

Information Management Resource Kit

Module on Digitization and Digital Libraries

UNIT 6. EXAMPLE OF DIGITAL LIBRARY SOFTWARE: GREENSTONE

LESSON 2. USING GLI: GATHERING AND ENRICHING DOCUMENTS

NOTE

Please note that this is a reference guide covering the main steps for using GLI as described in the corresponding lesson. It provides an excellent means for additional practice and for printing after you have completed the course.

We recommend that you first take the lesson using the courseware environment, as it contains interactive features such as exercises with feedback, pop-ups, animations etc.



© FAO and UNESCO, 2005

SCOPE NOTES

These practical exercises will help you build an operational digital library collection from a set of sample source documents. The collection will be built using the Greenstone Librarian Interface (GLI).

Building a new collection using GLI takes five steps:

- 1) Gathering
 - Copying the documents from the computers' file space, including existing collection/s, into the new collection. Any existing metadata remains attached to these documents. Documents may also be gathered from the web through a built-in mirroring facility
- 2) Enriching
 - Enriching the documents by adding further metadata to individual documents or groups of documents
- 3) Designing
 - Designing the collection by establishing its appearance and the access facilities it will support
- 4) Creating
 - Building the collection
- 5) Previewing
 - Previewing the newly created collection, which will be included in your Greenstone home page as one of the collections.

By using this document you will apply steps 1) and 2): Gathering and Enriching.

GLI HELP

Help for using GLI is available in several ways. Apart from the Greenstone User's Guide, page 20, Section 3.2 Librarian User Guide, you can get On-line help any time using the Help button at the top right corner of the GLI's main menu bar. For most GLI controls, you can also get quick tips (Tool tips) by simply pointing the mouse cursor at a related button or option.

INSTALLING GREENSTONE AND GLI

To use Greenstone and GLI, you will first need to install them on your computer system. Brief installation instructions for installing Greenstone on a Windows operating system are given below:

- Access the IMARK CD-ROM
- Go to Resources/Software&Tools/Greenstone Digital Library/ Software and follow the instructions for downloading the application.

The installation process begins by prompting you to choose the setup language. Note that this is the language to be used in the setup screens only, and is not the language of the document collection.

- Click OK to accept the default language, (e.g. English)
- Click Next>, to continue
- Click Yes, to accept all the license agreement terms
- Click Next, to accept C:\Program Files\gsdl as the default Installation (Destination) Folder

Greenstone has four different setup types:

- Custom
 - Custom setup allows you to install any or all of the features provided by the other three setup types. It also allows you to install additional features not included in the other three-setup types.
- Local Library [Recommended]
 - This version of Greenstone has a built-in web server and is suitable for building and viewing Greenstone collections on a stand-alone system. It has limited web serving abilities.
- Source Code
 - Installs the source code only [binary executables will not be installed]. You need to compile the source code before it can be used. This installation type does NOT compile Greenstone for you.
- Web Library [requires a separate web server]
 - This is the recommended setup type for those wanting to serve Greenstone collections on the web. It requires a separate web server [e.g. Apache].

Let us setup the Local Library edition of Greenstone for this practical exercise.

- Click Next> to install the Local Library [recommended] edition of Greenstone. This is the default setup type.
- In the next screen, enter the desired password - 'admin' is the default. Make sure that you take note of the password!

The installation process starts the moment you click on the Next button. A status bar will show the progress of the installation. Finally

- Click 'Finish' to complete the Local Library installation process.

