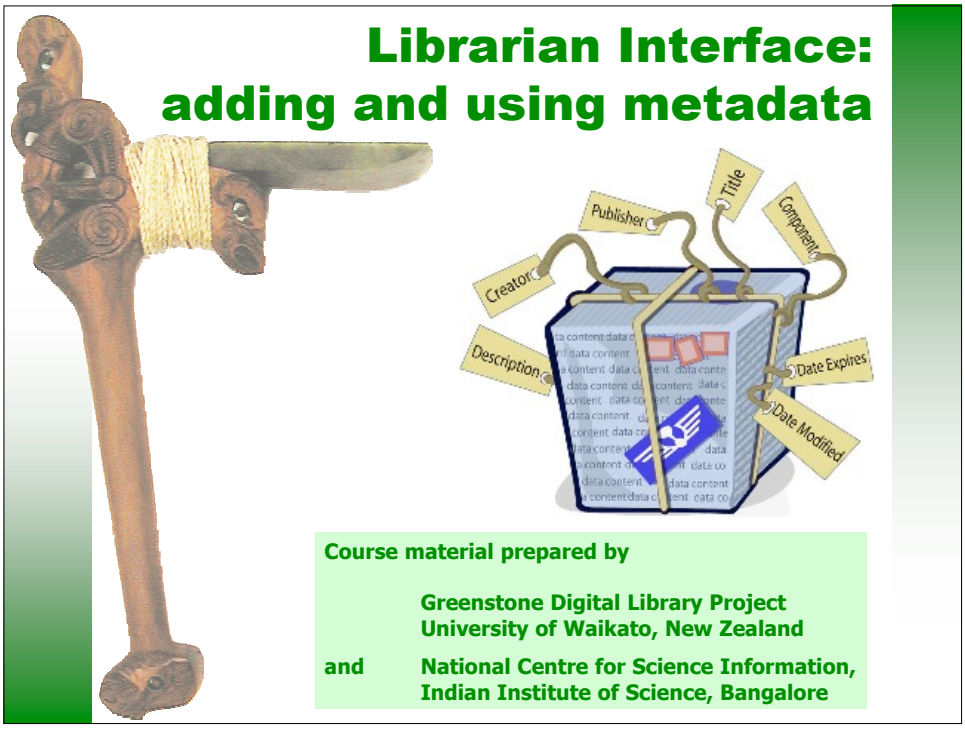

Librarian Interface:
adding and using metadata
Publisher
Title
Component
Creator
Description
Date Expires
Date Modified
Course material prepared by
Greenstone Digital Library Project
University of Waikato, New Zealand
and
National Centre for Science Information,
Indian Institute of Science, Bangalore


Agenda
Metadata sets
Dublin Core (dc)
Development Library Subset (dls)
Extracted (ex)
Adding classifiers
Full text tagging
Image collection
Greenstone archive format (GA)

