaviour?](#)
- [Are there other important species of locusts?](#)
- [Can locusts hurt humans?](#)
- [How can locusts be controlled?](#)
- [Who carries out locust control operations?](#)
- [Are there any non-chemical ways to kill locusts?](#)
- [Can locusts be detected by satellites?](#)
- [Why are Desert Locust so difficult to control?](#)
- [Do people really eat locusts?](#)
- [What is a Desert Locust composed of?](#)

FAO on the Internet (1998) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

HOME HELP PREFERENCES

phrases

food and agriculture organization

search titles a-z language phrases

Search for poisson Previous Next

FAO on the Internet (1998) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

food and agriculture organization

phrases

search titles a-z language phrases

Search for poisson Previous Next

poisson (first 10 of 11 phrases, first 10 of 25 documents) docs freq

	docs	freq
du poisson	17	57
poisson en Afrique	3	15
poisson et	6	12
de poisson	10	10
poisson trait	3	8
le poisson	4	4
sattaquent au poisson trait	3	3
poisson frais	3	3
en poisson	2	3
poisson dans	2	2

Get more phrases

FAO on the Internet (1998) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

food and agriculture organization

phrases

search titles a-z language phrases

Search for poisson Previous Next

poisson (first 10 of 11 phrases, first 10 of 25 documents) docs freq

	docs	freq
du poisson	17	57
poisson en Afrique	3	15
poisson et	6	12
de poisson	10	10
poisson trait	3	8
le poisson	4	4
sattaquent au poisson trait	3	3
poisson frais	3	3
en poisson	2	3
poisson dans	2	2

Get more phrases

du poisson (10 phrases, first 10 of 17 documents) docs freq

	docs	freq
technologie du poisson	6	16
du poisson en Afrique	3	14
du poisson et	5	11
commercialisation du poisson	7	11
Commerce du poisson	5	9
transport routier du poisson et des produits de la p	3	3
Traitement du poisson	3	3
n et destination de l'abondance du poisson	1	2
la technologie du poisson hach	2	2
et de l'inspection du poisson	2	2

FAO Fisheries Department list of Publications and Documents 10

**FAO on the Internet (1998) - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

HOME HELP PREFERENCES

**phrases**

search titles a-z language phrases

Search for poisson Previous Next

du poisson (10 phrases, first 10 of 17 documents)

	docs	freq
technologie du poisson	6	16
du poisson en Afrique	3	14
du poisson et	5	11
commercialisation du poisson	7	11
Commerce du poisson	5	9
transport routier du poisson et des produits de la p	3	3
Traitement du poisson	3	3
n et destination de l'abondance du poisson	1	2
la technologie du poisson hach	2	2
et de l'inspection du poisson	2	2
FAO Fisheries Department list of Publications and Documents	10	

commercialisation du poisson (3 phrases, 7 documents)

	docs	freq
Traitement et commercialisation du poisson	3	4
la commercialisation du poisson trait	3	3
de commercialisation du poisson	3	3

produced by FIU/FAO Documents techniques sur les pêches produits par FI 2  
produced by FIU/FAO Documents techniques sur les pêches produits par FI 2  
lement et commercialisation du poisson. Rapport technique 8. (Fr) / Amega 2  
INFOPECHE  
Section 4 - P&#234;ches  
CONSTITUTION OF THE CENTRE FOR MARKETING INFORMATION AND  
37 - Traitement des produits halieutiques. Rapport technique 3. (Fr) / Tall, A.

**Desert Locust Information Service of FAO: Locust FAQs - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

HOME HELP PREFERENCES

search titles a-z language phrases

**INFOPECHE**

INFOPECHE, a été créé en tant que Projet Régional exécuté par l'Organisation des Nations Unies pour l'Agriculture et l'Alimentation (FAO). Sa mission comprend la fourniture d'information et la coopération pour la commercialisation des produits de la pêche en Afrique. Le projet devient le réseau mondial des services d'information et de conseil sur la commercialisation du poisson. Le réseau comprend quatre INFOPECHE: INFOYU, EASTFISH, INFOPECA, INFOFISH couvrant l'Afrique Sud-est; et INFOSAMAK, GLOBEFISH banque de données mondiales, est basée au siège de l'INFO à Rome, complète le réseau INFO de la FAO.

