

CREATING DIGITAL LIBRARIES BASED ON CDS/ISIS DATABASES

by Pablo Morete and John Rose

This guide is intended to help users of UNESCO's CDS/ISIS software¹ to convert their databases to digital libraries under the Greenstone Digital Library program.

It is strongly recommended that users wishing to convert CDS/ISIS databases to Greenstone digital libraries utilize Greenstone version 2.70 or later (version 2.70 is available on the 2006 version of the UNESCO Greenstone CD-ROM or the latest version of Greenstone can be downloaded from the website at <http://www.greenstone.org>). Some conversion functionality can be obtained on earlier version of Greenstone,² but the conversion is likely to be more difficult and the results less satisfactory.

Like all Greenstone applications, libraries converted from CDS/ISIS can be disseminated on multiple Web platforms (Windows, Unix, Linux, Mac OS X) or on CD-ROM. Two types of conversion are possible:

1. "As is" conversion of a CDS/ISIS database to a Greenstone digital library. Since CDS/ISIS records are limited to 32,000 characters (in version 1.5), CDS/ISIS databases generally do not contain the full text of documents; this first type of conversion is thus normally used to provide easier Web or CD-ROM access to a bibliographic or referral database (through indexing or browsing on any of the CDS/ISIS fields). If any of the CDS/ISIS fields contain hyperlinks to external resources, they will be active in the resultant Greenstone library. The Greenstone library cannot be enlarged or edited in Greenstone; it will typically be generated periodically from the master CDS/ISIS database.

¹ <http://www.unesco.org/webworld/isis>

² The ISISPlug plugin enabling search and display of CDS/ISIS metadata in Greenstone has been available since version 2.50, and the explode function, enabling the creation of CDS/ISIS databases which can be updated in Greenstone and the integration of the full-text documents corresponding to bibliographic records, has been available since version 2.60. The following major improvements are available only in versions 2.70 and above:

- * Correct handling of "pseudo-repeatable" CDS/ISIS fields (non-repeatable fields with occurrences delimited by "<" and ">" (Indexing technique 2) or by "/" and "/" (Indexing technique 3). Prior to version 2.70 only Indexing technique 2 is handled, but only for a CDS/ISIS field named "Keywords".
- * Generation of a proper record display by ISISPlug. Prior to version 2.70 extensive reformatting was necessary for "as is" conversions of bibliographic databases (as in the Greenstone CDS/ISIS example application) unless the bibliographic records are in the CDS format of UNESCO.
- * The "replace" function enabling full-text documents associated with bibliographic records to be easily integrated into the Greenstone library of bibliographic records. Prior to version 2.70, if the documents were imported at the time of digital library creation from a CDS/ISIS field, each new document must be manually specified in the metadata.xml file. Those who, despite the shortcomings, want to continue to use versions 2.62 or 2.63 can obtain some of the 2.70 functionality by downloading i) the updated ISISPlug plugin at <http://www.cs.waikato.ac.nz/~mdewsnip/greenstone/temp-2.63/ISISPlug.pm> to replace the existing "ISISPlug.pm" in the "Greenstone\perl\plugins" directory, normally found under C:\Program Files\ in Windows, and ii) the updated "explode" program at http://www.cs.waikato.ac.nz/~mdewsnip/greenstone/temp-2.63/explode_metadata_database.pl to replace the existing "explode_metadata_database.pl" file in the "Greenstone\bin\script" directory, normally found under C:\Program Files\ in Windows.

2. Creation of a Greenstone digital library of full-text documents from a CDS/ISIS bibliographic database and the files containing the full-text documents associated with the database records. This means that the original document can be imported into Greenstone (from a storage device available to your computer or from the Internet); and that the text of the document and/or the CDS/ISIS metadata can be accessed through indexing or browsing on any of the CDS/ISIS fields. The resulting digital library can then be updated/maintained as an autonomous application, or if preferred can be periodically regenerated from the master CDS/ISIS database with associated documents. This method can also be implemented with dummy documents to enable a library based only on CDS/ISIS metadata to be enlarged or edited in Greenstone.

The second type of conversion is of particular interest for CDS/ISIS users, since it enables one to provide full digital library services (indexed access to original documents) while retaining all of the power and flexibility of CDS/ISIS in the handling of the corresponding metadata.³

Both types of CDS/ISIS to Greenstone conversion will be documented in the following two parts of this guide, while the third part will briefly address the issue of configuring the user interface once the database is converted. However, in order to follow these instructions, the user should first become familiar with the basic use of the Greenstone Librarian Interface (GLI) in creating digital libraries.⁴

The instructions and screen-shots in this guide are based on the default "Librarian" mode of operation of GLI. If another mode is used the appearance of the screens will be different, but the basic functionality will be the same.

1- IMPORTING A CDS/ISIS DATABASE INTO GREENSTONE

This procedure converts a CDS/ISIS database "as is" into a Greenstone digital library. It requires the ISISPlug plugin (loaded by default in a new Greenstone library or one based on the CDS/ISIS example collection).

The following steps should be followed:

- A- Load the Greenstone Librarian Interface (GLI) create a new collection by clicking on the **File/New...** main menu item, providing a short collection name (for convenience it could

³ Other FOSS/freeware options available for presentation of native CDS/ISIS databases "as is" on the Web include:

1. GenISIS is freeware available from UNESCO (<http://www.unesco.org/webworld/isis>) to create a customized server-end application for querying CDS/ISIS databases. It requires the WWWISIS software (also called WXIS from version 4.0) of the Latin American and Caribbean Center on Health Sciences Information (BIREME), but for serving under Windows, the freeware version of WWWISIS (version 3.0) is sufficient.
2. The JavaISIS package of UNESCO requires installation of the JavaISIS Server at the server end and the JavaISIS Client at the user end. Users can retrieve, create, modify and display records, and import and export records using the ISO2709 format. Multilingual encoding support is provided. Since JavaISIS also requires WWWISIS, it works only under a Windows based Web server unless the paid version of WWWISIS is acquired.
3. CLABEL can be used to serve a CDS/ISIS database over Linux/Unix. It requires OpenISIS and PHP-OpenISIS (all three are FOSS programs available at <http://www.sourceforge.org>).
4. Igloo (<http://igloo.lib.itb.ac.id/>) can be used either over OpenISIS and PHP-OpenISIS or over PHP-OpenIsis for Windows.

⁴ See, for example, Chapter 3 of the Greenstone User's Guide (<http://prdownloads.sourceforge.net/greenstone/User-en.pdf>) or the FAO IMARK training module on Digitization and Digital Libraries (<http://www.imarkgroup.org/>), followed as required by the more advanced sources in section 3 of this guide.

be the name of CDS/ISIS database) and a description. Keep the default setting for "Base this collection on:" as -- **New Collection** --, then click on **OK** (see Figure 1). You are proposed a list of metadata sets; uncheck Dublin Core Metadata Element Set and click **OK** and then **OK** again in the warning box (you do not need a metadata set for this step, metadata will be generated from the CDS/ISIS file structure).

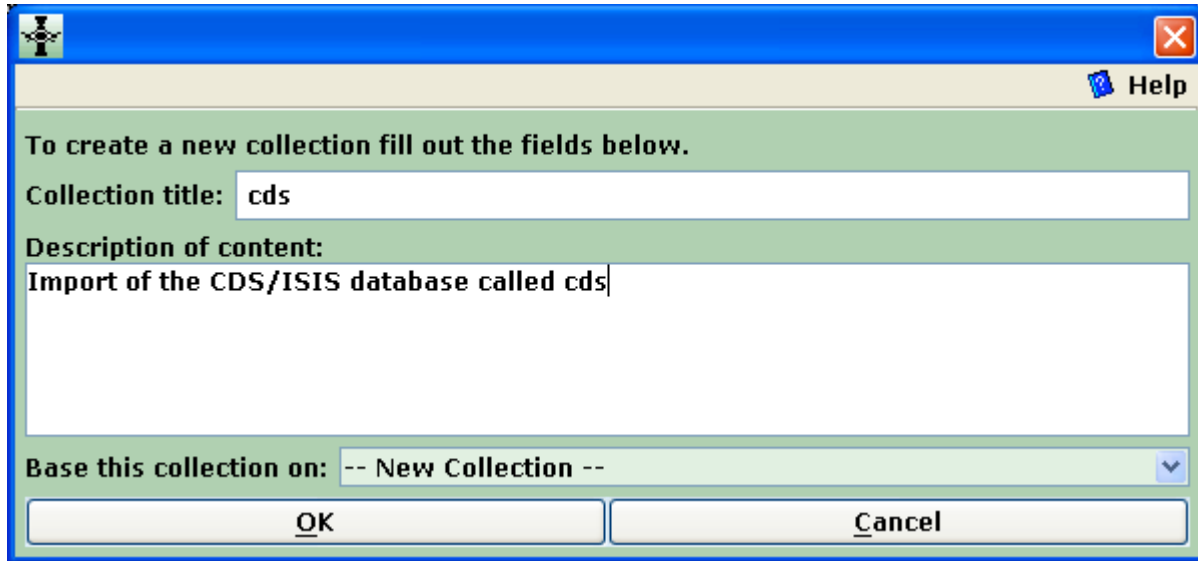


Figure 1: Creating a new collection

- B- The GLI **Gather** panel will appear. Locate the MST, XRF, and FDT files from your CDS/ISIS database in the directory tree of Workspace pane (normally they will be in the "C:\WINISIS\DATA\" directory of the Local Filespace) and drag them one by one into the **Collection** pane. When you copy MST file into the **Collection** pane Greenstone could propose to load ISISPlug (it will be loaded by default in version 2.70 or above); in that case accept the proposal. Note that if all three files do not show in the **Collection** pane, an error will be generated at the later "build" or "explode" steps; in that case just come back to the **Gather** panel, drag in the missing files and continue.⁵

⁵ For versions prior to 2.70, there may be a warning when dragging the XRF and FDT files to the effect that "None of Greenstone's plugins are expected to process the file" - simply click on **OK**.

You will then have the screen shown in Figure 2.

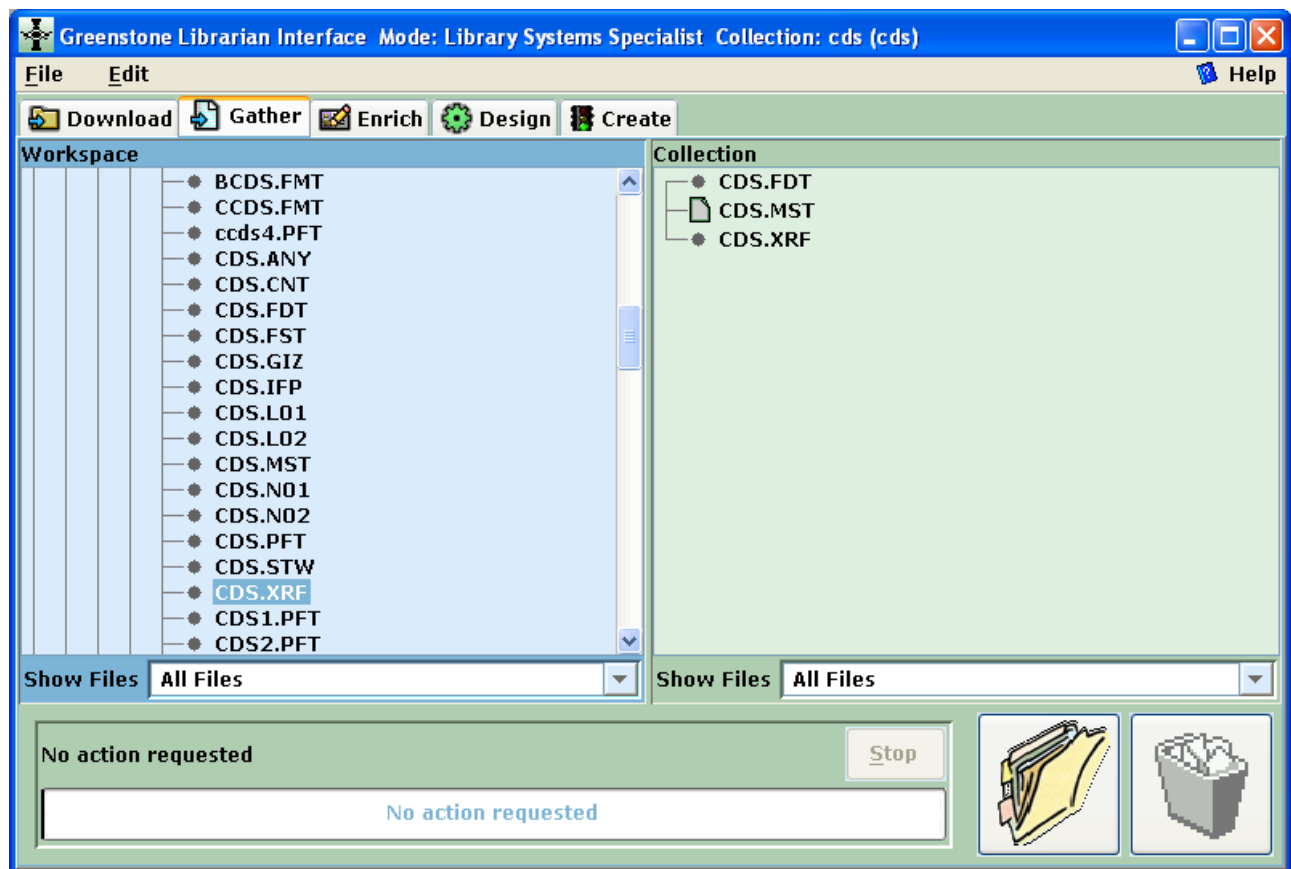


Figure 2: Collection files ready to process

- C- If you are using non-ASCII characters in your database, go to the **Design** panel and click on Document Plugins. Then click on ISISPlug and on **Configure Plugin** (see Figure 3). Check **input_encoding** and select the appropriate DOS codepage (for Latin alphabets using accented letters, it is "dos_850-DOS codepage 850 (Latin 1)") and click **OK**. This will set ISISPlug to correctly recognize the character set of your database (see Figure 4).

