
Rhythmyx

Enterprise Content Connector

Version 6.5.2

Copyrights and Trademarks

Copyright and Licensing Statement

All intellectual property rights in the SOFTWARE and associated user documentation, implementation documentation, and reference documentation are owned by Percussion Software or its suppliers and are protected by United States and Canadian copyright laws, other applicable copyright laws, and international treaty provisions. Percussion Software retains all rights, title, and interest not expressly stated. You may either (a) make one (1) copy of the SOFTWARE solely for backup or archival purposes or (b) transfer the SOFTWARE to a single hard disk provided you keep the original solely for backup or archival purposes. You must reproduce and include the copyright notice on any copy made. You may not copy the user documentation accompanying the SOFTWARE.

The information in Rhythmyx documentation is subject to change without notice and does not represent a commitment on the part of Percussion Software, Inc. This document describes proprietary trade secrets of Percussion Software, Inc. Licensees of this document must acknowledge the proprietary claims of Percussion Software, Inc., in advance of receiving this document or any software to which it refers, and must agree to hold the trade secrets in confidence for the sole use of Percussion Software, Inc.

The software contains proprietary information of Percussion Software; it is provided under a license agreement containing restrictions on use and disclosure and is also protected by copyright law. Reverse engineering of the software is prohibited.

Due to continued product development this information may change without notice. The information and intellectual property contained herein is confidential between Percussion Software and the client and remains the exclusive property of Percussion Software. If you find any problems in the documentation, please report them to us in writing. Percussion Software does not warrant that this document is error-free.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of Percussion Software.

Copyright © 1999-2007 Percussion Software.
All rights reserved

Licenses and Source Code

Rhythmyx uses Mozilla's JavaScript C API. See <http://www.mozilla.org/source.html> (<http://www.mozilla.org/source.html>) for the source code. In addition, see the *Mozilla Public License* (<http://www.mozilla.org/source.html>).

Netscape Public License

Apache Software License

IBM Public License

WebSphinx Licence

Other Copyrights

The Rhythmyx installation application was developed using InstallShield, which is a licensed and copyrighted by InstallShield Software Corporation.

The jTDS driver is licensed and copyrighted by CDS Networks, Inc.

The Sprinta JDBC driver is licensed and copyrighted by I-NET Software Corporation.

The Sentry Spellingchecker Engine Software Development Kit is licensed and copyrighted by Wintertree Software.

The Java™ 2 Runtime Environment is licensed and copyrighted by Sun Microsystems, Inc.

The Oracle JDBC driver is licensed and copyrighted by Oracle Corporation.

The Sybase JDBC driver is licensed and copyrighted by Sybase, Inc.

The AS/400 driver is licensed and copyrighted by International Business Machines Corporation.

The Ektron DHTML editor is licensed and copyrighted by Ektron, Inc.

WebSphinx webcrawling toolkit is licensed by Carnegie Mellon University.

AuthorIT™ is a trademark of Optical Systems Corporation Ltd.

Microsoft Word, Microsoft Office, Windows®, Window 95™, Window 98™, Windows NT® and MS-DOS™ are trademarks of the Microsoft Corporation.

This document was created using *AuthorIT™, Total Document Creation* (see AuthorIT Home - <http://www.author-it.com>).

Schema documentation was created using XMLSpy™.

Percussion Software

600 Unicorn Park Drive

Woburn, MA 01801 U.S.A

781.438.9900

Internet E-Mail: technical_support@percussion.com

Website: <http://www.percussion.com>

Contents

Introduction to the Rhythmyx Enterprise Content Connector 3

How The Rhythmyx Enterprise Content Connector Works.....	4
Starting the Rhythmyx Enterprise Content Connector.....	5

The Rhythmyx Enterprise Content Connector Interface 7

The Menu Bar	8
The Toolbar	10
Content Tree	11
Status View	13
Metadata View.....	14
Log View	15
Setting User Preferences.....	16