En 1991, les pays membres ont approuvé l'accord pour la création de l'Organisation Intergouvernementale d'Information et de Coopération pour la Commercialisation des Produits de la Pêche en Afrique (INFOPECHE).

**Membres d'INFOPECHE**

Des ratifications de l'accord portant création d'INFOPECHE ont été déposées au bureau du Directeur Général de la FAO par les pays ci-après : Cameroun, Congo, Côte d'Ivoire, Gambie, Ghana, Guinée Bissau, Liberia, Mauritanie, Maroc, Namibie, Nigéria, Sénégal et Sierra Leone. Le Gabon, Zaire et le Cap Vert continuent le processus de finalisation de ratification. Le marché commun pour l'Afrique Australe et l'Afrique de l'Est (COMESA) composé de 22 pays a pour l'intermédiaire du Département de l'Agriculture manifesté son désir de participer aux actes d'INFOPECHE en tant que membres à part entière.



# Agenda

- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

## CDS/ISIS

- ❖ Bibliography collections are typically fairly complex:
  - Form searching
  - Customised query result and browse lists
  - Customised document display
- ❖ Let's work through creating a simple collection using a small CDS/ISIS database describing a set of film slides  
(More information in the "Bibliography collection" and "CDS/ISIS" documented example collections)

## CDS/ISIS

- ❖ Add the CDS/ISIS files to a new collection
- ❖ After building, let's view the collection:
  - No metadata searching is available:

The screenshot shows the CDS/ISIS search interface. At the top, there are three tabs: "search" (highlighted in green), "titles a-z", and "filenames". Below the tabs is a search bar with the placeholder text "Search for [text] that contain [some] of the words". Underneath the search bar are two input fields: "text" and "filenames", both containing the word "text". To the right of these fields is a "Begin Search" button.

- The titles classifier is completely empty!

The screenshot shows the CDS/ISIS titles classifier interface. At the top, there are three tabs: "search" (highlighted in orange), "titles a-z" (highlighted in green), and "filenames". Below the tabs is a list area with the heading "titles a-z". There are no documents listed under this tab.

## CDS/ISIS

- ❖ More problems:
  - The filenames classifier is useless!

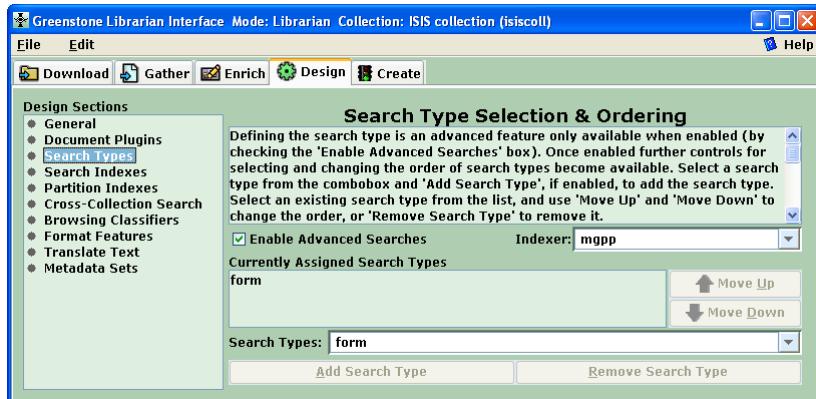
The screenshot shows the CDS/ISIS filenames classifier interface. At the top, there are three tabs: "search" (highlighted in orange), "titles a-z" (highlighted in orange), and "filenames" (highlighted in green). Below the tabs is a list area with the heading "filenames". Three documents are listed, each with the title "Untitled (slide.mst)" and a small document icon.