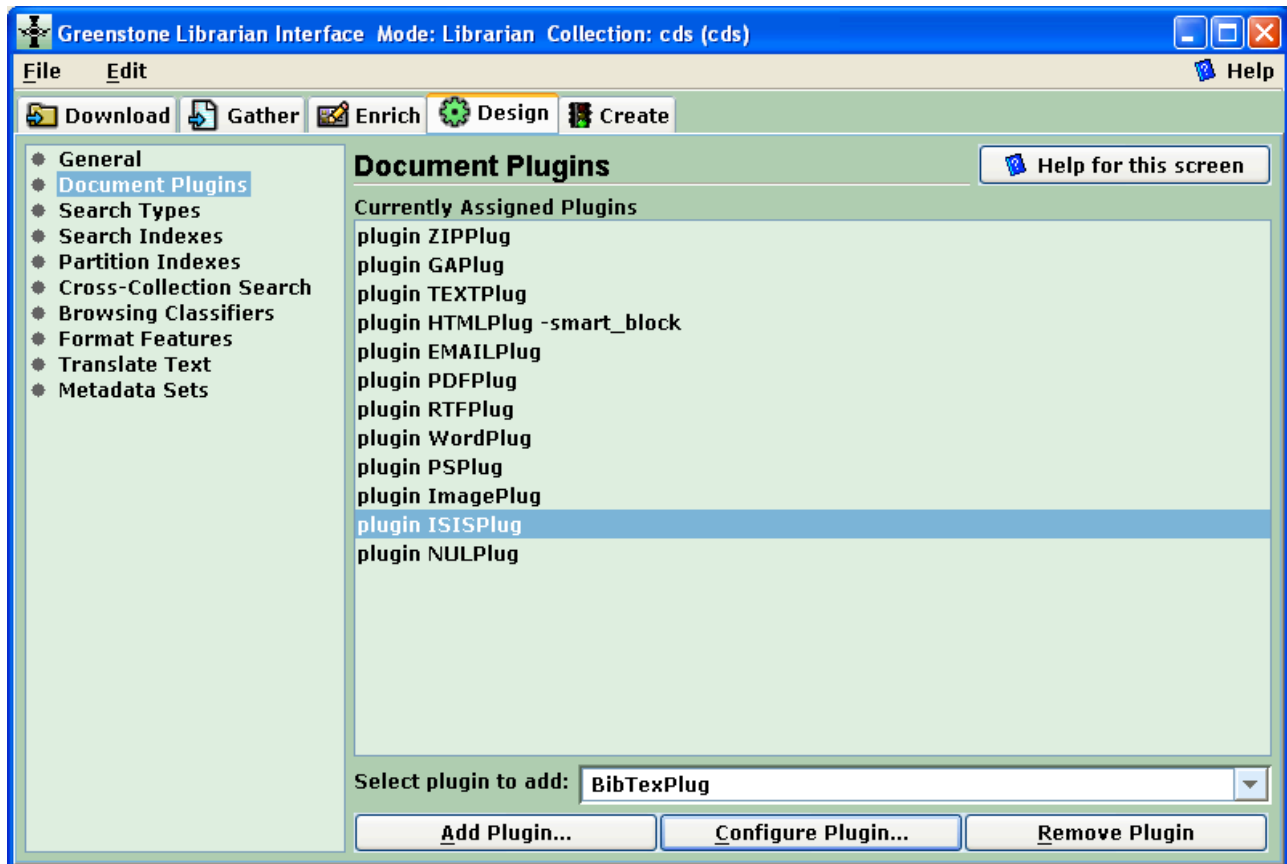


Figure 3: Select the ISISPlug under Document Plugins in the *Design* panel

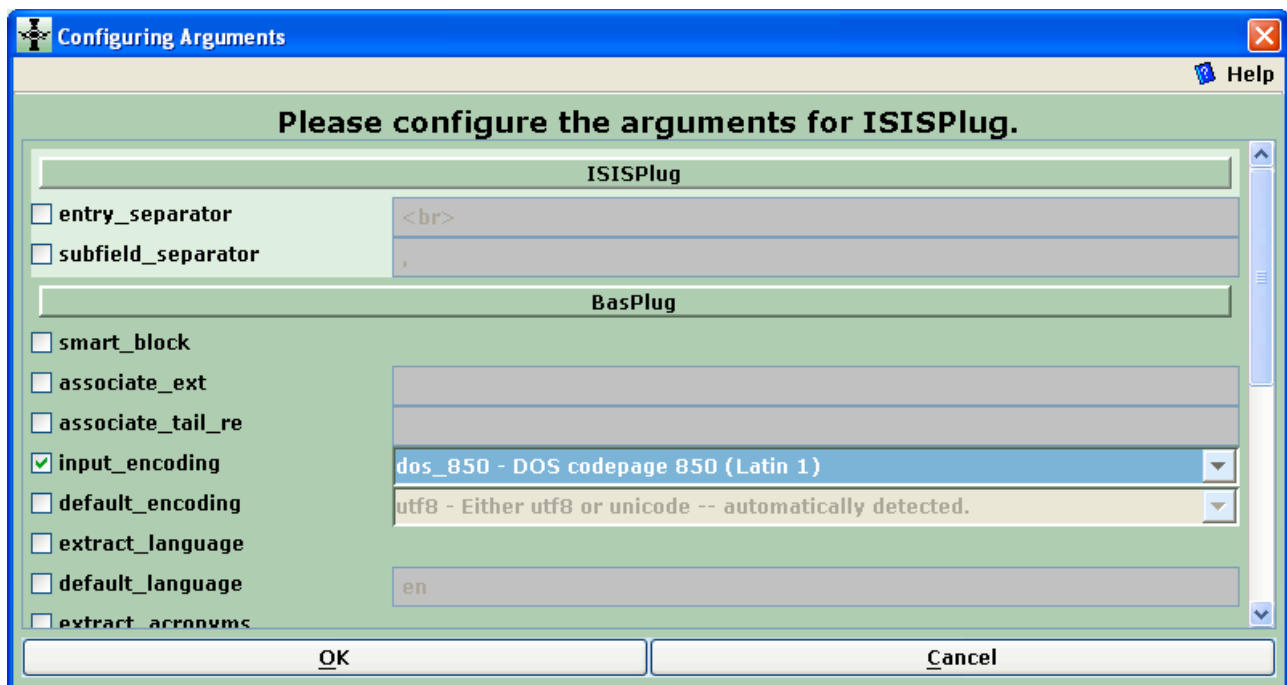


Figure 4: Select the appropriate character encoding scheme

If either i) you want to incorporate and index electronic document files into your Greenstone library or ii) you want to edit metadata in Greenstone (not enabled for "as is" conversion which is intended only for Web display or CD-ROM distribution of the CDS/ISIS database), then go to Part 2 on "Assigning CDS/ISIS Metadata to Electronic Documents". If on the other hand your CDS/ISIS database is a self-contained resource database pointing to paper documents, or to electronic documents which you do not need to physically integrate into the digital library,⁶ continue with the steps just below.

D- Go to the *Create* panel of GLI and click on **Build Collection** (see Figure 5).^{7,8}

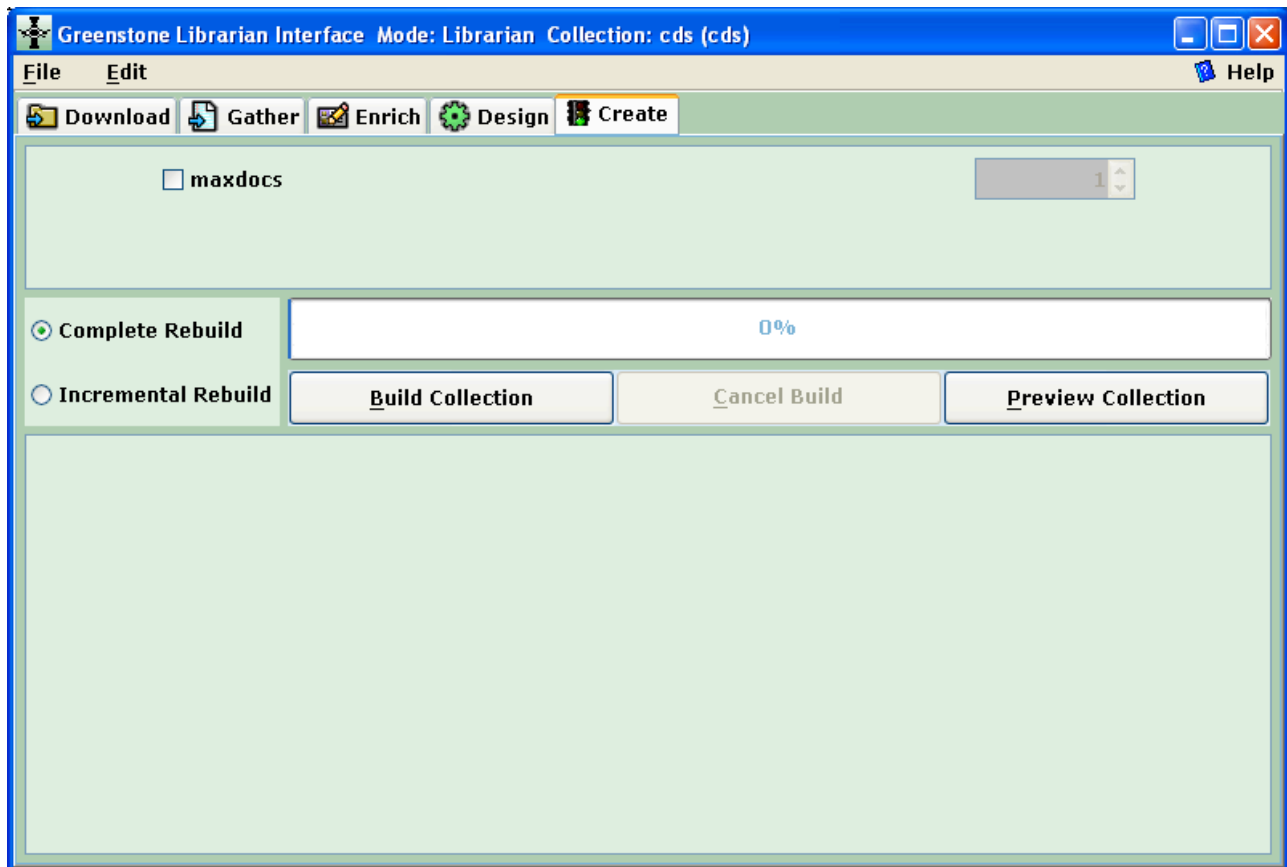


Figure 5: The **Build Collection** button in the *Create* panel

E- Normally after several seconds or minutes (depending on the size of your database, it could be even longer), a box may appear informing you that the collection has been built (see Figure 6), and you may then preview it by clicking on the "Preview Collection" button

⁶ Integration of the documents is required to index their full text within Greenstone, and may be useful to package the whole library and to compress the files; however, in an "as is" conversion, files identified in the bibliographic records by valid url's will be accessible from the Greenstone library.

⁷ Note that it is only after the collection is built the first time that the CDS/ISIS metadata are extracted and available to create indexes, browsing classifiers and display formats.

⁸ If your CDS/ISIS collection has several thousand records, it could be useful to switch to "Library Systems Specialist" mode, and to set the *groupsize* parameter to say, 200 before clicking on **Build Collection**. Bibliographic collections typically have many small documents, and grouping them together prevents Greenstone's internal file structures from becoming bloated and occupying more disk space than necessary.