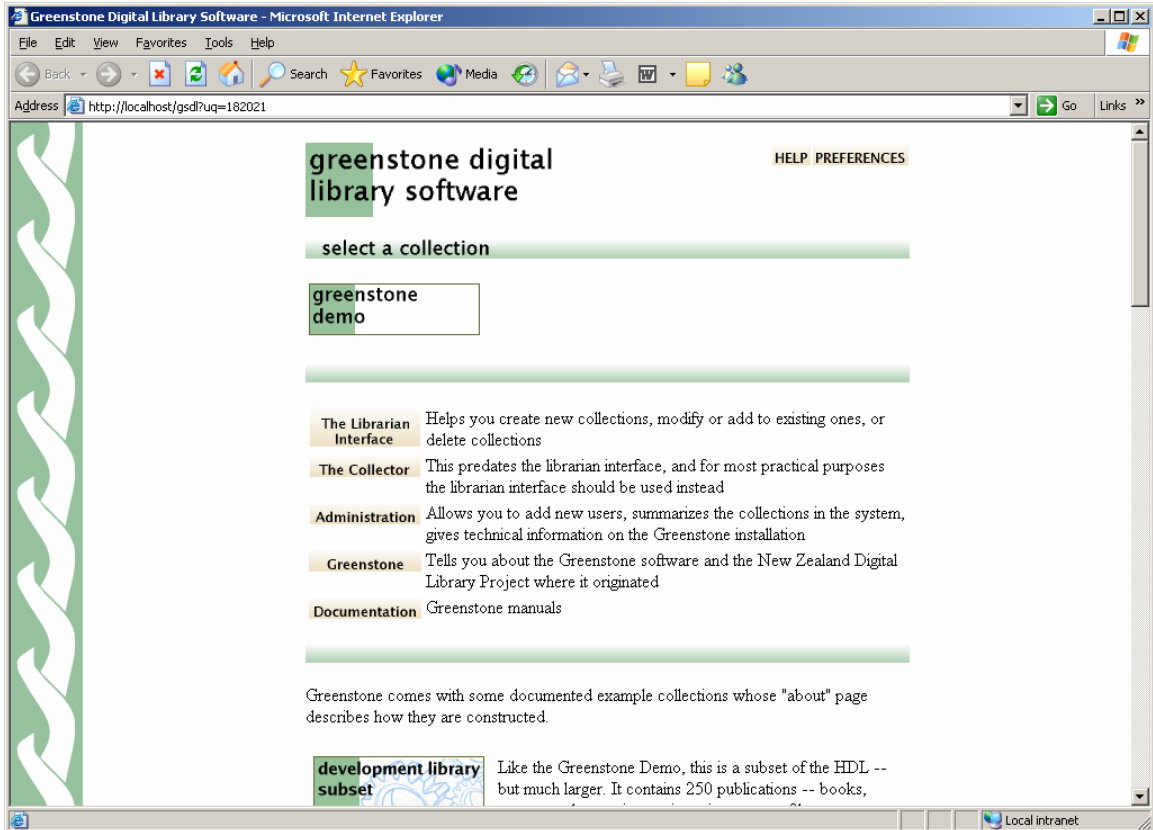
For the simplest installation, as illustrated above, just accept the default at each point by clicking the *Next* button. That's all you need to do! Greenstone is installed in the directory *C:\Program Files\gsdl*.

INSTALLING GREENSTONE AND GLI

Once the installation is complete, start your Greenstone system by:

- clicking the *Start* button → *Programs* menu → *Greenstone Digital Library*, and
- clicking *Enter Library* in the dialogue box.

This automatically opens your Internet browser and loads the Greenstone Digital Library home page. You can access the 'greenstone demo' collection by clicking on its icon.



Greenstone Homepage

(Note: "The Librarian Interface" button you see on the Greenstone home page is for your information only – you cannot start GLI from here).

ABOUT SAMPLE SOURCE DOCUMENTS (INPUT FILES FOR COLLECTION BUILDING)

The sample source documents you will be using in this exercise include documents in different file formats like PDF, Word, RTF, and Postscript. Using this sample source collection, let us go through all the required steps for building a collection using GLI. The collection building process will illustrate Greenstone's ability to handle documents in different file formats.

The source documents are located in the IMARK CD (software_tools/Greenstone/collect/wrdpdf-e/import).

GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

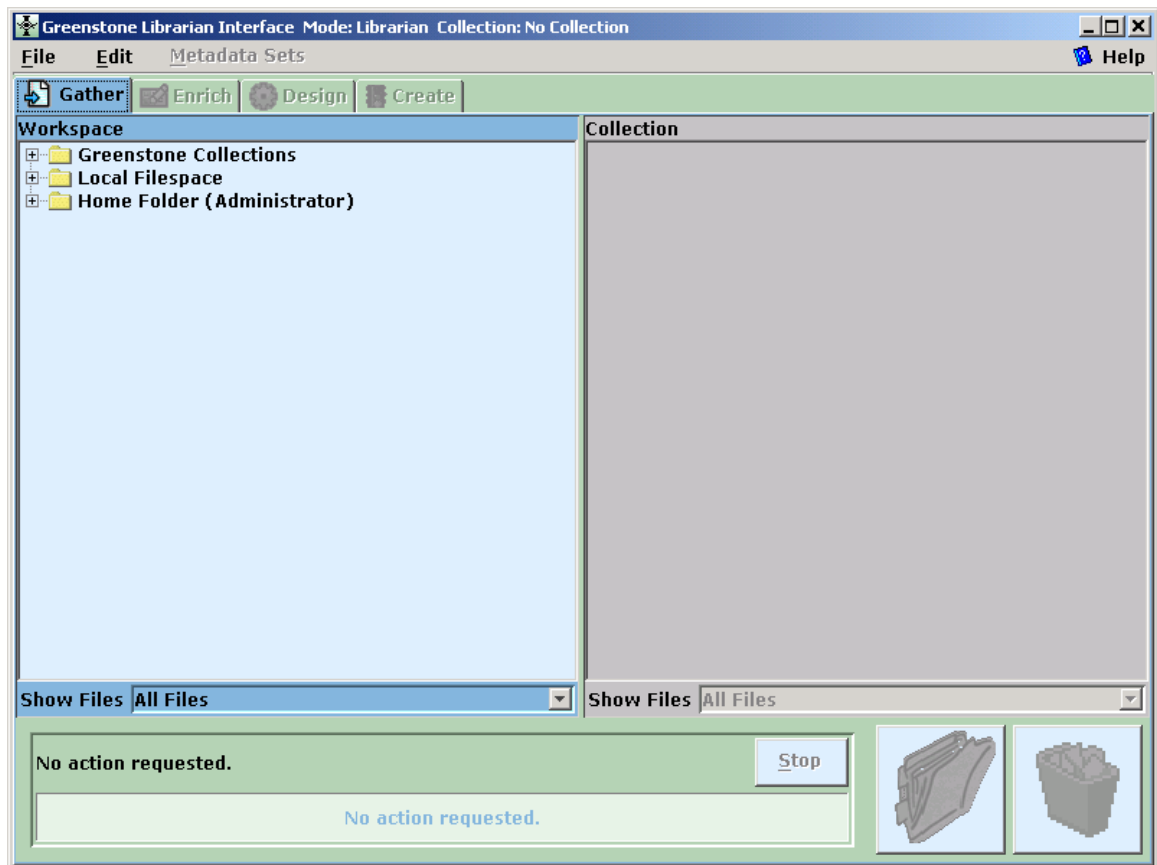
Starting GLI

Let us build a new collection called 'Word PDF' collection using the GLI.

Before you proceed any further, start the GLI in a separate window. To start the GLI:

- click Start → Program → Greenstone Digital Library → Greenstone Librarian Interface

Depending on the configuration of your PC, GLI will start almost immediately or may take several seconds. Once GLI is started, you should see the GLI interface below on your computer screen. This is the 'Gather View'.