- The document display isn't very pretty:

The screenshot shows the CDS/ISIS document display interface. At the top, there are three tabs: "search" (highlighted in orange), "titles a-z" (highlighted in orange), and "filenames" (highlighted in orange). Below the tabs is a list area with the heading "filenames". A single document is displayed with the title "Untitled". Below the title, there are buttons for "NO HIGH-LIGHTING" and "DETACH". To the right of the document, there is a large amount of text representing document metadata, including "tag=Year data=1997 tag=Month & Date data=Oct. tag=Notes data=contact plus negatives; Diwali photos tag=Program data=SECRETARIAT tag=Country data=<Fiji> tag=Photographer data=Woodward, P."

## CDS/ISIS: Metadata searching

- ❖ To enable form searching, go to the “Search Types” area in the GLI’s Design pane
  - Tick “Enable Advanced Searches” on
  - Add the “form” search type, and remove “plain”



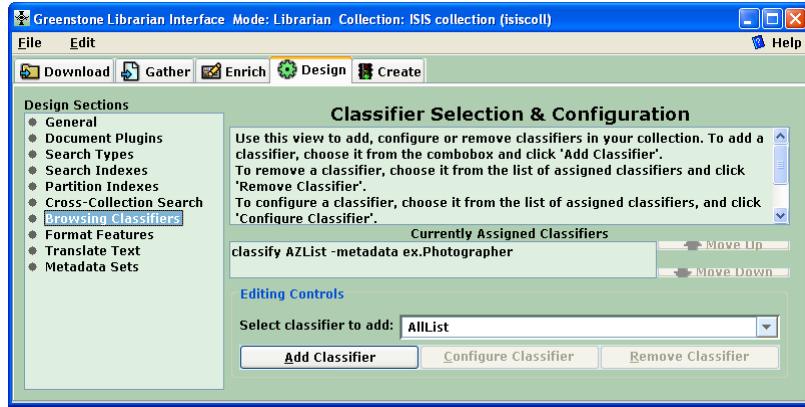
## CDS/ISIS: Metadata Searching

- ❖ Add metadata indexes in the “Search Indexes” part of the GLI’s Design pane
  - Add indexes for Photographer and Notes metadata
  - Remove the useless Source and Title indexes



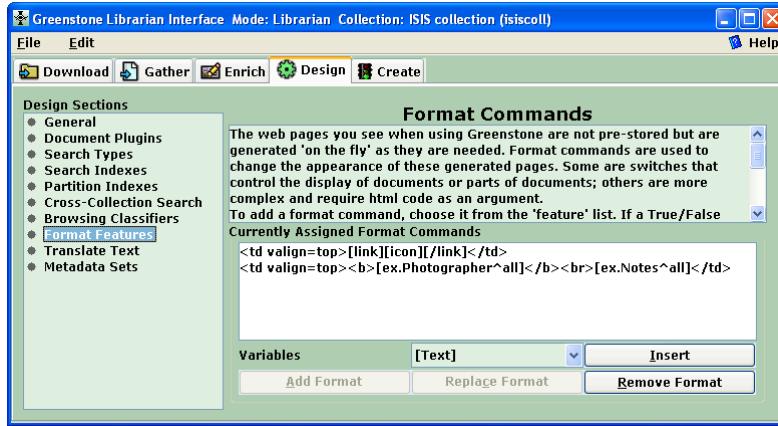
## CDS/ISIS: Better browsing

- ❖ Remove the existing (useless) classifiers for Title and Source metadata, and add a new one for Photographer



## CDS/ISIS: Better browsing

- ❖ Change the VList format statement to display the Photographer and Notes metadata:



## CDS/ISIS: Document display

- ❖ Next, let's change the DocumentText format statement to show the Photographer and Notes metadata:

```
<center><table width=_pagewidth><tr><td>Photographer:</td><td>[ex.Photographer^all]</td></tr><tr><td>Notes:</td><td>[ex.Notes^all]</td></tr></table></center>
```