Maintaining Rhythmyx Enterprise Content Connector Descriptors 17

Descriptor Setup Dialog.....	19
Creating a Rhythmyx Enterprise Content Connector Descriptor	21
Editing a Rhythmyx Enterprise Content Connector Descriptor	22
Creating a New Descriptor by Modifying an Existing Descriptor	23
Adding a Descriptor from another Host.....	24
Configuring Content Selectors.....	25
Choosing a Content Selector	25
Configuring a List Content Selector	27
Configuring a File Content Selector	27
Maintaining Content Extractors.....	31
Managing the Extractors Assigned to a Descriptor	32
Extractor Editors.....	33
Configuring Transformers	49
Configuring Field Transformers	50
Configuring Item Transformers.....	52
Configuring Content Loaders	54
Choosing a Content Loader	54
Configuring the Rhythmyx Loader.....	56
Configuring the Preview Loader	56

Configuring Logging	57
Configuring Descriptor Error Handling	59
Saving Descriptors	61

Running the Rhythmyx Content Connector 63

Running The Enterprise Content Connector Interactively	64
Modifying the Content Tree	67
Uploading Selected Items and Folder Contents	68
Scanning for New and Changed Content	70
Saving a Status File After Scanning	70
Scanning and Uploading Again to Correct Errors	71
Running The Rhythmyx Enterprise Content Connector Automatically	72

Examples of Rhythmyx Enterprise Content Connector Operations 73

Implementing Site Migration with Enterprise Content Connector: One Time Migration Example	74
Migration Example: Introduction	74
Migration Example: Setting up Rhythmyx Roles and Workflows	75
Migration Example: Creating a Descriptor	76
Migration Example: Scanning and Uploading	89
Migration Example: Reviewing an Upload	92
Incremental Upload Example	94
Incremental Upload Example: Introduction	94
Incremental Upload Example: Procedure	94
Incremental Upload Example: Creating the Syndicated Workflow	95
Incremental Upload Example: Creating the Descriptor and Performing the Initial Upload	96
Incremental Upload Example: Batch Processing	97

Configuring The Rhythmyx Enterprise Content Connector 100

Changing Configuration File Default Values	101
Adding New Plugins	111
Writing Content Selectors	118
Writing Extractors	119
Writing Transformers	122
Writing Loaders	123

The Enterprise Content Connector and Rhythmyx 125

Creating a Rhythmyx User to Load Content	126
Reconfiguring Workflows for Uploaded Content	127

Index 129

CHAPTER 1

Introduction to the Rhythmyx Enterprise Content Connector

The Rhythmyx Enterprise Content Connector is a stand alone tool that uploads content in HTML files, image files, and other file types from an external site into a Rhythmyx server or a filesystem. You may use the Enterprise Content Connector to upload the contents of an entire Web site when you initially switch to Rhythmyx or to regularly upload syndicated content or content stored in a data asset management system. During scheduled uploads, the Enterprise Content Connector compares current information to information that it previously uploaded and incrementally uploads new or modified content.

You can run the Enterprise Content Connector through the Content Connector Interface, which lets you create a description file for an upload, monitor your progress as you scan and upload files, modify scanned files before uploading, and view errors. You can also run the Enterprise Content Connector on a command line. If you are performing scheduled or incremental uploads, you can configure a batch file to run the Enterprise Content Connector at regular intervals with a scheduler program.

How The Rhythmyx Enterprise Content Connector Works

The Rhythmyx Enterprise Content Connector uploads content from an external site by scanning it and then uploading it to Rhythmyx or a file system. The uploading process includes retrieving data, modifying its contents to conform to Rhythmyx standards, and uploading it.