(This step can be eliminated in the future by checking the "Do not show this message again" box and clicking on **OK**).

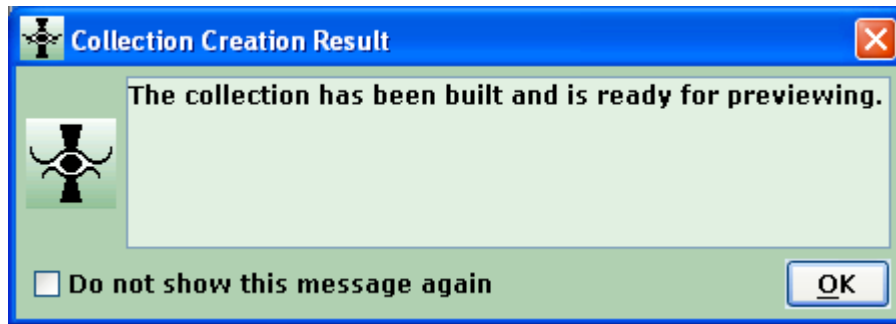


Figure 6: Box announcing that the collection has been built

- F- By default the new collection will be in simple search mode (single search index) with two browsing classifiers ("titles"⁹ and "filenames"¹⁰) and a standard record display format listing the CDS/ISIS field names and field content in order of CDS/ISIS tag number.¹¹ To provide an appropriate user interface to the collection, you will now need to configure your search types, search indexes and browsing classifiers in the *Design* panel.¹² This is a task common to all Greenstone collections, for which some specific guidance and general references are given in Part 3.

It is possible to convert more than one CDS/ISIS databases into a single Greenstone collection, combining the metadata records of each. To do this, simply drag the files of all of the source databases (which can have different metadata structures) into the *Collection* pane before building (or rebuilding).

BASING THE COLLECTION ON THE GREENSTONE CDS/ISIS EXAMPLE:

The CDS/ISIS example collection of Greenstone (called isis-e) was developed for version 2.50, before the explode functionality existed and before a formatted version of the CDS/ISIS record was included as the record text in an "as is" conversion (previously it was a raw record, see Part 3, Section E under *DocumentText*). This collection is now of interest mainly for didactic purposes, unless you are planning only an "as is" conversion and i) the metadata structure of your CDS/ISIS database is similar to that of the example collection (based on the sample database provided with CDS/ISIS which uses the bibliographic format of UNESCO's library) and ii) you wish to display your data as a user defined record display format with an option to additionally show or hide the full CDS/ISIS record.¹³

⁹ This classifier will work correctly by default only if the CDS/ISIS database has a field entitled "Title".

¹⁰ The filename for all of the records is the MST file, so this classifier is not useful.

¹¹ In collections built with Greenstone versions prior to 2.70 (including the isis-e example collection), by default the records will appear in raw format in which tag names and field data are strung together without line breaks between them; the improved standard presentation can be obtained by rebuilding these collections in version 2.70.

¹² If you are using a version of Greenstone prior to 2.70, you will also probably want to modify the display format features (see Part 3, Section E).

¹³ The example collections can be downloaded from <http://prdownloads.sourceforge.net/greenstone/gsdl-documented-collections-aug2005.zip> (if your CDS/ISIS example collection came with a version of Greenstone prior to 2.62, for example the UNESCO CD-ROMs published in 2004 (Greenstone version 2.50) or 2005 (Greenstone version 2.60), it will only work properly in version 2.62 and above if you download and install the updated version of the Greenstone\collect\isis-e\etc\collect.cfg file at <http://www.cs.waikato.ac.nz/~mdewsnip/greenstone/collect.cfg>).

To apply this method, specify when you create the Greenstone collection that it is to be based on the "CDS/ISIS example (isis-e)" by selecting this in the drop-down list for the parameter: "Base this collection on:" in the initial screen for creating the collection (see Figure 1). If your fields correspond exactly to those in the example, your library will work just as the example after you complete steps B-E above (provided that, for versions of Greenstone prior to 2.70, you replace "isis-e" in the two "&c=isis-e" specifications of the 'format DocumentText' statement (**Design** panel, **Format Features**) by the short name of your Greenstone collection (see Figure 7, and also Part 3, Section E concerning editing of formats in the **Format Features** view).

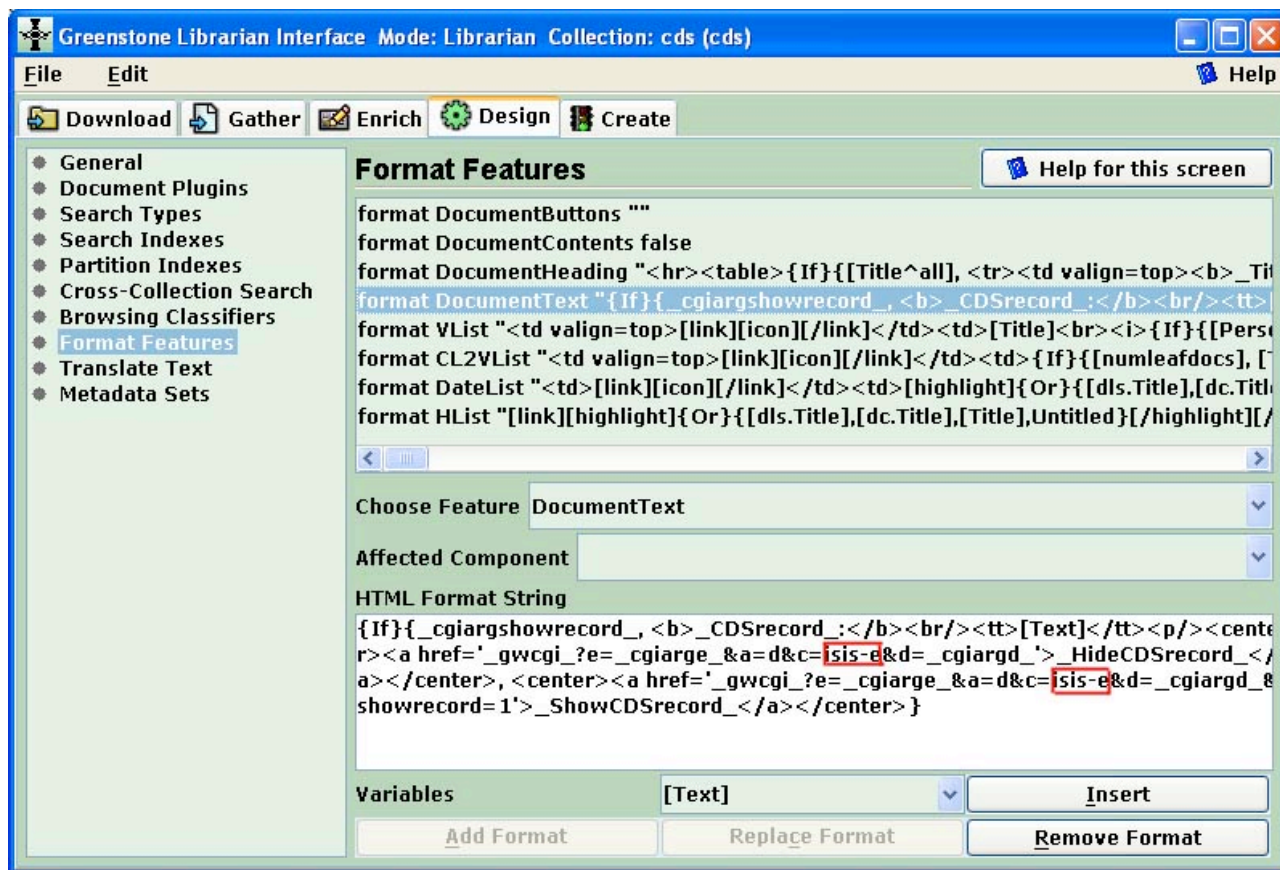


Figure 7: Format parameter to change (shown within the red box in **HTML Format String**)

If your CDS/ISIS fields differ from the example (based on UNESCO's bibliographic database), the full record display will be correct, but the record initially displayed will only show those fields whose names are identical to those of the example (and nothing if none are the same). In this case you should modify the initial user defined record display format by editing the **Document Heading** format in the **Design** panel, **Format Features**, as explained in the Bibliography collection (cltbib-e) (see also the general guidance on formatting in in Part 3, Section E, but note that, the **Document Heading** format will display starting from its end and has to be scrolled to backwards to be edited). Note also that no records may appear for display if the search indexes and browsing classifiers have not been customized (see Part 3); in that case you can fetch the records by searching on "raw record" in the search form.

Remember that if you are going on to assign CDS/ISIS metadata to electronic documents (second type of database conversion), there is no need to improve the display interface of the simple "as is" Greenstone collection based only on metadata, or to base your collection on the Greenstone example which is only for metadata display.

2- ASSIGNING CDS/ISIS METADATA TO ELECTRONIC DOCUMENTS

The "Explode Metadata Set" option" provides a way of reorganizing a Greenstone collection consisting of metadata only (e.g. "as is" conversion of a CDS/ISIS database) so that each record appears as an individual document with the associated metadata assigned to it. The explode option is a functionality of the ISISPlug plugin, which must be loaded as in the case of an "as is" conversion.¹⁴

Exploding metadata is an irreversible process,¹⁵ so that if you have built an "as is" CDS/ISIS library in Greenstone, and want to keep the data and/or the configuration (e.g. search types, search indexes, browsing classifiers, and display formats), save it before going on with this step (this is most easily done by duplicating the entire collection with another name in the Greenstone\collect\ directory).

In the *Gather* panel, you will notice that the MST file has a different coloured icon than the other files. This green icon indicates that the file it is a metadata database that can be exploded. Right-click on the icon and choose *Explode metadata database* by right clicking on this line in the menu (see Figure 8).

¹⁴ The explode option also works with other plugins designed to handle metadata databases, like BibTexPlug or MARCPlug.

¹⁵ The explode function erases the copies of the CDS/ISIS files which was dragged into GLI, but not the originals in your CDS/ISIS database (in versions prior to 2.70 only the MST file is erased).

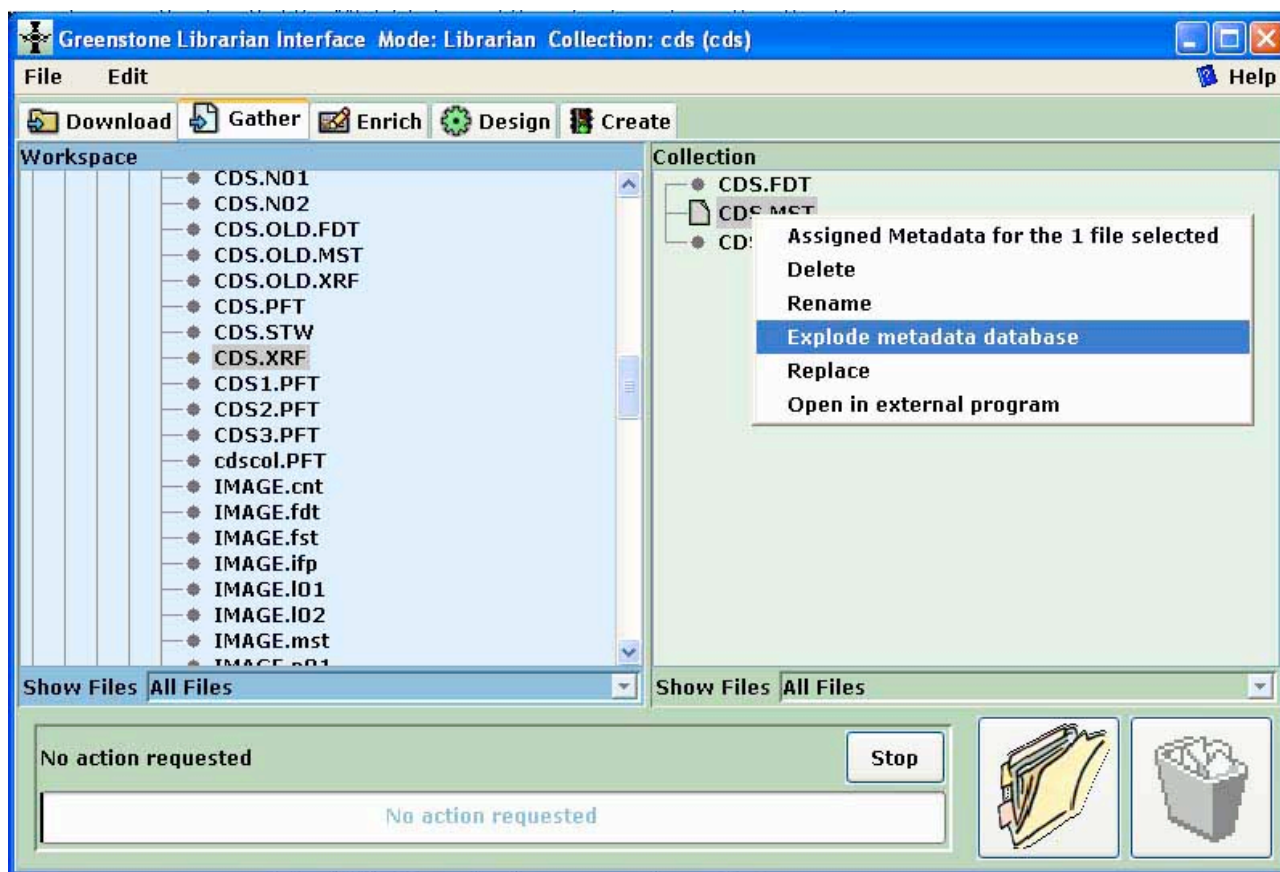


Figure 8: Menu presented after right click on the MST file

There are now two ways to integrate the electronic documents into the collection: A) automatically importing them through a hyperlink existing in the CDS/ISIS database or B) creating dummy files which are later replaced by the corresponding documents.

