

Lab 3 (2 hours): Librarian Interface: Using metadata, simple collection configuration

Part I – Configuring the collection of Word and PDF documents

1. In the Librarian Interface, open up the **reports** collection you built in **Lab 2**:

File→Open

[If you have not done **Lab 2**, please do **Part III** of **Lab 2** now, where you create a new collection of PDF and Word documents, and add some *dc.Title* metadata.]

2. Add *dc.Creator* metadata to the documents. Remember that you can double click on a document in GLI to open and view it. You can add more than one value for the same field: when you press Enter in a metadata value field, a new empty field of the same type will be generated. Add each Author separately as *dc.Creator* metadata.
3. If you build and preview your collection at this point, you will find that nothing has changed. You need to alter the collection design to use the new *dc.Creator* metadata.

The following exercises involve configuring the collection using the **Design** panel in the GLI. This includes adding search indexes and browsing hierarchies for the new metadata you have just entered.

Branding a collection with an image

4. Change to the **Design** panel by clicking on its tab. The panel is split into several sections. The first section **General Options** appears. This allows you to modify the values you provided when defining the collection, if desired. You can also brand the collection using a suitable image.
5. Click on the <**Browse**> button associated with “URL to about page icon”, browse to the image *My Computer → D → workshop_files → wrdpdf.gif*, and click <**Open**>. When you select this image, Greenstone automatically generates an appropriate URL for the image. **Preview** the collection: you should see the new image at the top left of the page.

[Information on the **General Options** page does not require a rebuild of the collection to take effect. Just go to the **Create** panel and click <**Preview Collection**>.

Document plugins

6. Back in the GLI, look at the **Document Plugins** section of the **Design** panel, by clicking on this in the list to the left. Here you can add, configure or remove plugins to be used in the collection. There is no need to remove any plugins, but it will speed up processing a little. In this case we have only Word and PDF

documents, and can remove the ZIPPlug, TEXTPlug, HTMLPlug, EMAILPlug, RTFPlug, PSPlug, ImagePlug and NULPlug plugins. To delete a plugin, select it and click **<Remove plugin>**. GAPug is required for all collections and should not be removed.

[The next section is **Search Types**. In this exercise, we will not make any changes to this.]

Search Indexes

7. The **Search Indexes** section specifies what parts of the collection are searchable (e.g. searching by text or Title). Delete the *ex.Source* index, which is not particularly useful, by selecting it and clicking **<Remove Index>**.
8. Modify the *ex.Title* index to include *dc.Title* by selecting the index in the **Assigned Indexes** box, and then selecting *dc.Title* from the **Build index on** box. Click **<Replace Index>**. Searching this index will search both *dc.Title* and *ex.Title* metadata. If you want to restrict searching to just the manually added *dc.Title* metadata, deselect *ex.Title* from the **Index Source** box and click **<Replace Index>**.
9. Now add a new index based on *dc.Creator*. Change the **Index Name** field to “authors”, and select *dc.Creator* from the **Index Source** box. Make sure you deselect the *ex.Title* and *dc.Title* metadata items. Click **<Add Index>**.

[The next two sections are **Partition Indexes** and **Cross-Collection Search**. In this exercise, we will not make any changes to these.]

Browsing classifiers

10. The **Browsing Classifiers** section adds “classifiers”, which provide the collection with browsing functions. Go to this section and observe that Greenstone has provided two classifiers, *AZLists* based on *ex.Title* and *ex.Source* metadata. These correspond to the *titles a-z* and *filenames* buttons on the collection’s access bar. Remove the *ex.Source* classifier by selecting it and clicking **<Remove Classifier>**.
11. Modify the *ex.Title* classifier to use the *dc.Title* metadata instead. Select the *ex.Title* classifier and click **<Configure Classifier>**. In the **metadata** box, select *dc.Title* instead of *ex.Title*. Click **<OK>**.
12. Add an *AZCompactList* classifier for *dc.Creator*. Select **AZCompactList** from the **Select classifier to add** drop-down list, and click **<Add Classifier>**. A popup window **Configuring Arguments** appears. Select *dc.Creator* from the **metadata** drop-down list. Activate the **mingroup** box and select 1. Click **<OK>**.

AZCompactList is like **AZList**, except that terms that appear multiple times in the hierarchy are automatically grouped together and a new node, shown as a bookshelf icon, is formed. Setting **mingroup** to 1 means that the bookshelf

appears even when there is just one item, and is done here to provide a more uniform display.

[The last three sections are **Format Features**, **Translate Text** and **Metadata Sets**. In this exercise, we will not make any changes to these.]