The Enterprise Content Connector knows where to find the external files, how to recognize which files correspond to Content Types in Rhythmyx, and how to reconfigure information in the files by looking at the information that you supply in a *Descriptor* (see "[Maintaining Rhythmyx Enterprise Content Connector Descriptors](#)" on page 17), the configuration file for a specific upload. The Enterprise Content Connector provides the *Descriptor Setup dialog* (on page 19) for configuring the settings for Java plugins that the Enterprise Content Connector uses to scan and upload data and for logging and error handling. The Enterprise Content Connector configuration file, `contentconnector.xml`, contains settings, including all of the options available for configuring a Descriptor in the Descriptor Setup dialog, but it is not a Descriptor and you cannot use it to scan and upload data.

After the Enterprise Content Connector scans for content on a source site, it assigns a Rhythmyx Content Type to each file that matches its filters. If you are using the Content Connector Interface, you may review the results and reassign Content Types or mark files to be excluded from uploading. When you perform the upload, the Enterprise Content Connector extracts the scanned files (except those marked Exclude and Error) from the source, modifies them for Rhythmyx, and uploads the files to Rhythmyx or a preview folder. It also records or displays any error messages, and sends success or error notifications to specified users.

Each file extracted represents a Rhythmyx static file (for example, a stylesheet) or a content item. If the file represents a content item, the content of the file becomes the body of the content item.

If you are uploading incrementally (for syndication), after the Enterprise Content Connector uploads content to Rhythmyx, if a file has not been previously uploaded, Rhythmyx creates a new content item; if a file has been previously uploaded but its content has changed, Rhythmyx replaces the existing Content Item (its Revision ID does not change). If you are uploading to perform a one-time migration, you can upload all content as new. Then Rhythmyx further processes the items by entering system values into metadata fields as it would for any new or revised items, applying Workflow Actions to new content items, and checking in all new and changed content items.

When the Enterprise Content Connector uploads content items to a Preview location that you specify in the file structure of the source site, it gives all of the content items the Content ID of 0 and the Revision ID of 1. You must view these files in file editors rather than Rhythmyx.

By default, the Enterprise Content Connector uploads external items into the same location in the folder under `RhythmyxServer`.

Starting the Rhythmyx Enterprise Content Connector

To start the Rhythmyx Enterprise Content Connector, in the <Rhythmyx root> run the executable file `RhythmyxContentConnector.exe`.

CHAPTER 2

The Rhythmyx Enterprise Content Connector Interface

The Rhythmyx Enterprise Content Connector interface consists of a Title Bar, a *Menu Bar*, a *Toolbar* (see "The *Toolbar*" on page 10), and windows that display a *Content Tree* (on page 11), a *Status View* (on page 13), a *Metadata View* (on page 14), and a *Log View* (on page 15).