A- Automatic importation of the electronic documents

When the ***Explode Metadata Database*** window opens, the explode parameters including the CDS/ISIS field containing the hyperlink should be specified as in Figure 9.

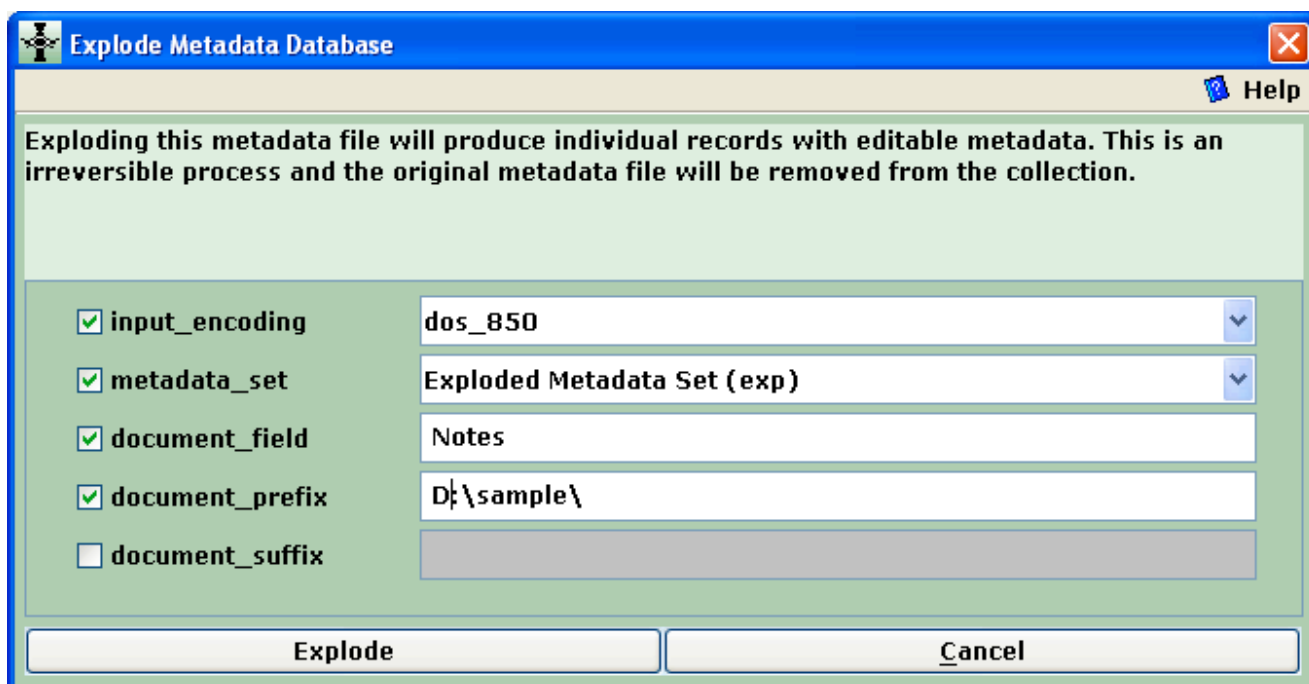


Figure 9: Specifying the *Explode Metadata Database* parameters

- ***input_encoding***: If you are using non-ASCII characters in your CDS/ISIS database select the appropriate character set from the dropdown menu (for Latin alphabets it is "dos_850").
- ***metadata_set***: This parameter (available only in version 2.70 and above) should be selected unless you want to ignore or combine some of the CDS/ISIS fields.
- ***document_field***: Indicate the label (name) of the CDS/ISIS field containing the file name (in which case the ***document_prefix*** parameter is also required, as in this example) or the full path and filename of the electronic version of the document associated with the record (the document can be in any format accepted by Greenstone: htm, pdf, doc, ppt, etc.). Here we have specified the "Notes" field of the example CDS/ISIS application in which the document names have been added.
- ***document_prefix***: Indicate the path (if any) to be prefixed to the contents of ***document_field***.
- ***document_suffix***: This parameter could contain the file extension if not included in the content of ***document_field***.

The full document file path (concatenating ***document_prefix***, ***document_field*** and ***document_suffix***) may be either valid path on the local computer or network or a valid url on the Internet.

Leave the other options blank and click on ***Explode***, then click on the ***OK*** button when informed that the explode process has been completed.¹⁶

¹⁶ If the ***metadata_set*** parameter has not been set in versions 2.70 and above, or if you are using version 2.6x, you will first be prompted to add, merge or ignore each of the CDS/ISIS fields through a dialogue box. For each metadata element, click on the ***Add*** button to assign data to it, the ***Merge*** button to combine this data with a ***Target metadata element***, or the ***Ignore*** button to ignore the element. Then click on the ***OK*** button when informed that the explode process has been completed.

At this point the full-text document file corresponding to each record will have been copied into the "Greenstone\collection\xxx\import\YYY\" directory, where xxx is the name of the Greenstone database (in this case "cds") and YYY is the name of the CDS/ISIS database (in this case "CDS"). The metadata for the records has been written to a new "metadata.xml" file in the "Greenstone\collection\xxx\import\YYY" directory, and the metadata for each record is available for editing in the **Enrich** panel; the metadata elements will have the same names as the corresponding CDS/ISIS fields, preceded by the prefix "exp." (as opposed to ".ex" when an "as is" conversion is done).¹⁷

B- Creation and replacement of dummy documents

If the user finds it inconvenient to place the paths and filenames of all or some of the associated electronic documents in the CDS/ISIS database,¹⁸ Greenstone can substitute the missing documents by dummy document files (files with zero length and ".nul" extension). In this case, the bibliographic metadata are attached to the dummy file. As and when the full documents become available, the "replace" function can be used to copy them into the collection to replace the corresponding dummy files.

The procedure is essentially the same as in Section A above. If the full document file path (concatenating *document_prefix*, *document_field* and *document_suffix*) is not valid, a dummy file will be created with filename derived from the concatenation and "nul" extension.¹⁹ If the concatenation is a null string (no data in *document_prefix*, *document_field* or *document_suffix*), dummy files will be sequentially created for the concerned records with filenames names 0001.nul, 0002.nul, etc. As before the metadata for each record is available for editing in the **Enrich** panel; the metadata elements will have the same names as the corresponding CDS/ISIS fields.

The library can now be configured, built and used just as if it contained actual documents (The NULPlug plugin must be loaded to process dummy documents, but this will normally not require action since NULPlug is loaded by default unless the collection is modeled on one without it.). When the document corresponding to a dummy file is available, one should right-click on the dummy file and then left-click on "Replace" as shown in Figure 10.²⁰

¹⁷ Prior to version 2.70 the exploded metadata was not automatically fixed into an editable metadata set. In versions 2.6x use the following method to be able to edit the metadata: Open the separate Greenstone Editor for Metadata Sets (GEMS) program and open a new metadata set (File/New... main menu item). Give it a full Name and a unique short name (Namespace) and click OK. Then select the metadata set in the left panel and click on the Save item in the File menu. A blank metadata set will be saved in a file named Namespace.mds (in this case, cds.mds). Now go back to your Greenstone collection in GLI and add the newly created blank metadata set in the Metadata Sets view of the Design panel. Add all of the extracted metadata set items one by one as prompted to the new dataset. The metadata with elements named Namespace.fieldname will now be editable in the **Enrich** panel.

¹⁸ For example, because the user wishes to set up a bibliographic database in Greenstone with the intention of gradually incorporating the associated documents, or simply because it is inconvenient to automate the transfer due to documents scattered among different storage units and directories.

¹⁹ In Greenstone versions 2.6x, there is an additional parameter called *filename_field* in the **Explode Metadata Database** window which is used to generate the dummy file names.

²⁰ This function is only available starting in version 2.70. In versions 2.6x it is necessary to delete the dummy file in the **Collection** pane of the **Gather** panel and drag in the document file. Then, before building, the user must use a text editor like WordPad to change the line "<FileName>filename.nul</FileName>" in the metadata.xml file (in the Program

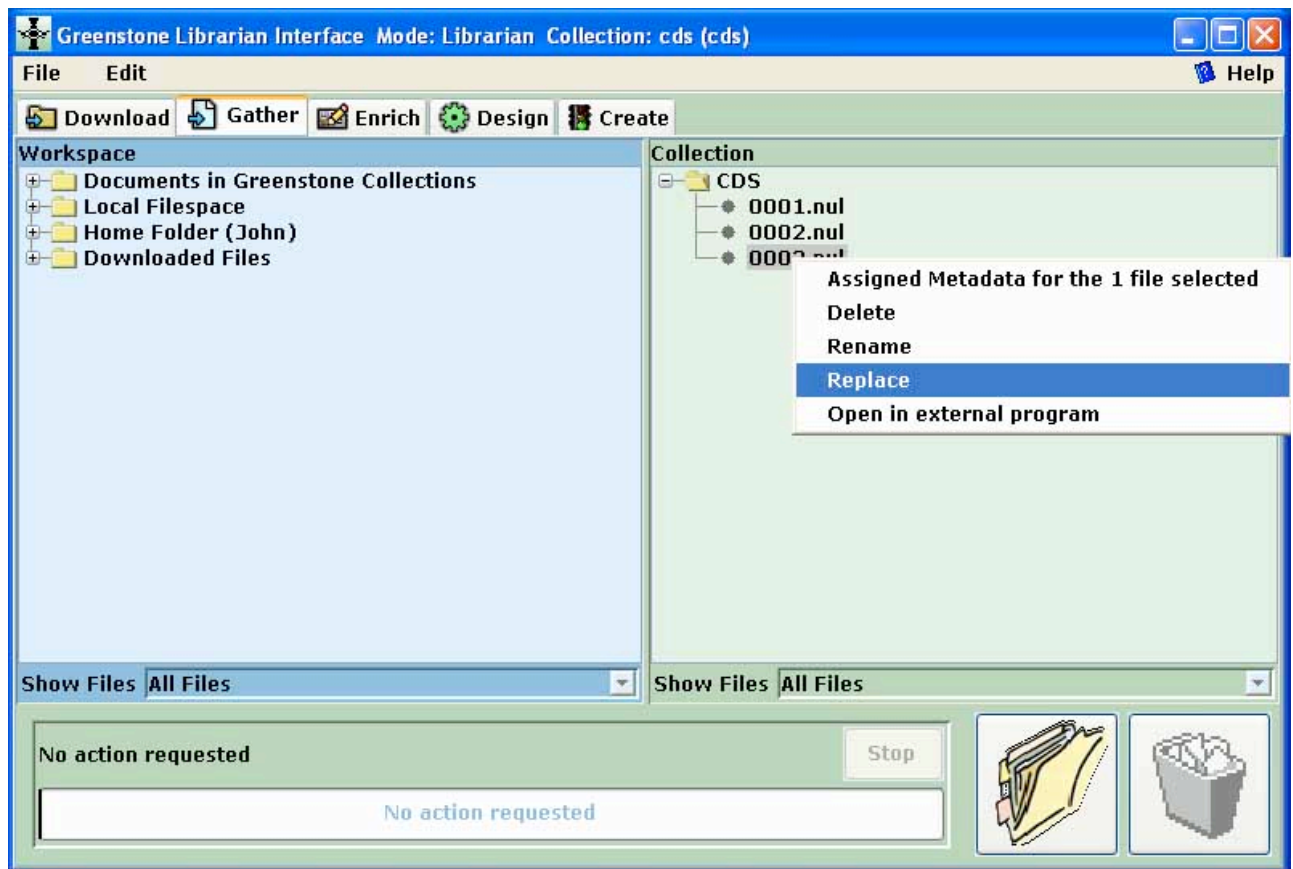


Figure 10: Menu presented after right click on a dummy file to be replaced

A browsing window (see Figure 11) will then enable the selection of the document file to replace the dummy file.