13. Switch to the **Create** panel, and **build** and **preview** the collection.
14. Check that all the facilities work properly. There should be three full-text indexes, called *text*, *titles* and *authors*. In the *titles a-z* list should appear all the documents to which you have assigned *dc.Title* metadata (and only those documents). In the *authors a-z* list should appear one bookshelf for each author you have assigned as *dc.Creator*. Clicking on a bookshelf should take you to all the documents they have authored.

Classifying on multiple metadata

15. The new *titles a-z* list only shows those documents which have assigned *dc.Title* metadata. For many documents, extracted Titles may be fine, and it is impractical to add the same metadata again as *dc.Title*. Fortunately there is a way we can use both metadata types in one classifier: specify a list of metadata names in the classifier.
16. In the **Browsing Classifiers** section of the **Design** panel, select the *dc.Title* AZList classifier in the **Currently Assigned Classifiers** box and click **<Configure Classifier>**. Note you can achieve the same result by double clicking on the classifier.
17. Type *,ex.Title* after the *dc.Title* in the metadata box—i.e. make it read

dc.Title,ex.Title
18. **Build** the collection again and **preview** it. Now all the documents should appear in the *titles a-z* list.

Extracted metadata is unreliable. But it is very cheap! On the other hand, manually assigned metadata is reliable, but expensive. The previous section of this exercise has shown how to aim for the best of both worlds by using extracted metadata but correcting it when it is wrong. While this may not satisfy the professional librarian, it could provide a useful compromise when you want to get your collection together with a minimum of effort.

Part II – Building an image collection

1. Start a new collection (File→New) called **backdrop**. Fill out the fields with appropriate information. For “**Base this collection on**”, select the item **Simple image collection (image-e)** from the drop-down menu. Then click <OK>.

[When you base a collection on an existing one, it inherits all the settings of the old one. You won’t be asked to choose a metadata set because the new collection inherits the one (or ones) used by the seed collection.]
2. Copy the images provided in the *workshop_files* folder *sample_images* into your newly-formed collection.
3. Change to the **Create** panel and **build** the collection.
4. **Preview** the result.
5. Click on **browse** in the navigation bar to view a list of the photos ordered by filename and presented as a thumbnail accompanied by some basic data about the image. The structure of this collection is the same as ‘image-e’, but the content is different.
6. Change to the **Enrich** panel, and view the extracted metadata for *Ascent.jpg*.

Adding a metadata set to the collection

7. We will now manually add our own metadata and use it to give users a new way to browse the collection
8. The collection (image-e) on which **backdrop** is based uses only extracted metadata. We will add a further metadata set that matches our needs. For this exercise we use the Dublin Core metadata set. Go to the **Design** panel of the Librarian Interface and click <Metadata Sets> in the list on the left. Click <Add Metadata Set> (lower left button). In the window that pops up, select **dublin.mds** and click <Add Metadata Set>.
9. Change to the **Enrich** panel by clicking this tab. The metadata for each file now shows the Dublin core ‘dc.’ fields as well as the extracted ‘ex.’ fields.

Adding Title and Description metadata

10. Let’s work with just the first three files (*Ascent.jpg*, *Autumn.jpg* and *Azul.jpg*) to get a flavour of what is possible.
11. First, we set each file’s **dc.Title** field to be the same as its filename but without the filename extension. For example, the **dc.Title** for *Ascent.jpg* would be **Ascent**.
12. Add a description for each image as **dc.Description** metadata. What description should we enter? To remind yourself of a file’s content, the Librarian Interface lets you open files by double-clicking them. It launches the appropriate

application based on the filename extension: Word for *.doc* files, Acrobat for *.pdf* files and so on. Double-click *Ascent.jpg*: the image will normally be displayed by Microsoft's Photo Editor (although this depends on how your computer has been set up).

Back in the Librarian Interface enter the text **Moon rising over mountain landscape** as the **dc.Description** for *Ascent.jpg*. Add suitable descriptions for *Autumn.jpg* and *Azul.jpg*.

Change format features to display new metadata

13. Building or previewing the collection at this point won't reveal anything new. That's because we haven't changed the design of the collection to take advantage of the new metadata.

14. Go to the **Design** panel and select **Format Features** from the left-hand list.

If the size of the GLI window is too small, the **Currently Assigned Format Commands** list may not be visible. In this case, enlarge the window by clicking on the maximize button (top right corner) or by dragging on the edge of the window.