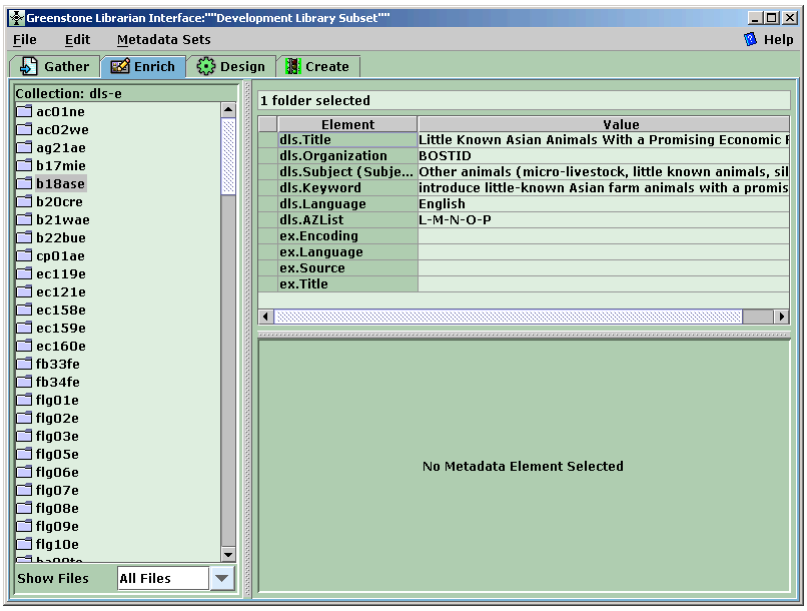
Dublin Core metadata

Metadata	Tag	Definition
Title	dc.Title	Name given to the resource
Creator	dc.Creator	Entity primarily responsible for making the content of the resource
Subject and keywords	dc.Subject	Topic of the content of the resource
Description	dc.Description	Account of the content of the resource
Publisher	dc.Publisher	Entity responsible for making the resource available
Contributor	dc.Contributor	Entity responsible for making contributions to the content of the resource
Date	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
Resource identifier	dc.Identifier	Unambiguous reference to the resource within a given context: this is the object identifier or OID
Source	dc.Source	Reference to a resource from which the present resource is derived
Language	dc.Language	Language of the intellectual content of the resource
Relation	dc.Relation	Reference to a related resource
Coverage	dc.Coverage	Extent or scope of the content of the resource
Rights management	dc.Rights	Information about rights held in and over the resource