Files\Greenstone\collect\xxx\import\YYY directory) by replacing "filename\.nul" by the full file name (name and extension, e.g. "actualfilename\.doc") of the actual electronic document.

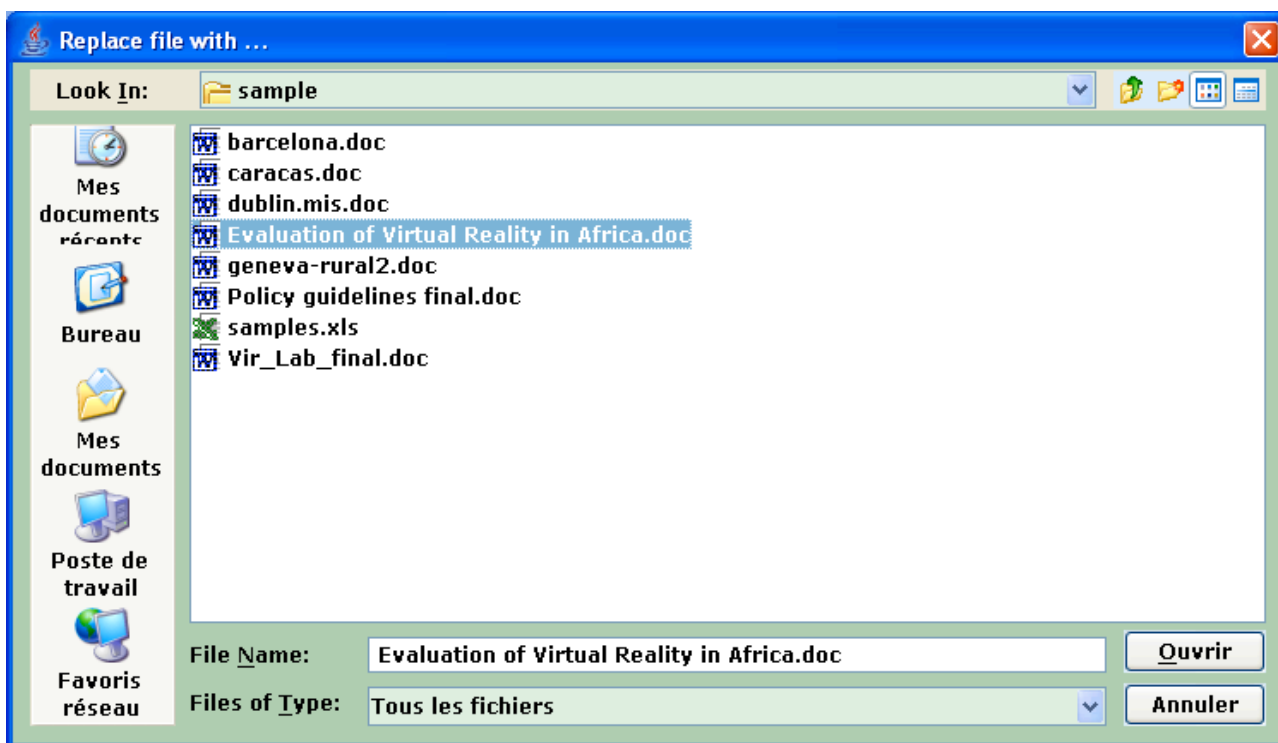


Figure 11: Choosing the document file to replace the dummy file

The collection can then be built, making sure in the **Document Plugins** view of the **Design** panel that the plugins needed to process the new documents (e.g. WordPlug and HTMLPlug for Word documents) have been loaded.

After the explode step as in either Section A or Section B above, you will be able to search and to browse the collection using the default interface. The collection can be finalized by configuring the search types, search indexes, browsing classifiers and format features in the **Design** panel (see Part 3).

It is possible to explode more than one CDS/ISIS databases into a single Greenstone collection, combining the documents of each and their metadata records. To do this, simply drag the files of an additional database (which can have different metadata structures) into the **Collection** pane and explode it. This technique can be used to update an existing collection with new documents (without having to rebuild the entire collection as is necessary when adding an additional database to an "as is" collection).

3- CONFIGURING THE USER INTERFACE

This guide cannot go into all of the details of how to configure the end-user interface for your collection converted from CDS/ISIS. This part will provide only guidance for obtaining a basic acceptable configuration of search types, search indexes, browsing classifiers and format features, as well as a list of Greenstone documentation for more detailed or advanced configuration. The discussion that follows will refer to the default configuration parameters that are obtained by creating a new "as is" collection (see step 1.A.) or exploded collection. If you base your collection on an

existing collection, the parameters will be those of the model collection (and thus in principle will be closer to those required for your collection and will require a lesser degree of modification).

Remember that, in Greenstone, search indexes and browsing classifiers (browsing lists) are different and must be specified separately. Any given metadata element may be indexed and/or be presented in a browsing classifier.

As shown in Figure 12, the *search* button is always at the upper left on the function bar of the collection homepage, and the browsing classifier buttons are to its right.

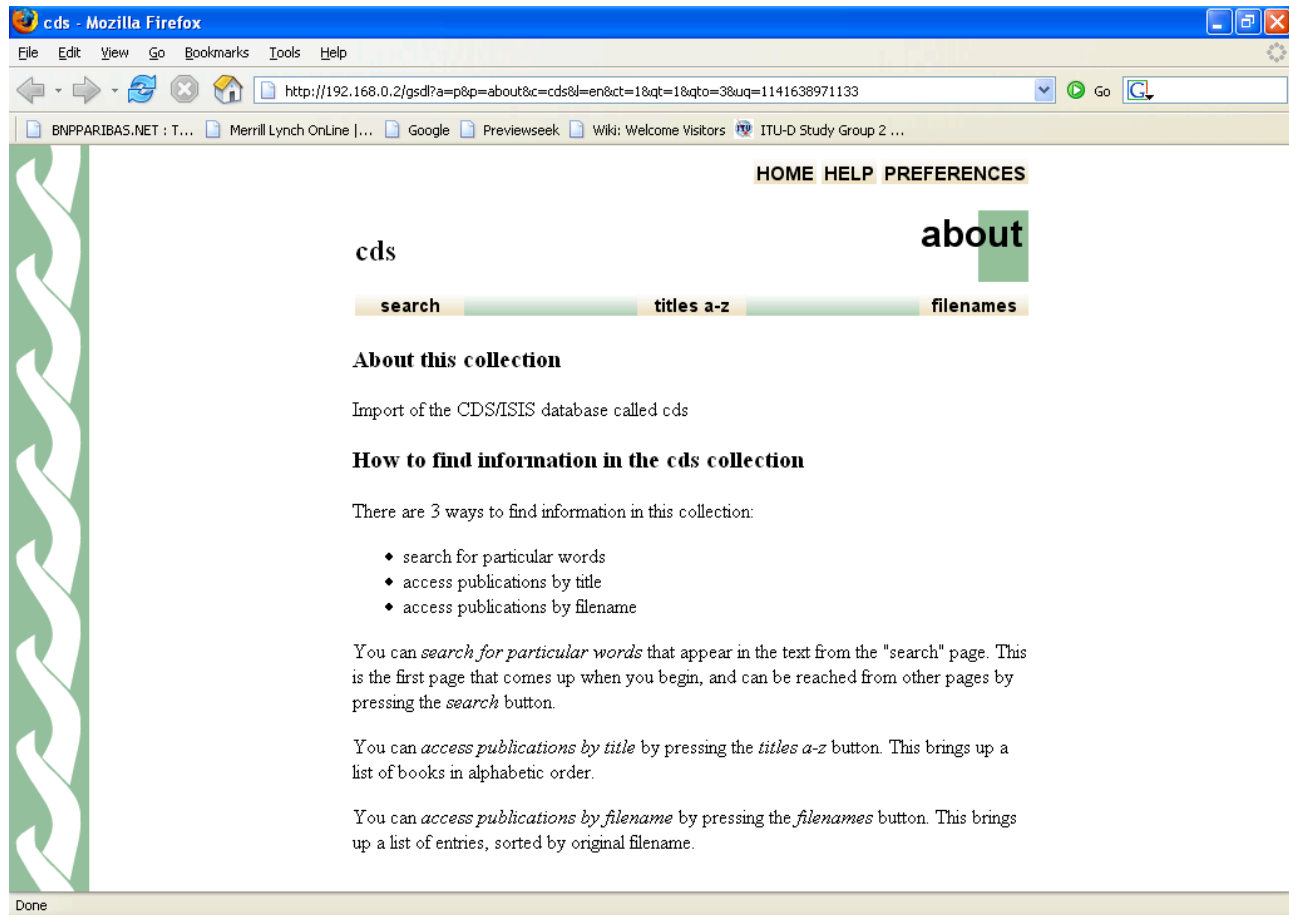


Figure 12: Homepage of the "cds" collection

A- Search Types

The default configuration is simple search (all search fields in the same index) and plain search type (a single box for search terms). In order to configure for advanced (multi-field) search with a the multi-field search form as default for the end user, go to the *Search Types* view of the *Design* panel and check the *Enable Advanced Searches* box. Then click on the *Add Search Type* button to add "form" search. Select "form" in the *Currently Assigned Search* types and click the *Move Up* button. You will then have the correct configuration shown in Figure 13.

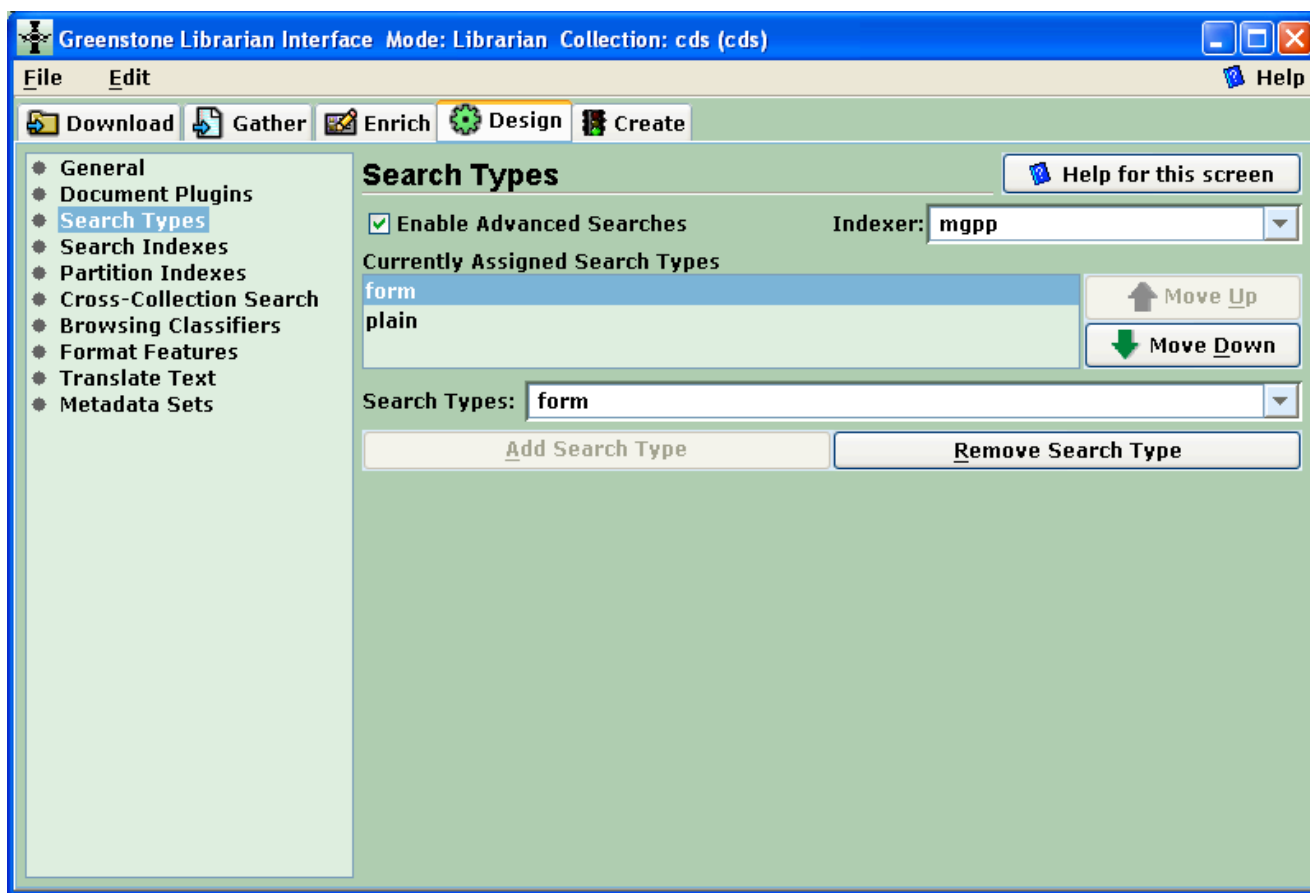


Figure 13: Parameters set for form search as default using multi-field searching

B- Available Metadata

To be able to follow the instructions in the following sections it is necessary to be able to recognize the metadata elements made available by Greenstone for the search indexes, browsing classifiers and display formats. For a given CDS/ISIS fieldname, the metadata extracted in an "as is" conversion have the prefix "ex" (e.g. ex.fieldname) and exploded metadata have the prefix "exp" (e.g. exp.fieldname).