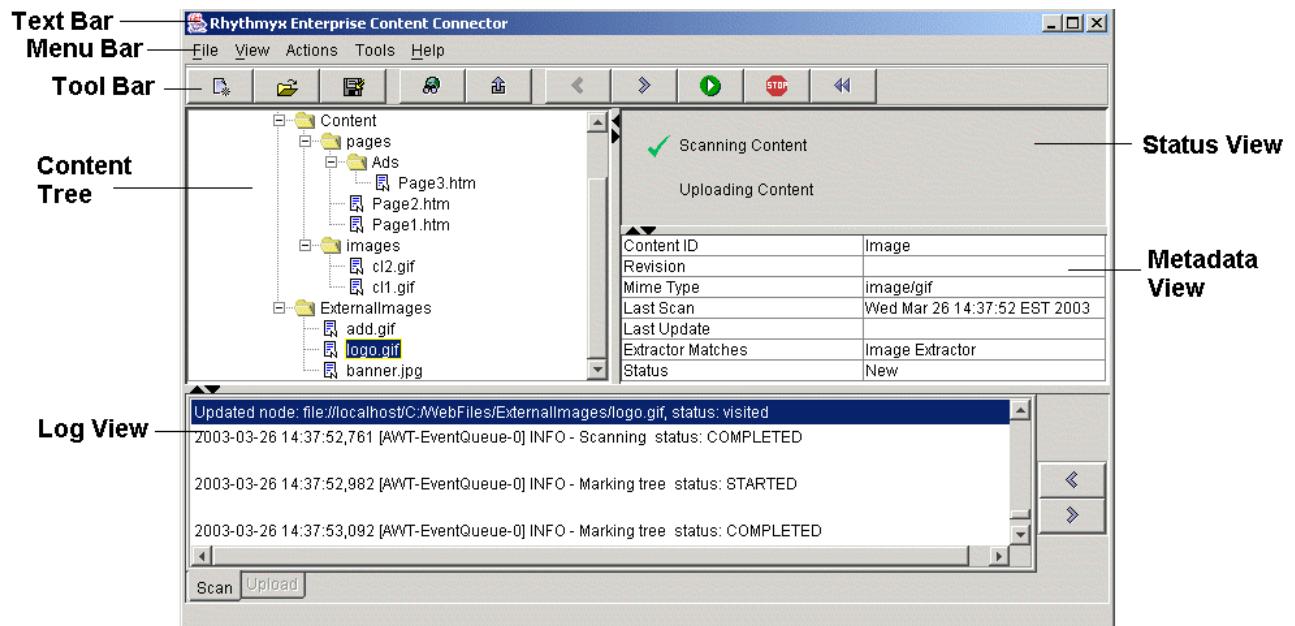


Figure 1: Enterprise Content Connector interface

The Menu Bar

The Menu Bar consists of five drop down menus:

- File Menu

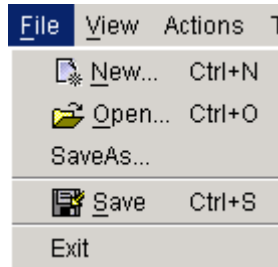


Figure 2: File menu

The File Menu includes the following options:

- *New* - Opens the Descriptor Setup dialog so that you can **create a new Descriptor** (see "[Creating a Rhythmyx Enterprise Content Connector Descriptor](#)" on page 21).
 - *Open* - Opens a file chooser dialog so that you can **load an existing Descriptor** (see "[Editing a Rhythmyx Enterprise Content Connector Descriptor](#)" on page 22).
 - *SaveAs* - Save the current Descriptor under a new name.
 - *Save* - Opens a Save dialog so that you can save the current Status file. The Descriptor refers to the Status file to determine which source content is new or changed. Status files are always saved to [Descriptor Name].cls in the [Descriptor Name] folder. When you open a previously used Descriptor, the Status file opens. When you choose *File > Save* after running the Descriptor, the newly created Status file overwrites the previous one.
 - *Exit* - Close the Enterprise Content Connector user interface.
- View

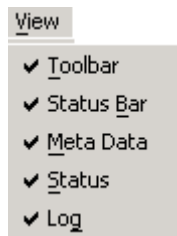


Figure 3: View menu

The View Menu includes the following interface options:

- *Toolbar* - toggles the Toolbar between view and hide modes;
- *Status Bar* - toggles the Status Bar between view and hide modes;
- *Log* - toggles the Log View between view and hide modes;
- *Meta Data* - toggles the Metadata View between view and hide modes;
- *Status* - toggles the Status View between view and hide modes.

- Actions

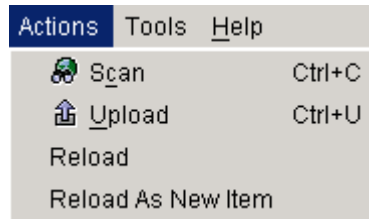


Figure 4: Actions menu

The Actions Menu includes the following options:

- *Scan* - Begins scanning based on the open Descriptor. Disabled if no Descriptor is loaded.
 - *Upload* - Begins uploading based on the current Content Tree. Disabled if no Descriptor is loaded, scanning is not complete, or the scanned content does not require uploading.
 - *Reload* - Uploads the selected file(s). Files must be in New, Changed, or Unchanged Status.
 - *Reload as New Item* - Uploads the selected file(s) as a New items whether or not they were uploaded previously. (By default, when previously uploaded items are reloaded, they update the original item.) Files must be in New, Changed, or Unchanged Status.