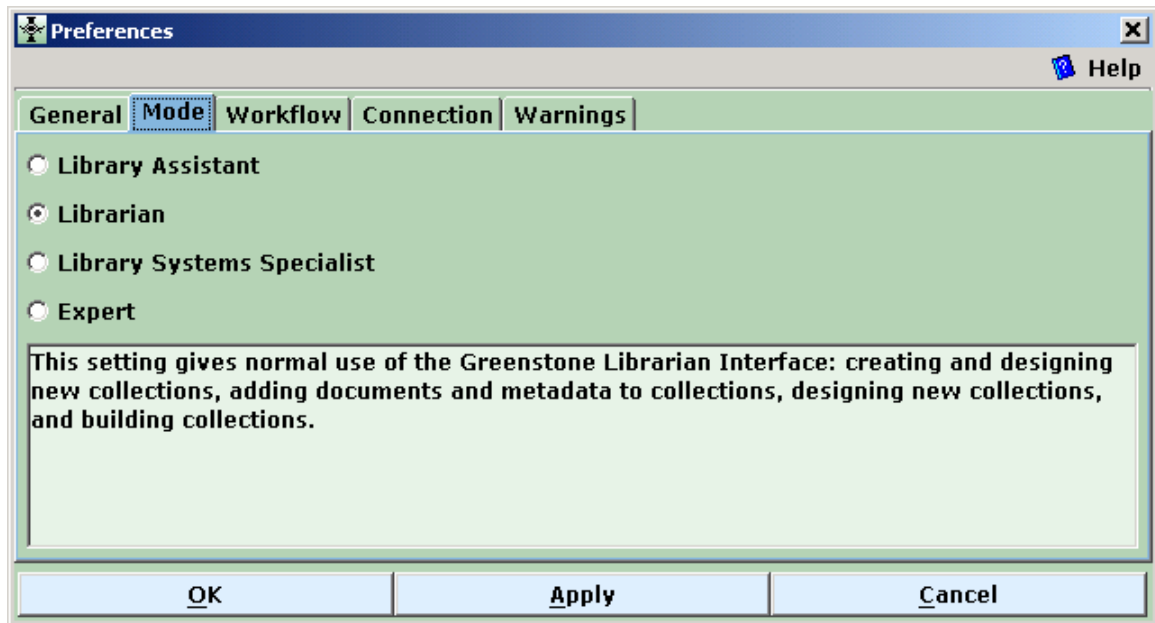
Gather View

GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

'Mode' Settings

Before proceeding, make sure you have selected the correct 'Mode'. This setting is available in the 'File' -> 'Preferences' menu option in GLI.

GLI supports four modes of usage: Library Assistant, Librarian, Library Systems Specialist and Expert. These assume different levels of expertise for collection building: Library Assistant being the lowest and Expert the highest. Supported functionalities also vary (e.g. Library Assistant cannot configure the collection using the 'Design' step). For this exercise, make sure you select the 'Librarian' mode (see figure).



GLI 'Mode' Setting

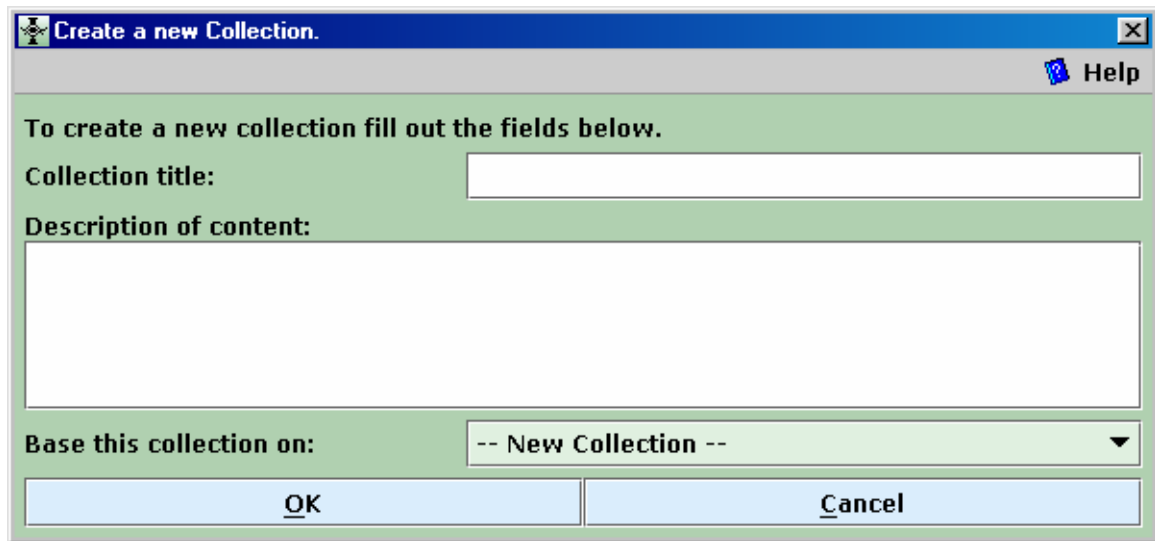
GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

Creating a new collection/1

You are now ready to begin the collection building process using GLI. During the collection building process you will need to go back and forth between this self-learning material and the GLI, which should be started by now. Go to the GLI and:

- Select the New option from the File menu located at the top left hand corner.

A pop-up window 'create a new collection' as shown below will come up.



Create a New Collection

Now, you need to provide information for 'Collection title' and 'Description of content'.