15. Leave the **Editing Controls** at their default value, so that **Choose Feature** remains blank and **VList** is selected as the **Affected Component**. In the **HTML Format String**, edit the text as follows:

Change '`_ImageName_:`' to '`Title:`'
Change '`[Image]`' to '`[dc.Title]`'

Place your cursor after the text that says

```
[dc.Title]<br>
```

and add the following text:

```
Description: [dc.Description]<br>
```

[Note: Metadata names are case-sensitive in Greenstone: it is important that you capitalize "Title" and "Description" (and not "dc").]

16. Next click **<Replace Format>**. The new format statement will be displayed in the **Currently Assigned Format Commands** list. The first substitution alters the fragment of text that appears to the right of the thumbnail image, the second alters the item of metadata that follows it. The addition displays the description after the Title.
17. Go to the **Create** panel and click **<Build Collection>**. Now **preview** the collection. Now when you click on **browse** in the navigation bar the presentation has changed to "Title: Ascent", and so on. Each image's description should appear beside the thumbnail, following the title.

Because we only assigned metadata to the first three items, after this the title and description become blank because the subsequent items have no dc.Title or dc.Description metadata. You need to spend more time entering the metadata to get a full listing.

[Note: For some design parameters the collection must be rebuilt before the effect of changes can be seen. However, changes to format statements take place immediately and you can see the result straightaway by pressing <**Preview Collection**> in the **Create** panel. Above, you were asked to build before previewing because you had also added metadata.

Changing the size of image thumbnails

18. Let's change the size of the thumbnail image and make it smaller. Thumbnail images are created by the *ImagePlug* plug-in, so we need to access its configuration settings. To do this, switch to the **Design** panel and select **Document Plugins** from the list on the left. Double-click **plugin ImagePlug** to pop up a window that shows its settings. (Alternatively, select **ImagePlug** with a single click and then click <**Configure Plugin**>.) Currently all options are off, so standard defaults are used. Select **thumbnailsize**, set it to **50**, and click <**OK**>.
19. **Build** and **preview** the collection.
20. Once you have seen the result of the change, return to the **Design** panel, select the configuration options for **ImagePlug**, and switch the **thumbnailsize** option off so the thumbnail will revert to its normal size when the collection is re-built.

Adding a browsing classifier based on Description metadata

21. Now we'll add a new browsing option based on the descriptions. Switch to the **Design** panel and select **Browsing Classifiers** from the left-hand list. Set the menu item for **select classifier to add** to **AZList**; then click **Add classifier**.
22. A window pops up to control the classifier's options. Set the menu item for metadata to **dc.Description**, and click <**OK**>. Now switch to the **Create** panel, **build** the collection, and **preview** it. Choose the new **descriptions** link that appears in the navigation bar.

Only three items are shown, because only items with the relevant metadata (dc.Description in this case) appear in the list. The original **browse** list includes all photos in the collection because it is based on ex.Image, extracted metadata that reflects an image's filename, which is set for all images in the collection.

Creating a searchable index based on Description metadata

23. We now build a searchable index based on dc.Description metadata. Switch to the **Design** panel and select **Search Indexes** from the left-hand list. Enter the text "descriptions" as the **Index Name**, select **dc.Description** and click <**Add Index**>.

24. Switch to the **Create** panel, **build** the collection, then **preview** it. As an example, search for the term “mountain” in the *descriptions* index.

Controlling the building process

25. Developing a new collection usually involves numerous cycles of building, previewing, adjusting some enrich and design features, and so on. While prototyping, it is best to temporarily reduce the number of documents in the collection. This can be accomplished through the ‘maxdocs’ parameter to the building process.

Click on the **Create** panel, in the options that appear, select **maxdocs** and set its numeric counter to **3**. Now **rebuild**. The build output will only show three documents being processed.

26. **Preview** the newly rebuilt collection’s **browse** page. Previously this listed more than a dozen photos, but now there are just three—the first three files encountered by the building process.

In this case the image collection, being intended for demonstration purposes, is small to begin with and doesn’t take long to build in its entirety. Real collections often contain thousands of documents, and setting maxdocs to (say) 50 yields a good preview in a fraction of the time taken to build the whole collection.

Part III – Your custom collection

1. If you haven’t done **Part IV** of **Lab 2**, do so now.
2. Add extra metadata to your collection, and add new indexes and browsing classifiers to utilize this new metadata.

Part IV – Extra work

Have a look at the collection configuration file for one of the collections you have been working on. Using WordPad, open
C:\Program Files\Greenstone\collect\reports\etc\collect.cfg. This contains all the settings for the **reports** collection that you specified in the **Design** panel of GLI. Before we had the Librarian Interface, this is what collection designers had to edit by hand!