- Tools

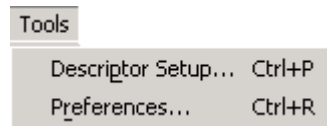


Figure 5: Tools menu

The Tools Menu includes the following options:

- *Descriptor Setup* - Opens the **Descriptor Setup dialog** (on page 19) for the currently opened Descriptor. Disabled if no Descriptor is opened.
 - *Preferences* - Opens the Preferences dialog so that you can **specify the Descriptor** (see "[Setting User Preferences](#)" on page 16) (new, last, or none) that the Enterprise Content Connector opens when it starts.

- Help













Figure 6: Help menu

The Help Menu includes the following options:

- *About* - opens the Rhythmyx About dialog, which displays the Rhythmyx release, build, and license number;
 - *Content Connector Help* - opens Enterprise Content Connector online help.

The Toolbar

The Toolbar provides shortcuts for the most common Rhythmyx Content Connector tasks. By default the Enterprise Content Connector displays the Toolbar. Choose *View > Toolbar* to toggle it on and off.

Button	Description
	Same as <i>File > New</i> menu command.
	Same as <i>File > Open</i> menu command.
	Same as <i>File > Save</i> menu command.
	Same as <i>Actions > Scan</i> menu command.
	Same as <i>Actions > Upload</i> menu command.
	Go back to the previous task. Disabled if no previous task exists. Currently, the only possible previous task is Scanning.
	Go to the next task. Disabled if no task follows. Currently, the possible next tasks are Scanning and Uploading.
	Perform a scan and an upload. Does not stop after the scan to allow you to modify the Content Tree.
	Stop the current scanning or uploading task.
	Clear the Content Connector Interface of the previous scanning and uploading tasks. After clearing, the Content Tree appears empty, and the currently opened Descriptor remains open.

Content Tree

Before you scan content, the Content Tree in the Content Connector Interface is empty. After you scan content, the Content Tree displays a *Servers* folder with sub-folders for scanned sites. Expand the server folders to view a duplicate of the directory structure on the source site containing nodes for scanned files. Each scanned file has a Content Type and other metadata associated with it.

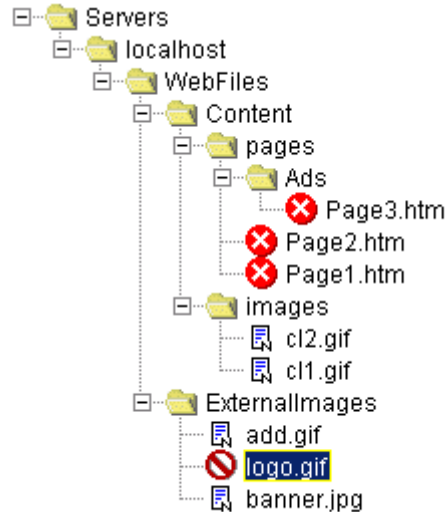



Figure 7: Servers Folder in Content Tree

A file's icon in the Content Tree indicates its status:


Content Tree Icon	Status and Description
	New. The Rhythmyx Enterprise Content Connector has scanned this file for the first time. When the Enterprise Content Connector initially scans a site, all files that <i>match extractors</i> (see " Maintaining Content Extractors " on page 31) have this icon. Items in New files are uploaded and added to the target site.
	Changed. The file contents have changed since the Enterprise Content Connector last uploaded it. Items in Changed files are uploaded to the target site and replace the existing item.
	Unchanged. The file has not changed since the Enterprise Content Connector last uploaded it. The Enterprise Content Connector will not upload it.
	Exclude. The Rhythmyx Enterprise Content Connector will not upload this scanned file. It gives this status to files that do not match extractors.
	Error. The Rhythmyx Enterprise Content Connector will not upload this scanned file. It gives this status if a file is required to have well-formed XML and the XML is not well-formed.