The collection title is a short phrase which identifies the collection's content. Let us call our collection 'WordPDF'. In the Description of content box, you can provide a brief summary of the nature of documents which comprise this collection. For the WordPDF collection you can provide the following description:

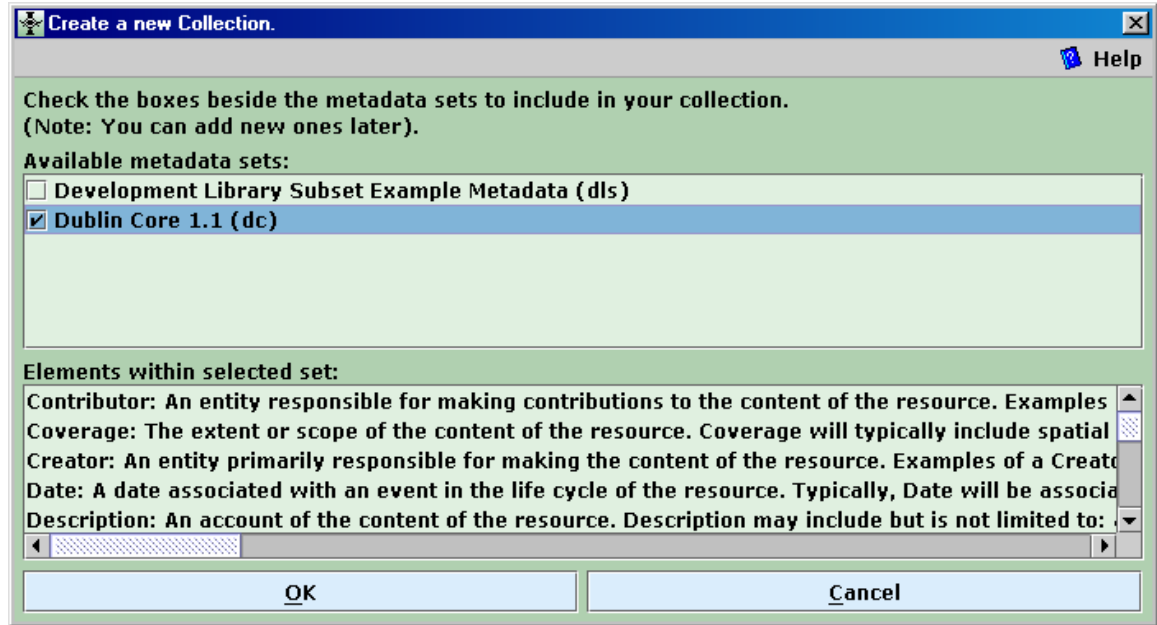
This collection demonstrates Greenstone's ability to handle documents in different file formats. The source collection contains several papers written by various members of the NZDL project. The papers are in PDF, MSWord, RTF, and Postscript formats

Note: *The present version of GLI does not allow you to use the mouse right button to copy and paste the values for collection title and description of content. You will have to manually key-in the text or use Control-C to copy the input text and Control-V to paste.*

In this step, as an extra option, you can base the configuration of the WordPDF collection ('base this collection on' option) on any of the existing collections. For example you can base your collection on the 'greenstone demo (demo)' collection, or build an entirely new type of collection. This means that the new collection will use the same metadata scheme, indexing, search, browse and display features as that of the 'greenstone demo (demo)' collection. Let us build our collection with an entirely new collection configuration. Click on OK to continue with the collection building process.

Creating a new collection/2

Another 'Create a new collection' pop-up window will come up. This screen shows the available metadata sets. You need to check the appropriate metadata set check box. Let us select the 'Dublin Core 1.1 (dc)' metadata set for our collection. Once you check the Dublin core metadata set, the elements (fields) within this set will be displayed in the lower half of the screen. The corresponding screen shot is shown below.



Assigning a Metadata Set

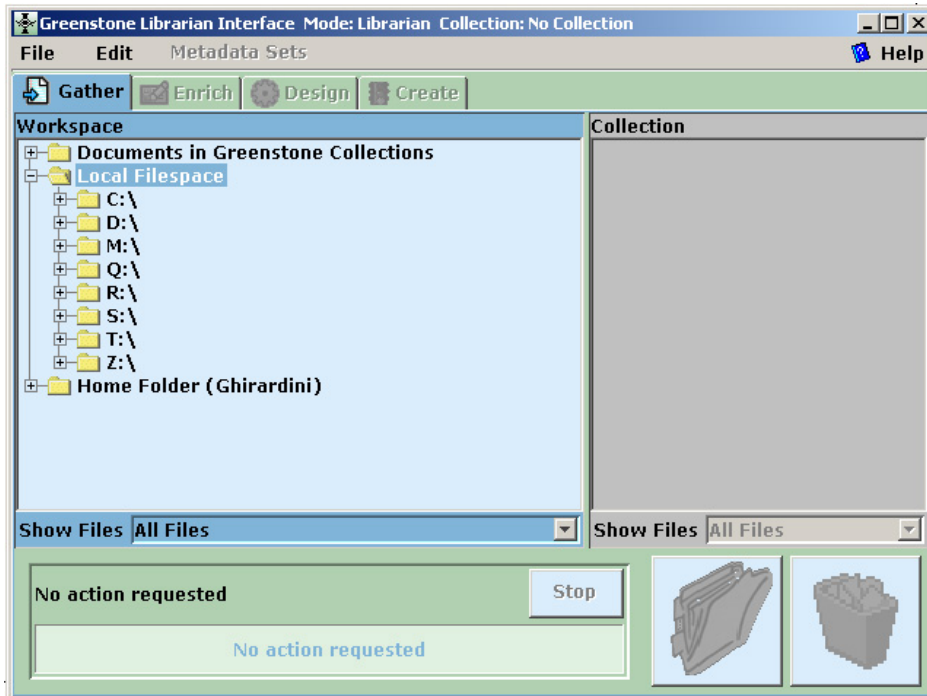
Click on OK to continue. Remember to save your work regularly by opening the 'File' menu and clicking on the 'Save' option.