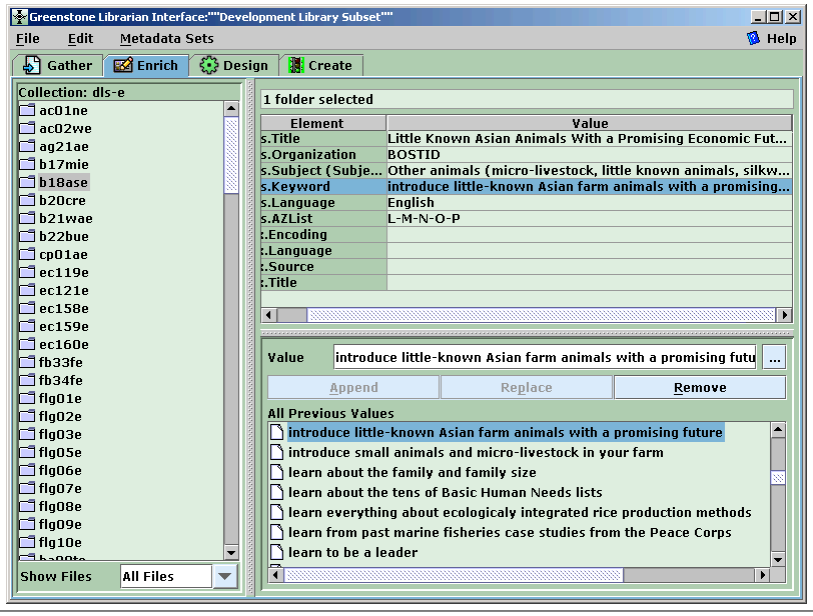
DLS metadata

Metadata	Definition
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

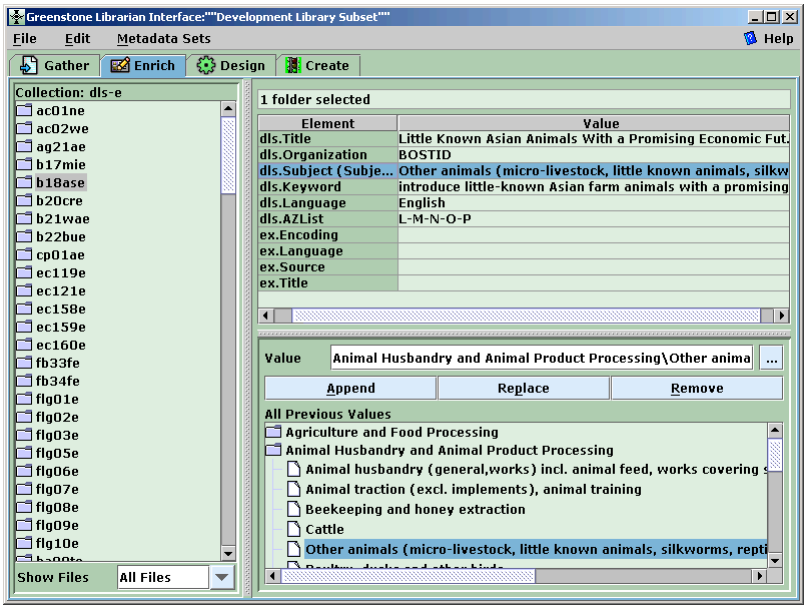
Example of DLS metadata



Example of DLS metadata



Example of DLS metadata



Greenstone Librarian Interface (GLI): adding metadata

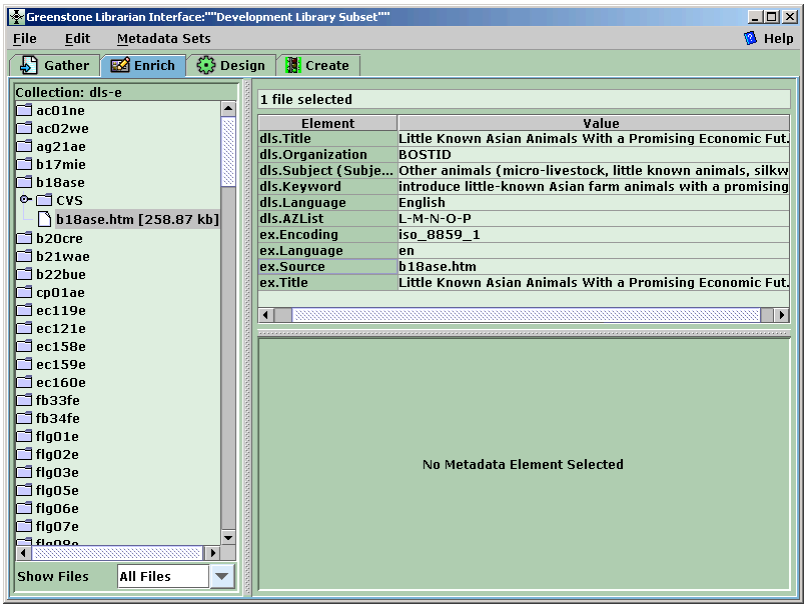
Invoke GLI on Demo collection.