Right-click on files or folders in the Content Tree to access a menu of commands for editing the file or folder. You may select multiple files and right-click to select a file menu option to apply to all of the selected files.

File menu options may include:

- *Exclude* - Changes the file's icon to  indicating that the file will not be uploaded.
- *Assign > [content type]* - Changes the Content Type assignment of the file to the Content Type selected. The Content Type options are the Content Types of all extractors in the Descriptor. If a file has no Content Type and you give it one, its status changes to New.
- *Preview* - Displays a preview of the file's contents in the default system browser.
- *Upload* - Uploads the file. Does not upload any dependents.
- *Reload* - Uploads a file in New, Changed, or Unchanged Status.
- *Reload as New Item* - Uploads a file in New, Changed, or Unchanged Status as a New item whether or not it was uploaded previously. (By default, when previously uploaded items are reloaded, they update the original item.)

Folder menu options may include:

- *Exclude All* - Changes the icons of all files in the folder to , indicating that the files will not be uploaded.
- *Assign > [content type]* - Changes the Content Type assignment of all files in the folder to Content Type selected. The content type options are the Content Types of all extractors in the Descriptor. If a file has no Content Type and you give it one, its status changes to New.
- *Upload All* - Uploads all files in the folder. Does not upload their dependents.
- *Reload* - Uploads all files in the folder that have New, Changed, or Unchanged Status.
- *Reload as New Item* - Uploads all files in the folder that have New, Changed, or Unchanged Status as New items whether or not they were uploaded previously. (By default, when previously uploaded items are reloaded, they update the original item.)

Status View

Status View lists upload tasks:

- Scanning Content
- Uploading Content

If a task is complete, a check appears in front of it. If a task is in progress, an arrow appears in front of it and the task appears bold. The Status View window displays the task's progress.

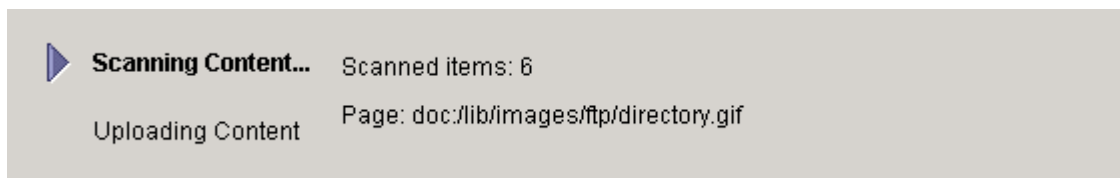


Figure 8: Status View

Metadata View

The Metadata View displays the metadata for a file selected in the Content Tree. If you have not yet uploaded the item(s) in the file, some of the values appear empty.

Content ID	Image
Revision	
Mime Type	image/gif
Last Scan	Wed Mar 26 14:37:52 EST 2003
Last Update	
Extractor Matches	Image Extractor
Status	New

Figure 9: Metadata View

Metadata View Fields

Content ID - Content Type assigned to file.

Revision - Revision ID.

Mime Type - MimeType of file.

Last Scan - Date and time last scanned.

Last Update - Date and time last updated.

Extractor Matches - The Extractor that matches with the file or the Extractor of the assigned Content Type. During scanning, if more than one Extractor matches with the file, the first matching Extractor specified in the Descriptor is assigned.

Status (see "[Content Tree](#)" on page 11) - New, Changed, Unchanged, or Excluded.

Log View

The scanning and uploading tasks have separate summary logs on the tabs in Log View. Each log contains entries that may refer to one or more files in the Content Tree. When you select an entry that refers to a file, the Enterprise Content Connector selects that file in the Content Tree.