You have now created a new collection called the 'WordPDF' Collection. The next logical step is to put some files (input source documents) into the collection.

GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

Putting documents into the collection

The input files can come from existing Greenstone collections, local file space (located on disc drives c:, d:), etc. When creating our collection, we will use input files from the local file space. Double click on 'Local File space'. The root directories of local drives, as shown below will be displayed.



Listing of Local File Space

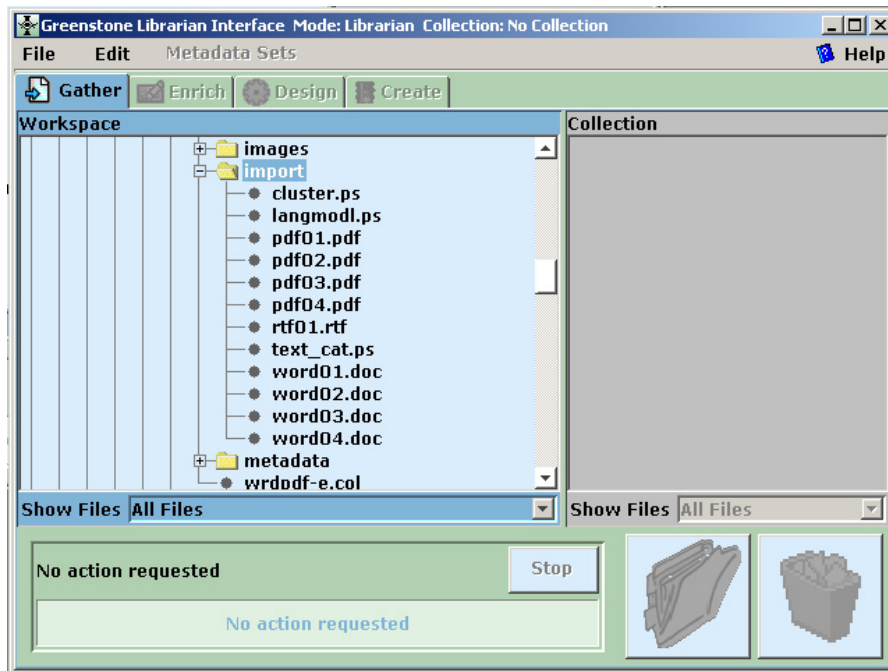
The display you see on your system depends on what hard disk drive partitions you have on your PC.

GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

Now you can import the files you need for your collection. For example, let's copy all the files of an existing Greenstone demo:

- go to: CD drive/software_tools/Greenstone/collect/wrdpdf-e, and
- click on the folder named 'import'.

All the files in this folder will be displayed as shown in the figure below.