In the case of a CDS/ISIS repeatable field, the following metadata elements are generated:

| | |
|--|--|
| ex.fieldname (or exp.fieldname): | The individual occurrences of fieldname |
| ex.fieldname^all (or exp.fieldname^all): | Delimited list of all of the occurrences |

In the case of a CDS/ISIS field with subfields, the following metadata elements are generated:

| | |
|--|--|
| ex.fieldname^a (or exp.fieldname^a): | Subfield "a" |
| ex.fieldname (or exp.fieldname): | Delimited list of all of the subfields |
| ex.fieldname^all (or exp.fieldname^all): | Same as above |

In the case of a CDS/ISIS "pseudo-repeatable" field, the following metadata elements are generated:

| | |
|--|---|
| ex.fieldname^sub (or exp.fieldname^sub): | The individual delimited terms in fieldname |
| ex.fieldname (or exp.fieldname): | The raw total content of fieldname |
| ex.fieldname^all (or exp.fieldname^all): | Same as above |

C- Search Indexes

Go to the **Search Indexes** view of **Gather** panel.

The **Assigned Indexes** by default will be (see Figure 14, but *attention*, if you have not enabled advanced searches as described under point A above, the screen will look different):

- text "text"** [index on the text of the record (in "as is" conversion) or of the linked document (in exploded conversion)]
- ex.Title "Title"** [index on the extracted title - this is only assured to be the real title if you have a CDS/ISIS field called Title and your collection has not be exploded]
- ex.Source "Source"** [index on the source file name (for an "as is" conversion it is YYYY.MST for all records, where YYYY is the name of the CDS/ISIS database, and thus useless)]

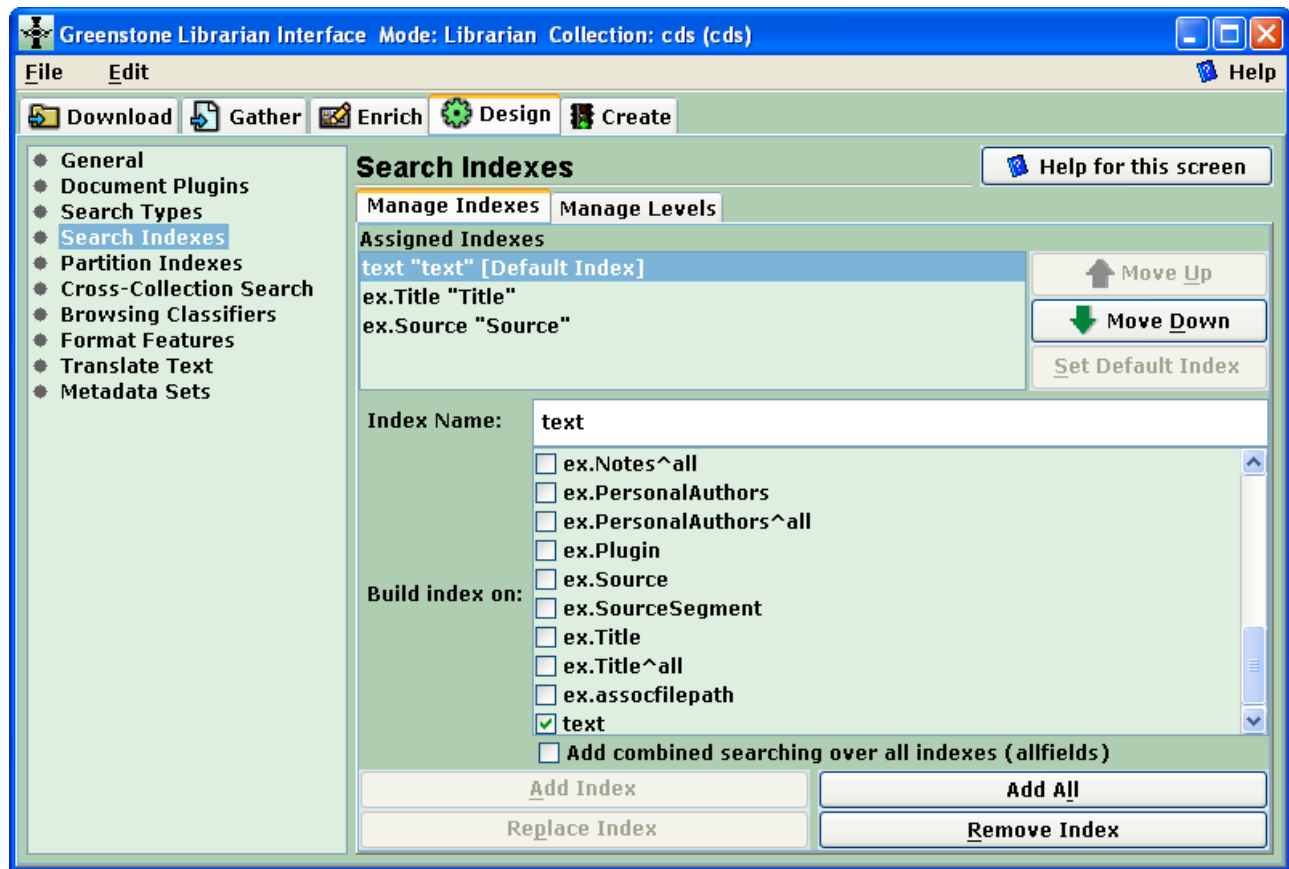


Figure 14: The view for setting search indexes

For any index, you can change the "Index Name" (presented in quotes in the corresponding line in the **Assigned Indexes** box) and/or "Build index on" (the metadata element to be indexed) by selecting the target line in the **Assigned Indexes** box, and then changing the data in the two corresponding spaces below it (*NOTE* that the **Add Index** and/or **Replace Index** buttons only become active when you have made a change in "Index Name" or "Build index on" parameters).

For an "as is" conversion or if your library has been exploded to contain textual documents, you will normally keep the *text* index line to index the full record or the full associated document.

For an "as is" conversion, *ex.Title* should be kept for if there is a CDS/ISIS field called title; otherwise you should change to the element which does represent the title (e.g. "ex.Name"). For an exploded conversion, *ex.Title* will normally be useless and should be changed to "exp.Name" where Name is the CDS/ISIS field name containing the title of the document.

If you wish to search on file names in the case of an exploded database, then you may keep the *ex.Source* line.

For an exploded conversion, the list of metadata elements presented for selection as indexes includes the basic Greenstone extracted elements (*ex.Title* and *ex.Source* and the elements of the "exploded" metadata set (exp.zzz where zzz is the name of a CDS/ISIS field). In addition, if the *metadata_set* parameter has not been set (see Part 2, Section A) or for Greenstone versions prior to 2.70, you may also see other elements extracted by ISISPlug (ex.zzz where zzz is the name of a CDS/ISIS field). **In that case do not specify the ex.zzz elements which are not operative; use the exp.zzz elements instead.**

D- Browsing Classifiers

The classifiers are configured by selecting the **Browsing Classifiers** view in the *Design* panel which is shown in Figure 15.

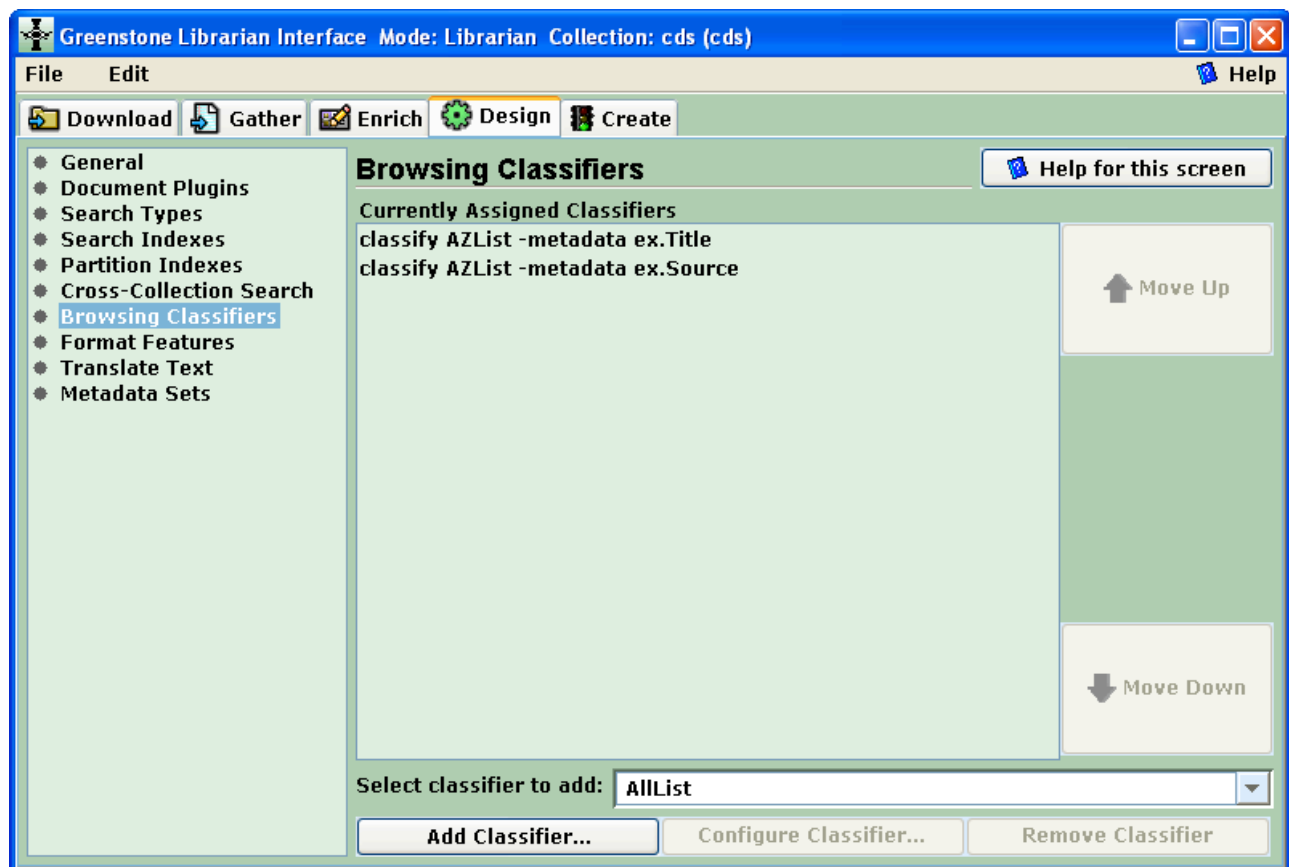


Figure 15: *Browsing Classifiers* view showing default classifiers

The classifiers proposed by default are:

| | |
|--|---|
| <i>classify AZList -metadata ex.Title</i> | [alphabetically sorted list ordered according to the extracted title - this is only assured to be the real title if you have a CDS/ISIS field called Title and your collection has not been exploded] |
| <i>classify AZList -metadata ex.Source</i> | [alphabetically sorted list ordered according to the source file name (for an "as is" conversion it is YYY.MST for all records, where YYY is the name of the CDS/ISIS database, and thus useless)] |

There are numerous classifier types available in Greenstone but for simplicity this guide will refer only the AZList (display of simple vertical list of records in alphabetical order) and AZCompactList (display of an alphabetically sorted vertical list of metadata values - clicking on one of the values yields a vertical list of the corresponding records in alphabetical order, similar to the list provided by AZList). The AZCompactList is used to browse metadata for which the same value can occur in several records, such as for authors and keywords.