- ❖ Then, let's remove those annoying "Detach" and "Highlight" buttons by setting DocumentButtons to empty
- ❖ Lastly, clear DocumentHeading to remove the "untitled" at the top of the document

## CDS/ISIS: Finished!

- ❖ Metadata searching now available:

ISIS collection  Photographer

Photographer

Search for  of  
 Word or phrase

	text
	photographers
	notes
	text

- ❖ Better browsing facilities:

ISIS collection  Photographer

Photographer

A-E F-G H-L M-R S T-X

 Australia. Bureau of Mineral Resources

 United States Geological Survey

Slides for presentations on the petroleum potential of the SOPAC region (New Ireland (PNG), Solomon Islands, Tonga and Vanuatu), obtained from N.F. Exxon

 Barclay, W.

10 col. slides of hydrocarbon potential of the Southwest Pacific

## CDS/ISIS: Finished!

### ❖ Document display improved:

search [Photographer](#)

Photographer Australia. Bureau of Mineral Resources  
United States Geological Survey

Slides for presentations on the petroleum potential of the SOPAC region  
(New Ireland (PNG), Solomon Islands, Tonga and Vanuatu), obtained  
from N.F. Exon

### ❖ What could still be improved?

- More metadata indexes, classifiers
- Display all fields in the document display
- Nice images for classifiers
- ...?

## Agenda



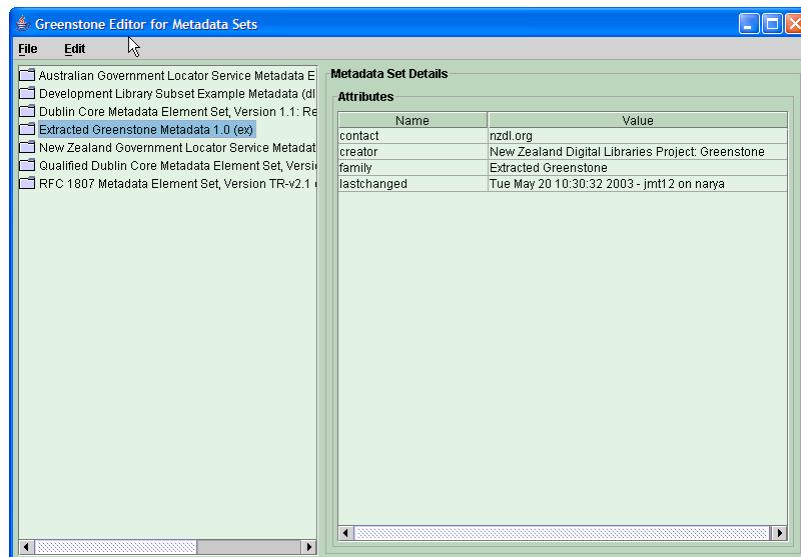
- ❖ Overview of collection building
- ❖ Metadata sets
- ❖ Adding metadata in GLI (+demo)
- ❖ GLI tricks
- ❖ Review: Searching and browsing
- ❖ General Options
- ❖ Plugins
- ❖ Searching – indexes
- ❖ Browsing – classifiers
- ❖ Simple formatting
- ❖ Form search
- ❖ Partitioning indexes
- ❖ PHIND phrase index
- ❖ CDS/ISIS
- ❖ GEMS: metadata set editing