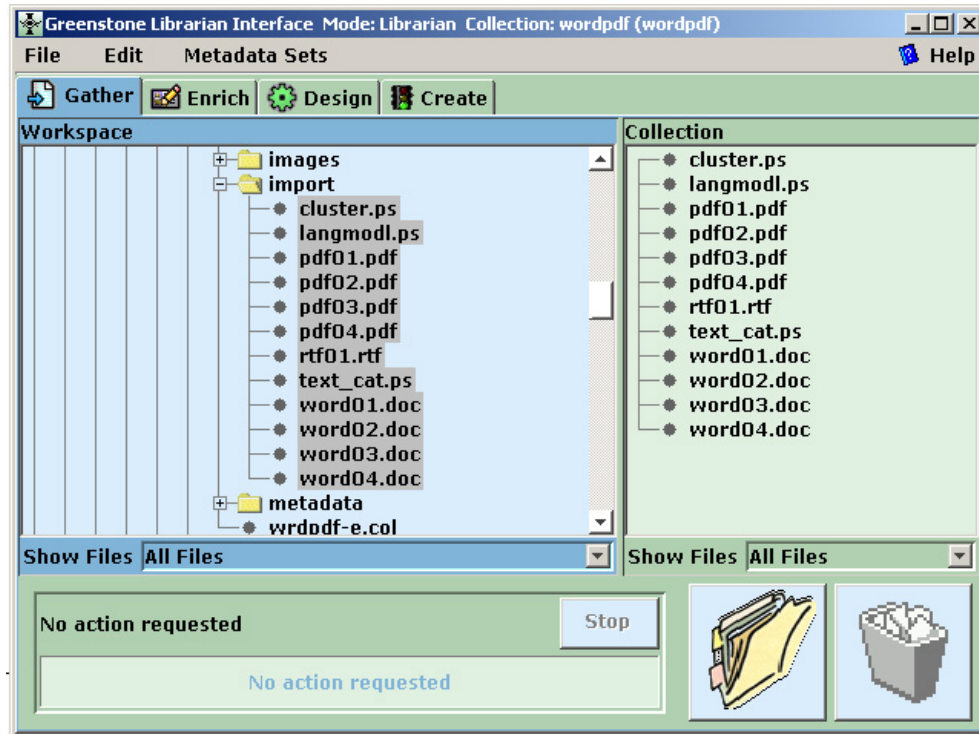


Listing of Source Files for WordPDF Collection

The Gather view has two areas called 'Workspace' and "Collection'. In the workspace, we see a list of files from the import directory in the self-learning CD. These files should be copied into the collection area by dragging and dropping.

GLI STEP 1: COLLECTION DEFINITION AND GATHERING OF DOCUMENTS

You can drag and drop one file at a time or multiple files at a time. Make sure you have dragged and dropped all the files. Once this process is complete, the 'Gather View' will appear as shown in the following figure.