For an exploded conversion, the list of metadata elements presented for selection as classifiers includes the basic Greenstone extracted elements (*ex.Title*, *ex.Source*, *ex.Encoding* and *ex.Language*) and the elements of the "exploded" metadata set (exp.zzz where zzz is the name of a CDS/ISIS field). In addition, if the *metadata_set* parameter has not been set (see Part 2, Section A) or for Greenstone versions prior to 2.70, you may also see other elements extracted by ISISPlug (ex.zzz where zzz is the name of a CDS/ISIS field). **In that case do not specify the ex.zzz elements which are not operative; use the exp.zzz elements instead.**

Classifiers are added, configured or removed using the *Browsing Classifiers* view (Figure 15). As a model, we will assume that the user wishes to enable browsing on title, authors and keywords, for which the following steps should be followed (in any order):

- Remove the source file browser by selecting the *ex.Source* line in the *Currently Assigned Classifiers* box and clicking on the *Remove Classifier* button.
- If the collection has been exploded or if the name of the title field of your CDS/ISIS database is other than "Title", then change the metadata element to be browsed in the title classifier line to the correct metadata name (e.g. if the title field in CDS/ISIS is "Name", then change the metadata element from *ex.Title* to *ex.Name* for an "as is" conversion or to *exp.Name* for an exploded conversion). This is done by selecting the line specifying the desired element in the *Currently Assigned Classifiers* box and clicking on the *Configure Classifier* button to display the "Configuring Arguments" window (Figure 16):

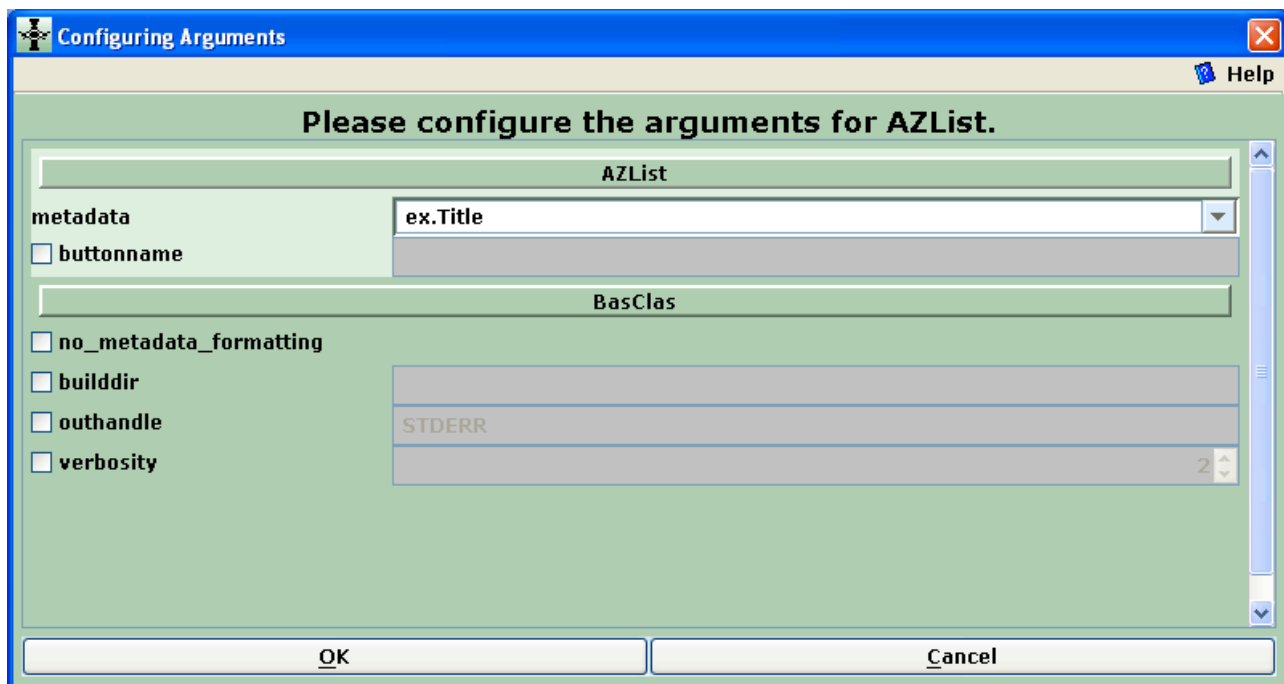


Figure 16: Configuration window for AZList classifier

Then you change the name of the metadata element to be browsed on in the top field and, if you want the browsing classifier to be given a name other than the metadata element name minus the prefix (in this case "Title"), check the **buttonname** box and type in the desired name. Then click **OK**.

- Add the additional authors and keywords fields using the AZCompactList classifier type. For each classifier, **Select AZCompactList** in the **Select classifier to add** field of the **Browsing Classifiers** view (Figure 15), then click on the **Add Classifier** button to get the "Configuring Arguments" window (Figure 17). Select the name of the metadata element to be browsed on in the top field (if the field is repeatable or pseudo-repeatable, choose the occurrence metadata element rather than the entire field, e.g. ex.PersonalAuthors or "ex.Keywords^sub" rather than ex.PersonalAuthors^all or "ex.Keywords^all"). Set the **mingroup** parameter to "1" and, if you want the browsing classifier to be given a name other than the metadata element name minus the prefix (in this case "Personal Authors" instead of "PersonalAuthors"), check the **buttonname** box and type in the desired name. Then click **OK**.

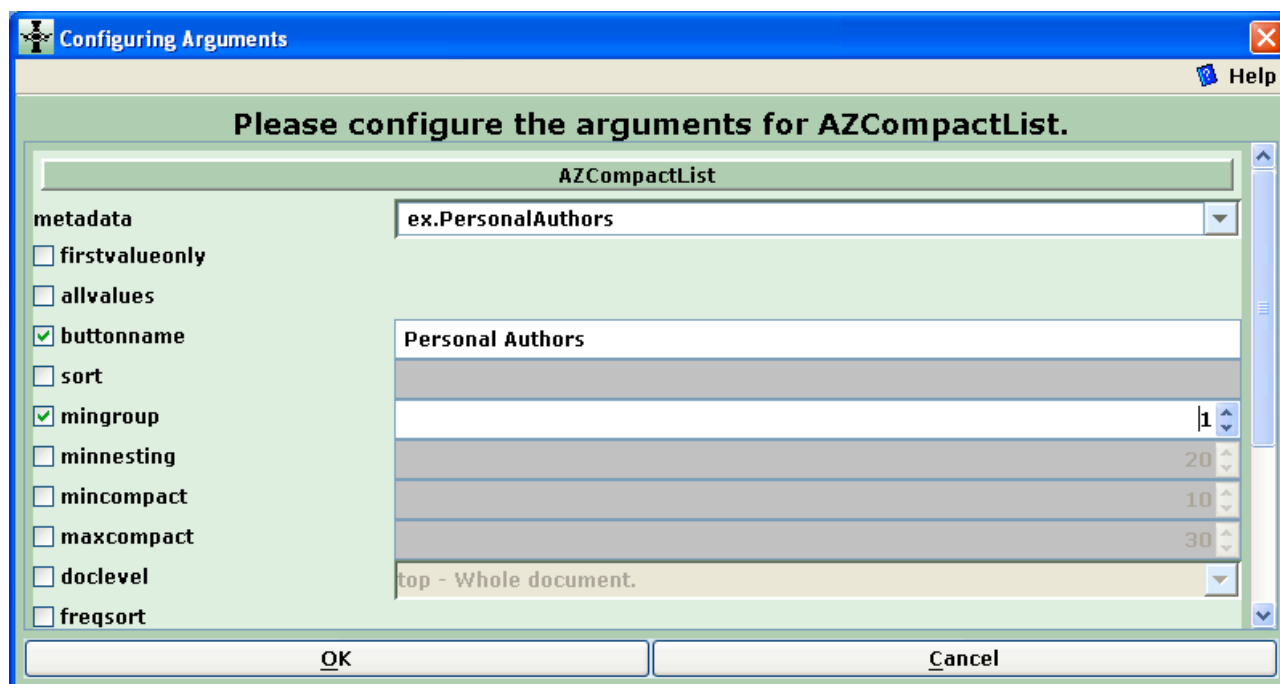


Figure 17: Configuration window for AZCompactList classifier

If the collection is rebuilt and previewed in the *Create* view, the new browser structure will appear in the homepage of the collection as seen in Figure 18.²¹ The order of the browser classifiers can be changed by selecting a line in the *Currently Assigned Classifiers* box, clicking on the *Move Up* or *Move Down* buttons as appropriate, and rebuilding the collection.

²¹ In Greenstone versions prior to 2.63, specifying the buttonname parameter may cause the changed classifier names to appear as underlined text rather than as buttons, because there is no button in Greenstone corresponding to the names of the classifiers (here PersonalAuthors and Keywords). Buttons are available for all of the metadata elements of the metadata sets provided with Greenstone (one can see the elements of any metadata set by temporarily adding it in the Metadata Sets view of the Design panel, and then removing it after review, or by using the Greenstone Editor for Metadata Sets (GEMS)). For example, you will see that "Keyword" rather than "Keywords" exists in the Development Library Subset (dls) metadata set, and knowing this, you can set the buttonname parameter in the classifier configuration window to "Keyword" and rebuild to get this button on the collection homepage. If you want to insert a button with an unsupported name, there is a page on the Greenstone website that can be used to generate a new button in the default Greenstone style (<http://www.cs.waikato.ac.nz/~mdewsnip/greenstone/make-images.html>).

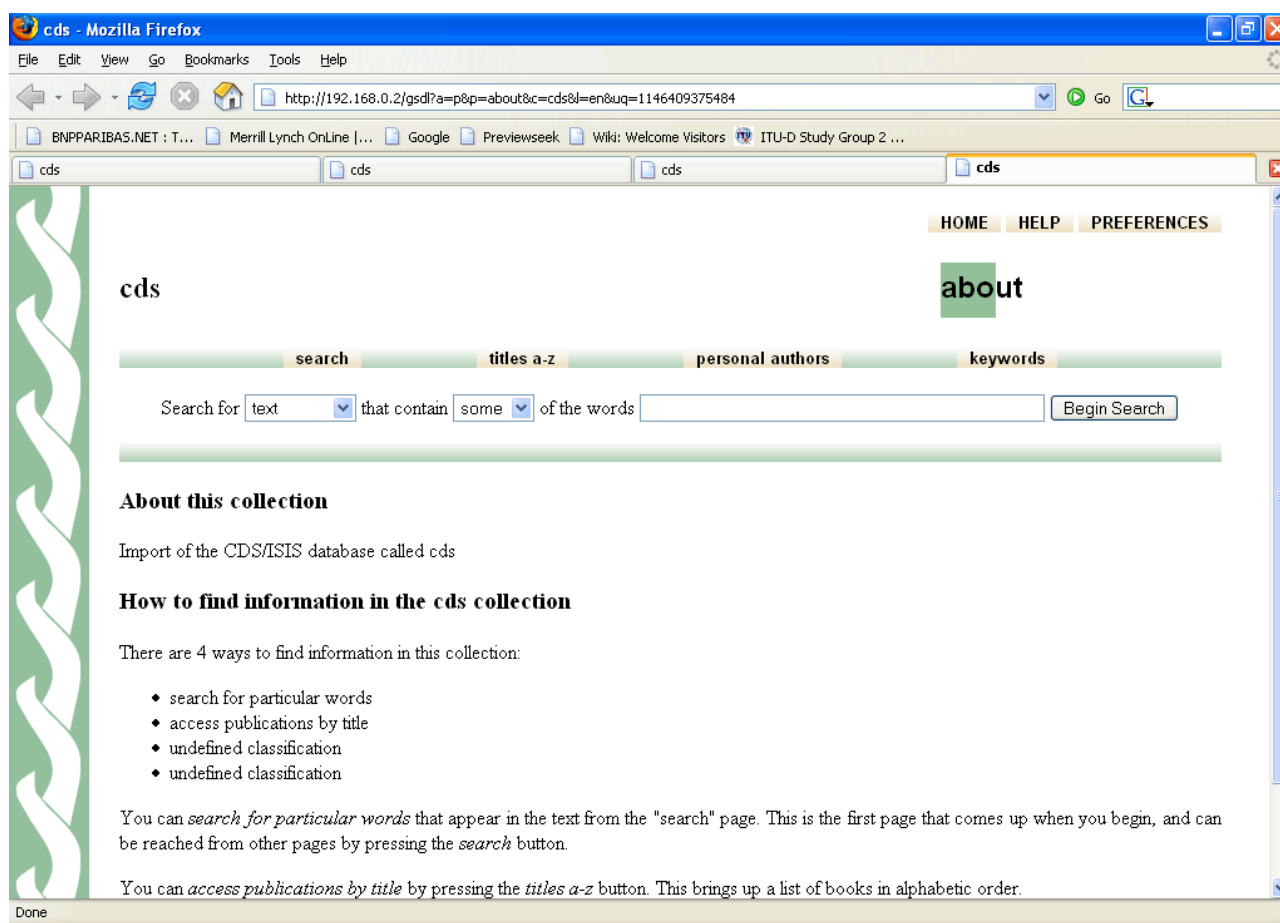


Figure 18: Modified browsing classifiers on the homepage of the "cds" collection

E- Format Features

Greenstone uses six principal display formats in Greenstone formatting language (modified html) to present metadata and documents to the end user. These display formats can be edited in the **Format Features** view of the **Design** view (note that they can be changed and previewed in GLI without rebuilding the collection).