- **Merge metadata values**
- **Add/edit various bits of metadata**

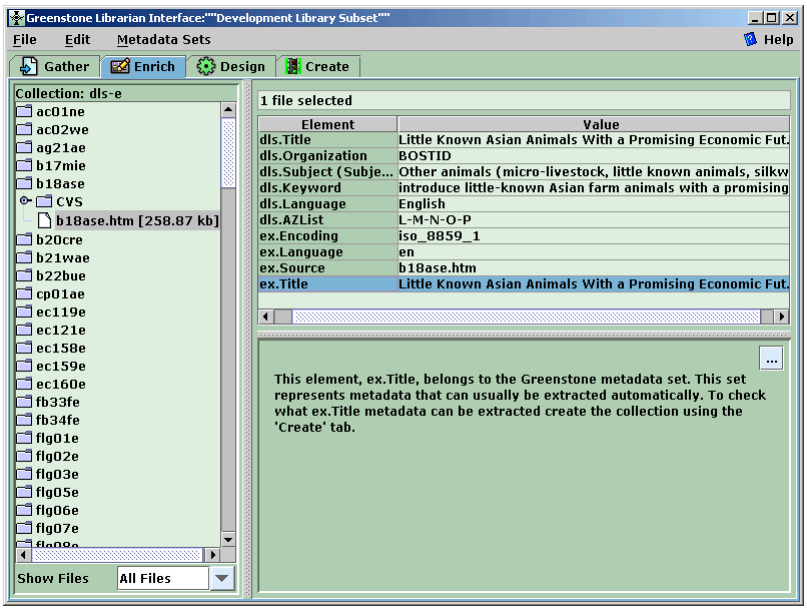
Extracted metadata

Metadata	Definition
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



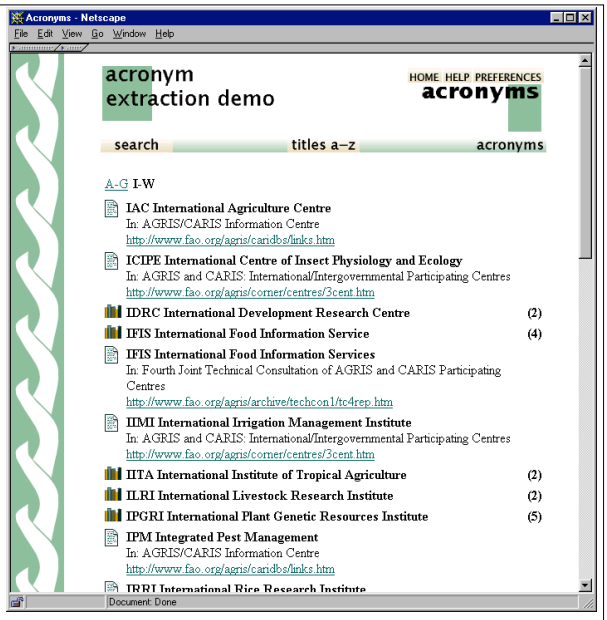
Example of extracted metadata





Language identification plugin

Metadata extraction plugins

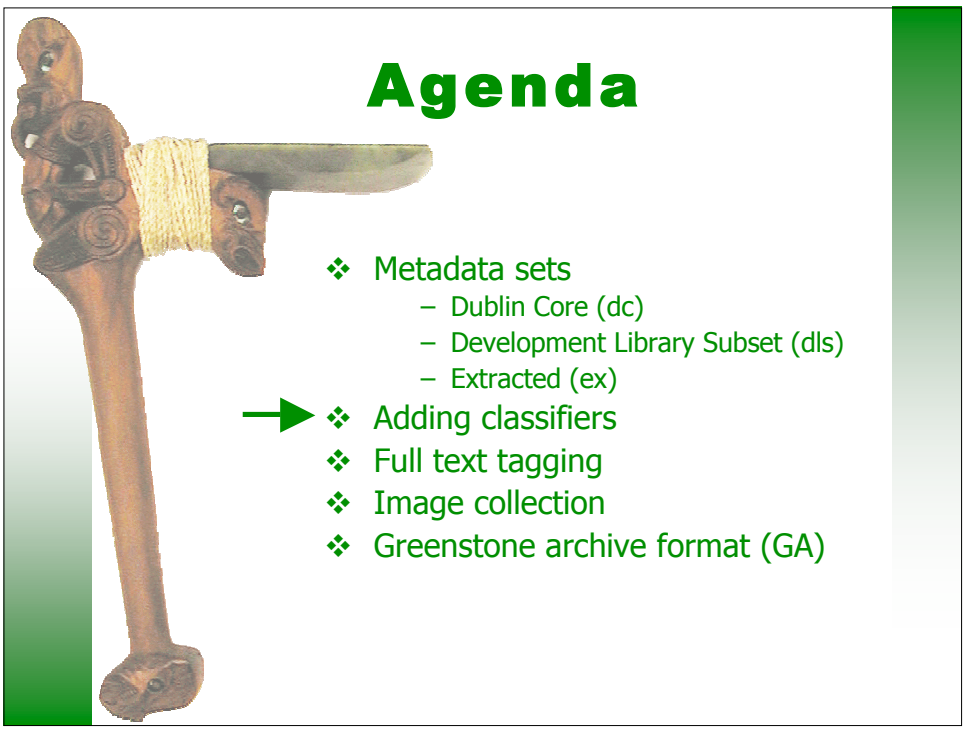


Acronym extraction plugin

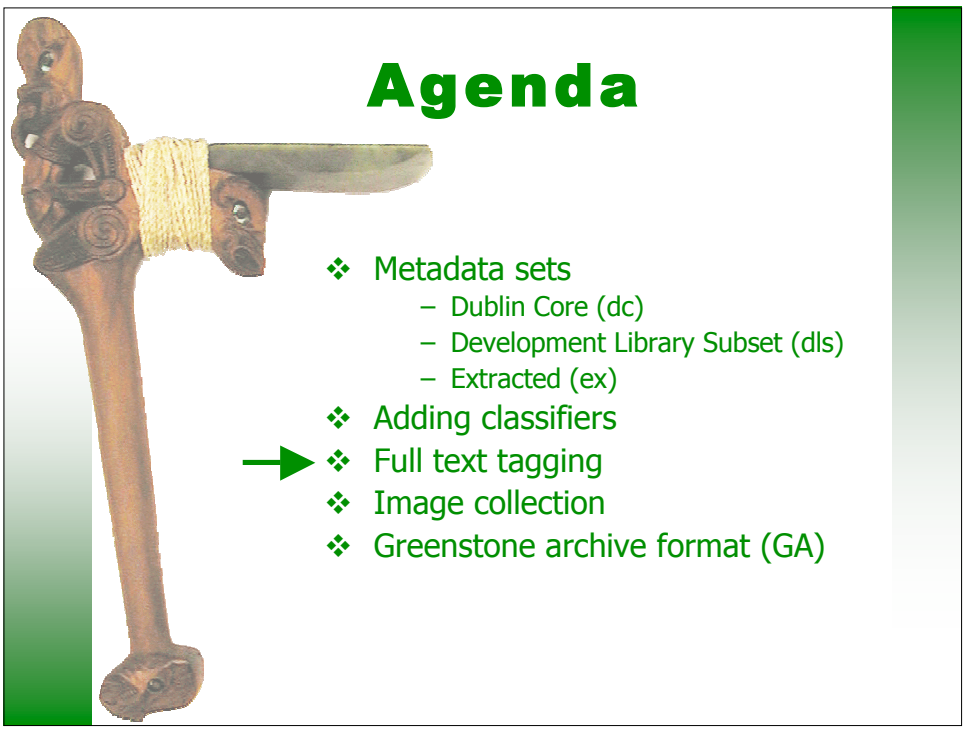
Greenstone Librarian Interface (GLI): extracted metadata

Invoke GLI on small HTML collection.