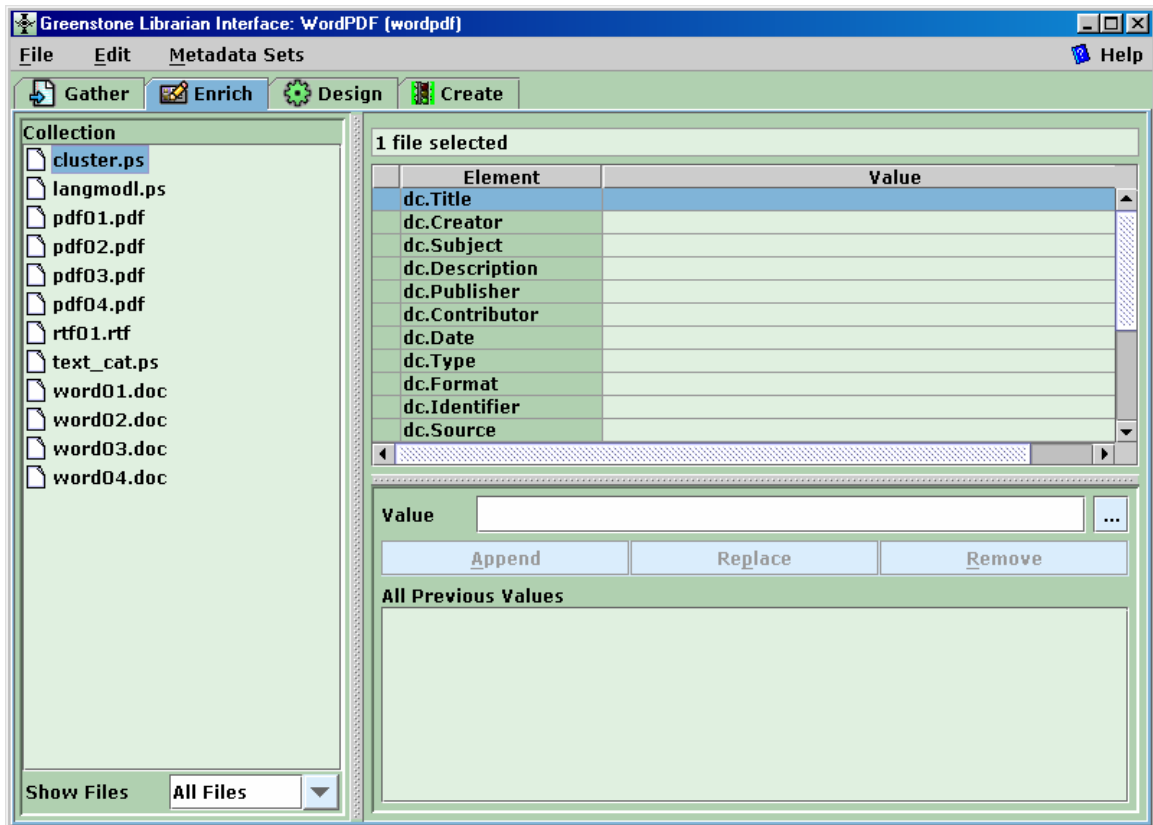


Gather View with Collected Files

GLI STEP 2: ENRICHING THE COLLECTION WITH METADATA

After gathering the files into the collection, you now need to enrich them by providing suitable 'metadata' information. Content in these metadata fields serve the very important purpose of defining the indexing, search, browse and display features for the collection in the **Design** step.

- Click on the GLI's **Enrich** tab at the top. The 'Enrich View' as shown below will be displayed:



Enrich View

This view shows the input source documents on the left and the Dublin Core metadata fields' template on the right. Lets us add metadata information to one of these documents.

- Click once on the **word03.doc** file.

The metadata element names will be displayed in the right side.

Just above the metadata names the display will let you know how many files have been selected. In our case it will show as 1. Open the file 'word03.doc' by double clicking it. This file will be opened in Microsoft WORD application (*Note that to open a Word, PDF, Postscript or file in any other format, respective application packages should be installed on your PC*).

Let us now add metadata information for dc.title and dc.author.

GLI STEP 2: ENRICHING THE COLLECTION WITH METADATA

Step A – Adding dc.title metadata

In the Word document, select the title of this paper (by highlighting the title text) and copy the selected text (using 'Control C' or right mouse click and 'copy' or by selecting 'copy' from the Edit menu). Go back to the GLI's enrich view and click on dc.title. Paste the copied text in the 'value' box. You cannot use the mouse's right button to paste. You will have to use the GLI's Edit option and click on the paste option to paste the value into the box (or use 'Control V'). Click on Append to add the metadata for dc.title field.

Step B – Adding dc.creator metadata

This document has four authors. Let us add them one by one as four different values for dc.creator. Click on dc.creator to select this field. Copy the first author (from the Word document) and paste it in the value box and click on Append to add the first author. Similarly add the other three authors. *It is important to click on the Append button after you enter the metadata value – otherwise this value will not be added to the metadata record.*

Add values for other metadata fields (subject, description, etc.) to this record. Repeat this process for the other documents in the collection.

Once you have finished adding the required metadata, the 'Enrich View' will appear as shown here:

The screenshot shows the Greenstone Librarian Interface (GLI) window titled "Greenstone Librarian Interface: WordPDF (wordpdf)". The interface has a menu bar with "File", "Edit", "Metadata Sets", and "Help". Below the menu bar are four tabs: "Gather", "Enrich", "Design", and "Create". The "Enrich" tab is active. On the left, there is a "Collection" pane listing various files: cluster.ps, langmodl.ps, pdf01.pdf, pdf02.pdf, pdf03.pdf, pdf04.pdf, rtf01.rtf, text_cat.ps, word01.doc, word02.doc, word03.doc, and word04.doc. The main area displays "1 file selected" and a table of metadata elements and their values. The table has two columns: "Element" and "Value". The current value for "dc.Title" is "Greenstone: A Comprehensive Open-Source Digital Library Software System". Below the table, there are three buttons: "Append", "Replace", and "Remove". Below these buttons is a section titled "All Previous Values" which shows the current value for "dc.Title".

Element	Value
dc.Title	Greenstone: A Comprehensive Open-Source Digital Library Software System
dc.Creator	David Bainbridge
dc.Creator	Ian H. Witten
dc.Creator	Rodger J. McNab
dc.Creator	Stefan J. Boddie
dc.Subject	Computer Science
dc.Subject	Digital Library
dc.Description	This paper describes the Greenstone digital library software system
dc.Publisher	
dc.Contributor	
dc.Date	