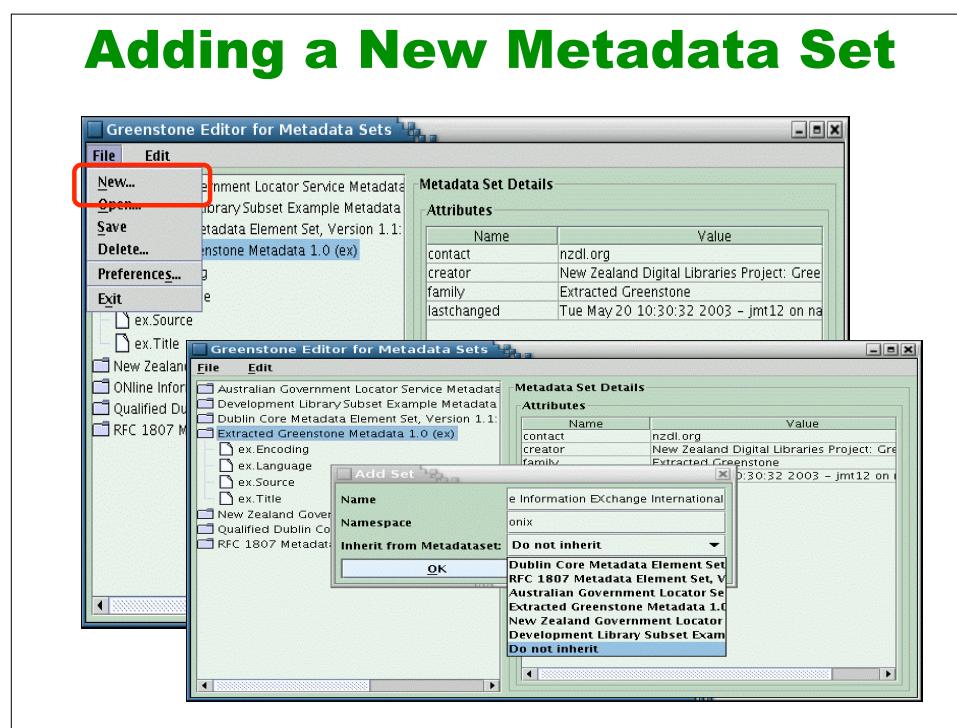
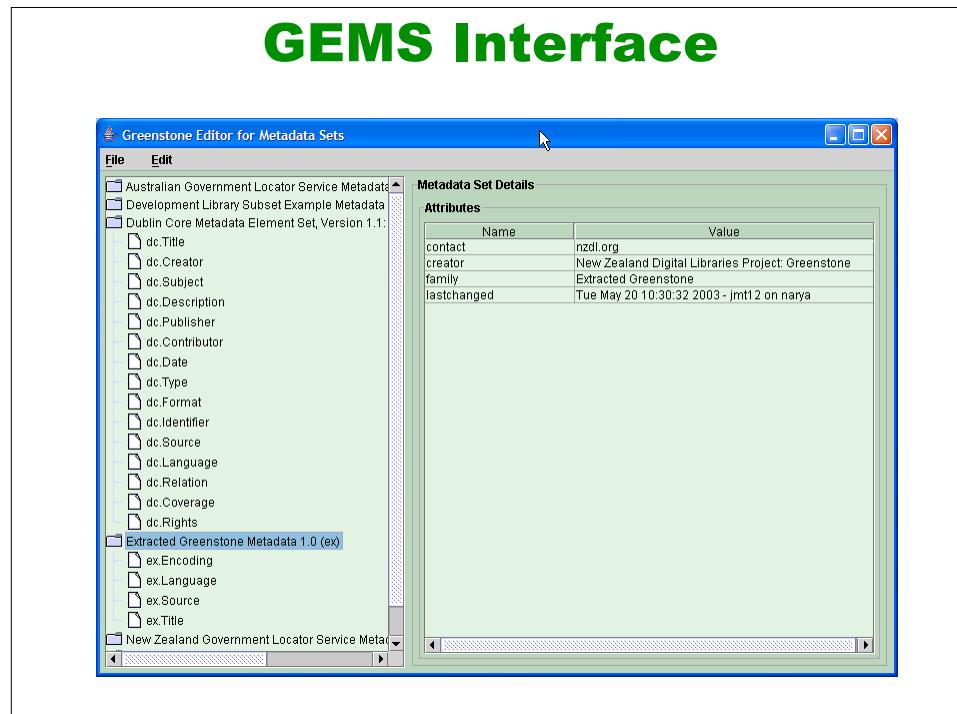
## Greenstone Editor for Metadata Sets - GEMS