- **Look at extracted metadata**
- **Build**
- **Look at extracted metadata again**
- **Add acronym extraction switch, rebuild, look again**

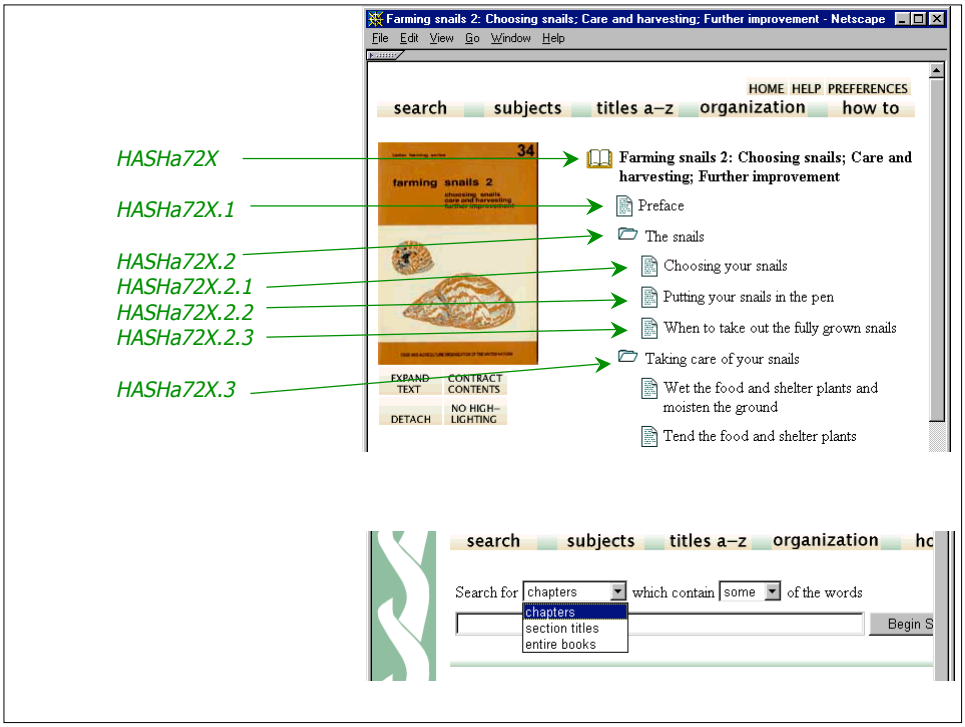


Greenstone Librarian Interface (GLI): adding classifiers



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



Full Text Tagging...

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

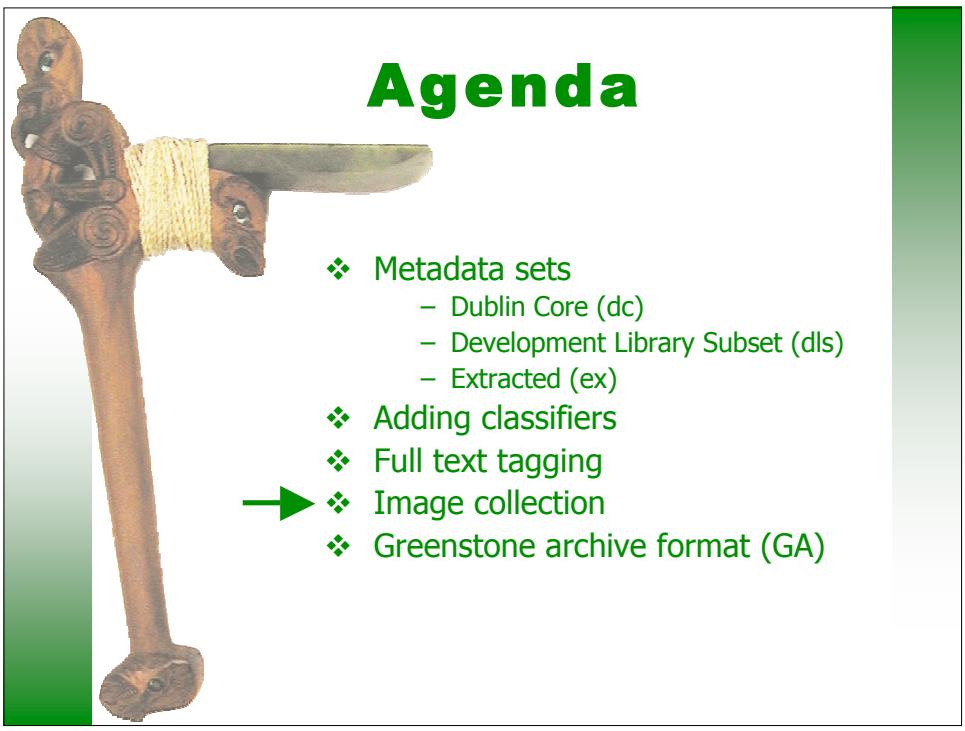
```
<!--
 <Section>
 <Description>
 <Metadata name="Title">Realizing human
  rights for poor people: Strategies for
  achieving the international development
  targets</Metadata>
 </Description>
 -->
  (text of section goes here)
 <!–
 </Section>
 -->
```