Four of the display formats (**DateList**, **Hlist**, **VList**, and **DocumentButtons**) determine the display of the record references in browsing lists and search results, while the two others (**DocumentHeading** and **DocumentText**) determine the display of the text of the full record (in "as is" conversion) or of the associated full-text document (in exploded conversion).

In most cases of conversions discussed in this guide, the user will find default display formats set by Greenstone to be acceptable. The two formats which users are most likely to wish to change are **VList** and **DocumentText**, which will be treated as examples below.

A third relevant formatting point covered below is the display of repeatable fields, fields with subfields, and "pseudo-repeatable" fields.

For more detailed formatting needs, the user is referred to the resources in Section F.

- VList is a format which determines how the record reference (including the title and other metadata determined by the user) is displayed vertically in the search results and browsing lists. It can be edited by selecting the line starting with "format VList" in the main **Format Features** box (DO NOT select it in the "Affected Component" field which is for adding new formats rather than modifying existing ones). The box with VList selected for editing is shown in Figure 19 (when working in Library Systems Specialist or Expert mode, MAKE SURE to expand the window to full screen to see the contents of the currently assigned format features).

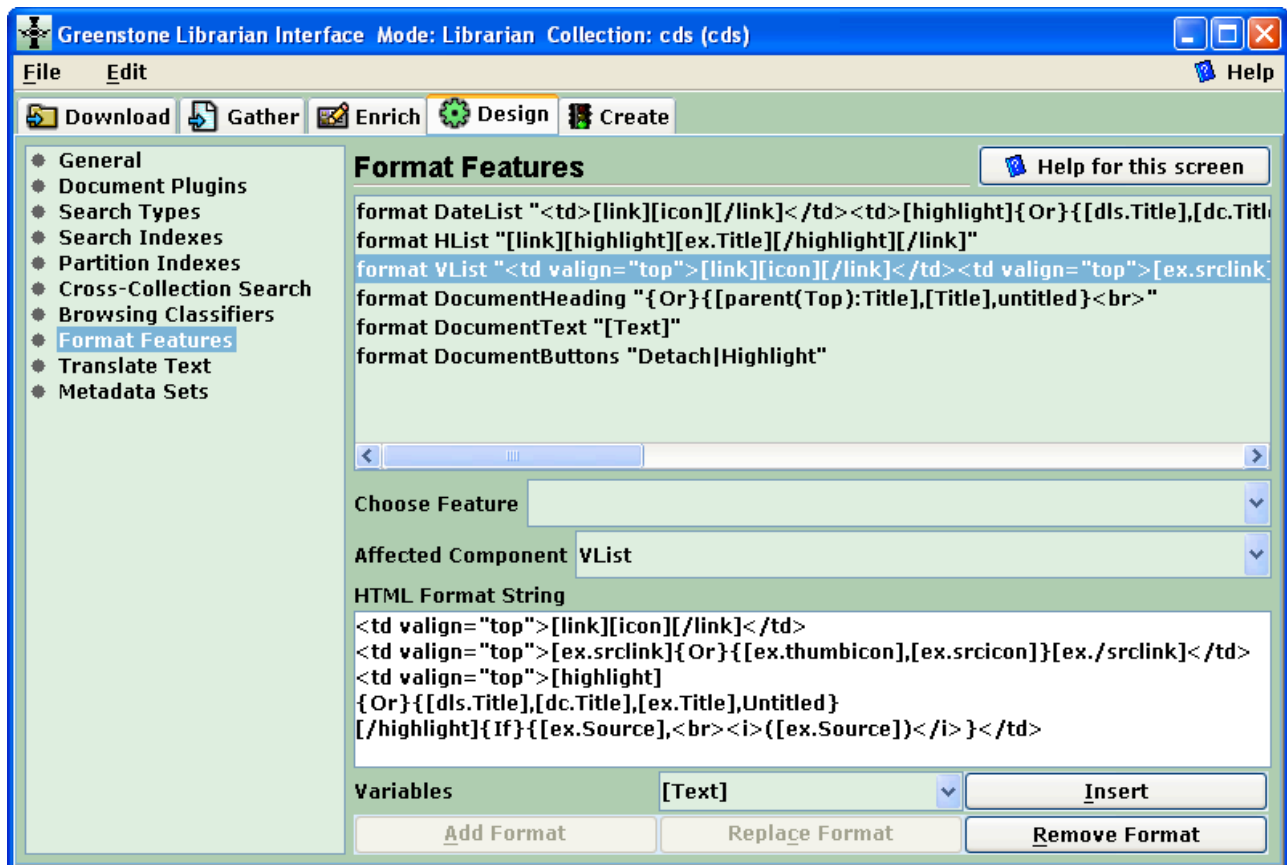


Figure 19: Format Features window ready for the editing of VList

VList can now be edited in the **HTML Format String** box by simply typing in the box and/or inserting specified metadata elements or standard format elements by choosing them in the drop-down list of **Variables** and clicking on the **Insert** button. When the editing is complete, click on the **Replace Format** button.

The default version of **VList** is:

```
<td valign=top>[link][icon][link]</td>
<td valign=top>[ex.srclink]{Or}{[ex.thumbicon],[ex.srcicon]}[ex./srclink]</td>
<td valign=top>[highlight]
{Or}{[dls.Title],[dc.Title],[ex.Title],Untitled}
[/highlight]{If}{[ex.Source],<br><i>([ex.Source])</i>}</td>
```

For example, if one wants the list of authors in parentheses instead of the source file name in parentheses, the last line may be changed to:

```
[/highlight]{If} {[ex.PersonalAuthors^all],<br><i>([ex.PersonalAuthors^all])</i></td>
```

The only major problem with the default VList format would be if the CDS/ISIS title field is called something other than "Title". For example if this field is called "Name", the fourth line of the Vlist format should be changed to the following:

```
{Or} {[ex.Name],Untitled} [This displays the value of ex.Name if this field exists,  
else the mention "Untitled".]
```

To get back the default setting of the format, remove the format, then reopen the collection clicking on the **File/Open...** main menu item

- The **DocumentText** format displays the full text associated with a selected record (the full CDS/ISIS record in the case of "as is" conversion, or the full electronic document if it has been imported through the explode method). If an electronic document has been imported, then the default value of this format will normally suffice.

However, if the display text is intended to be the CDS/ISIS record (not the associated document), the user may wish to modify the display. In an "as is" conversion this is less likely since the default is a specifically designed **text** metadata element generated by ISISPlug from the raw CDS/ISIS record (named **ex.ISISRawRecord** and searchable as **raw record**).²² However, in the case of an exploded conversion in which there are dummy documents (for which the **text** metadata will be null), the **DocumentText** format should be edited (in the same way as was done for VList above) to show the metadata elements in the desired way.

For example, for an exploded conversion, if the default format ([Text]) is changed as follows:

```
Title: [exp.Title]<br>Authors: [exp.PersonalAuthors^all]<br>Publisher:  
[exp.Imprint^all]<br>Keywords: [exp.Keywords^all]<br>[Text]
```

the text of one specific record would display as:

```
Title: Policy Guidelines for the Development and Promotion of Governmental Public Domain Information  
Authors: Uhler, P.F.  
Publisher: Paris, UNESCO, 2004  
Keywords: public domain, digital information, open access, copyright
```

followed by the full text of the associated document if it is in the collection.

- The ISISPlug plugin has two formatting parameters which affect the display of repeatable field and subfield metadata. These are the **entry_separator** and **subfield_separator**

²² However, in Greenstone versions prior to 2.70, **text** contains the raw record which the user is likely to want to modify.

parameters which can be edited by checking the corresponding boxes in the "Configuring Arguments" window for ISISPlug (Figure 20)²³ which is obtained by selecting ISISPlug in the *Document Plugins* view of the *Design* panel, and then clicking on the *Configure Plugin* button (see Figure 3). These parameters are useful in changing the content of metadata element for full field in the case of repeatable fields and fields with subfields.

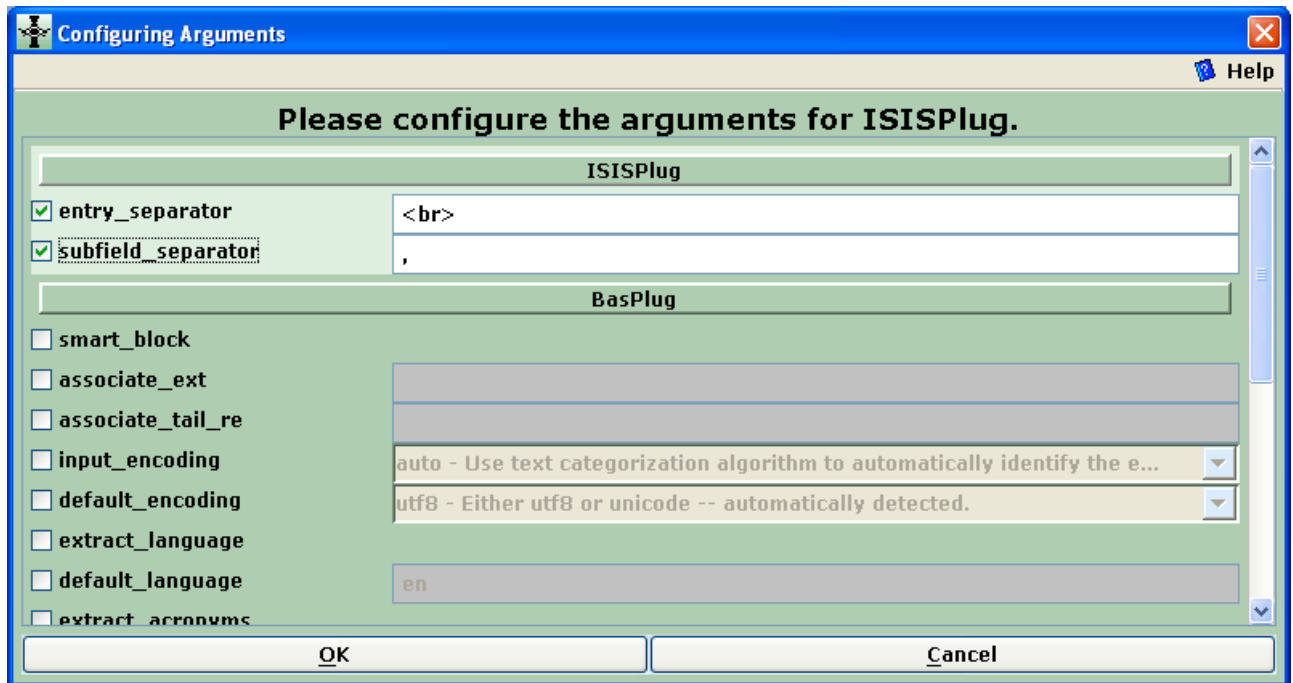


Figure 20: Formatting parameters selected in ISISPlug

The *entry_separator* parameter controls how the combined content of a repeatable field (designated in Greenstone as ex.fieldname^all) is derived from the individual occurrences; the default is
²⁴ which generates a combined field with the a new line between occurrences.

The *subfield_separator* controls how combined content of a field with subfields (the identical metadata elements designated as ex.fieldname and ex.fieldname^all) is derived from the individual subfields; the default value is a comma followed by a space (the space is entered in the parameter but not explicitly displayed in the input line), which generates a combined field consisting of the subfields separated by comma-space.

For "pseudo-repeatable" fields there is no metadata element for delimited occurrences (ex.fieldname and ex.fieldname^all contain the raw field with delimiters). One can generate a string of the occurrences with a chosen separator (e.g. " - ") by using the following format:

[sibling(All' - '):ex.fieldname^sub]

²³ When a parameter's box is not checked, the values shown in gray characters on gray background are active but cannot be edited.

²⁴ In version 2.63 and before, this has to be written as "
" using the sequences to represent the "less than" (<) and "greater than" (>) brackets in HTML.

F- Resources for detailed guidance

The following resources may be consulted for more complete/advanced guidance on configuring Greenstone search types, search indexes, browsing classifiers and format features.

- A summary of all formatting features and commands:
http://greenstone.sourceforge.net/wiki/index.php/How_to_format_the_output_of_your_collection
- Some example collections which have documentation about their configuration:
http://greenstone.sourceforge.net/wiki/index.php/Example_collections
- Section 2.3 of the Greenstone Developer's Guide:
<http://prdownloads.sourceforge.net/greenstone/Develop-en.pdf>
- Fiji workshop, greenstone tutorials, downloadable from:
http://greenstone.sourceforge.net/wiki/index.php/Tutorial_exercises