### ❖ What can GEMS do?

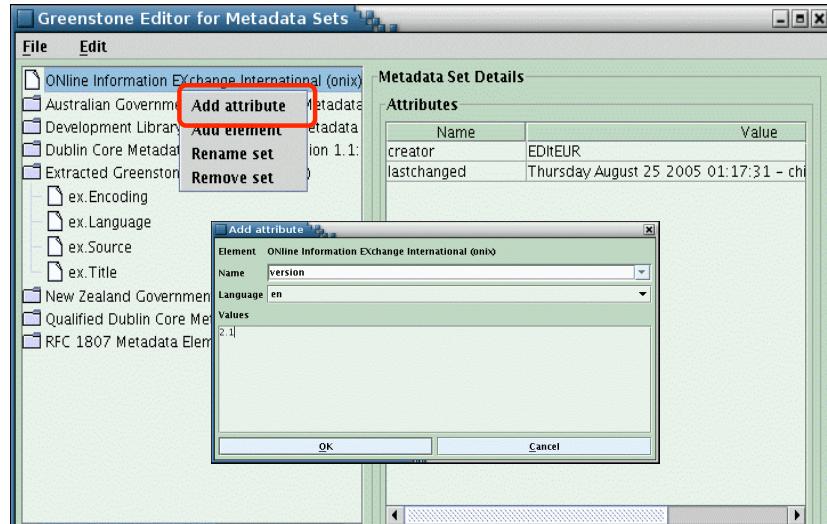
- Create a new metadata set
- Add metadata elements to an existing set
- Define and edit metadata attributes
- Inherit from an existing metadata set
- Import an existing metadata set
- Sets can be used in GLI

## GEMS Interface

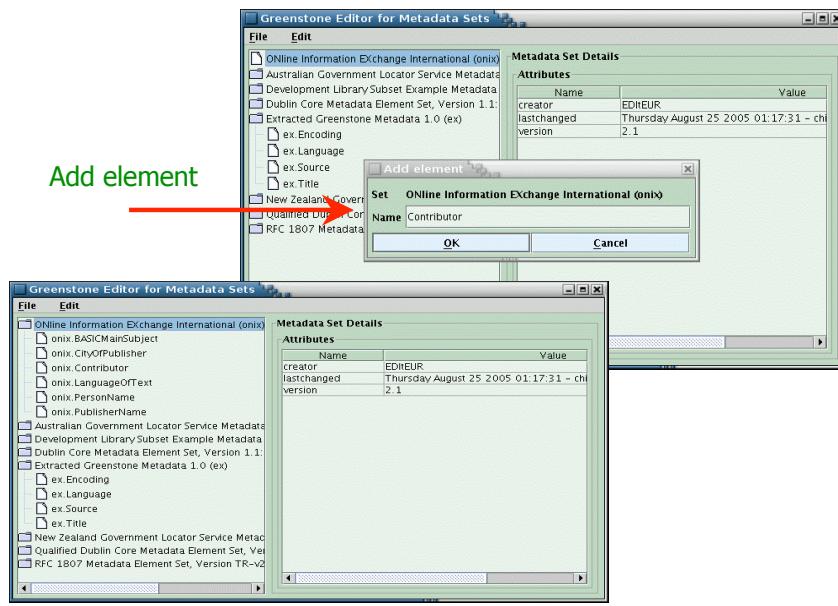




## Adding Metadata Attribute



## Adding Metadata Element



## Using an existing metadata set

- ❖ Using a pre-defined metadata set
  - Must comply with Greenstone metadata XML file format
    - ❖ Write a script to convert?
  - Copy to /gsdl/gli/metadata directory