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
```

Greenstone Librarian Interface (GLI): full-text tagging demo



Greenstone Librarian Interface (GLI): create an image collection



Greenstone Archive Format: Example document

```
<?xml version="1.0" ?>
<!DOCTYPE GreenstoneArchive SYSTEM
"http://greenstone.org/dtd/GreenstoneArchive/1.0/GreenstoneArchive.dtd">
<Section>
  <Description>
    <Metadata name="gsdlsourcefilename">ec158e.txt</Metadata>
    <Metadata name="Title">Freshwater Resources in Arid Lands</Metadata>
    <Metadata name="Identifier">HASH0158f56086efffe592636058</Metadata>
    <Metadata name="gsdlassocfile">cover.jpg:image/jpeg:</Metadata>
    <Metadata name="gsdlassocfile">p07a.png:image/png:</Metadata>
  </Description>
  <Section>
    <Description>
      <Metadata name="Title">Preface</Metadata>
    </Description>
    <Content>
        This is the text of the preface
    </Content>
  </Section>
  <Section>
    <Description>
      <Metadata name="Title">First and only chapter</Metadata>
    </Description>
    <Section>
      <Description>
        <Metadata name="Title">Part 1</Metadata>
      </Description>
      <Content>
        This is the first part of the first and only chapter
      </Content>
    </Section>

</Section>
```

Greenstone archive format

Document type definition (DTD)

```
<!DOCTYPE GreenstoneArchive [
  <!ELEMENT Section (Description,Content,Section*)>
  <!ELEMENT Description (Metadata*)>
  <!ELEMENT Content (#PCDATA)>
  <!ELEMENT Metadata (#PCDATA)>
  <ATTLIST Metadata name CDATA #REQUIRED>
]>
```

Specifying metadata: XML metadata file

```
<?xml version="1.0" ?>
<!DOCTYPE GreenstoneDirectoryMetadata SYSTEM
"http://greenstone.org/dtd/GreenstoneDirectoryMetadata/1.0/GreenstoneDirectoryMetadata.dtd">
<DirectoryMetadata>
  <FileSet>
    <FileName>nugget.*</FileName>
    <Description>
      <Metadata name="Title">Nugget Point, The Catlins</Metadata>
      <Metadata name="Place" mode="accumulate">Nugget Point</Metadata>
    </Description>
  </FileSet>
  <FileSet>
    <FileName>nugget-point-1.jpg</FileName>
    <Description>
      <Metadata name="Title">Nugget Point Lighthouse</Metadata>
      <Metadata name="Subject">Lighthouse</Metadata>
    </Description>
  </FileSet>
</DirectoryMetadata>
```

XML metadata format

Document type definition (DTD)

```
<!DOCTYPE GreenstoneDirectoryMetadata [
  <!ELEMENT DirectoryMetadata (FileSet*)>
  <!ELEMENT FileSet (FileName+,Description)>
  <!ELEMENT FileName (#PCDATA)>
  <!ELEMENT Description (Metadata*)>
  <!ELEMENT Metadata (#PCDATA)>
  <ATTLIST Metadata name CDATA #REQUIRED>
  <ATTLIST Metadata mode (accumulate|override) "override">
]>
```

Documentation and Help

- *User's Guide* (user.pdf)
  (Section 3.3 on full-text tagging)
- Tooltips
- On-line help
- nzdl.org: Simple image collection

Questions?
Comments?
Discussion?
Feedback form!