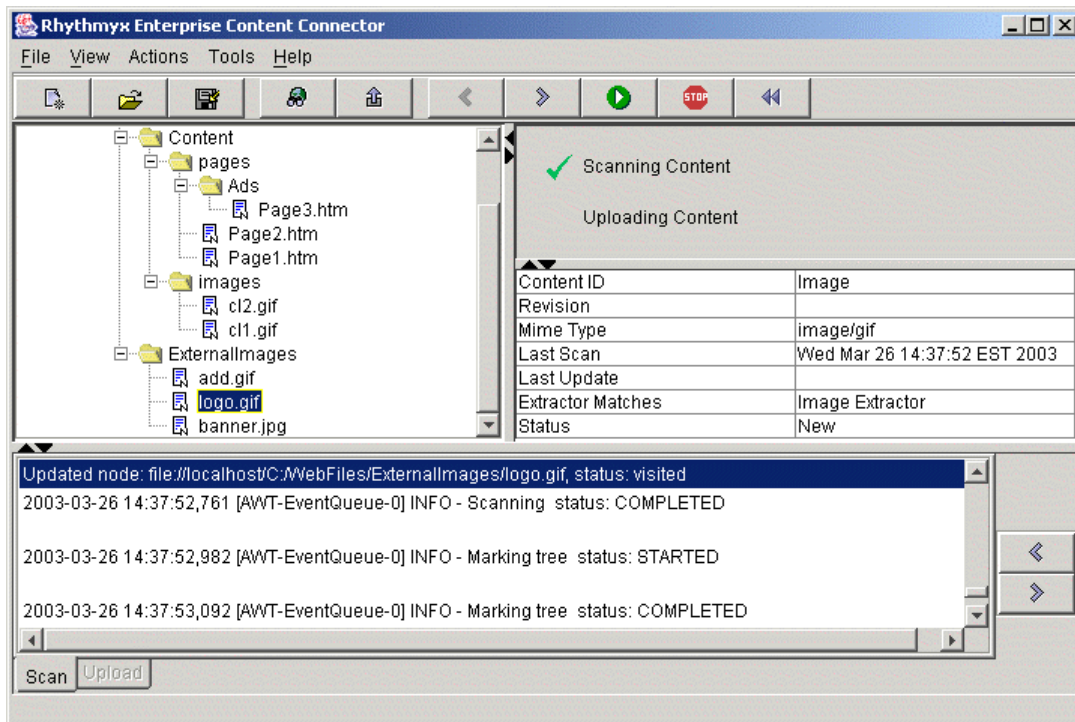




Figure 10: Log View

If a tab does not have log entries, it is disabled. If logging is disabled, the first tab shows the entry Log disabled and the other tab is disabled.

Click  and  beside Log View to move up and down the lists of log entries.

Setting User Preferences

User Preferences let you specify the Descriptor that opens (new, last, or none) when you open the Content Connector Interface.

To set User Preferences:

- 1 Choose *Tools > Preferences*.

The User Preferences dialog opens.

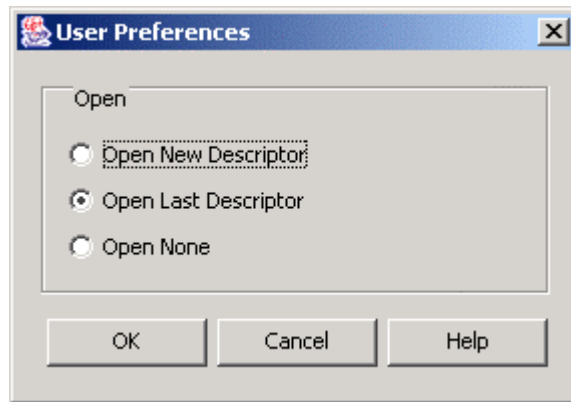


Figure 11: User Preferences dialog

- 2 Specify the Descriptor that the Content Connector Interface opens when you start it:
 - Open New Descriptor - When the Content Connector Interface starts, it opens the Descriptor Setup dialog for creating a new Descriptor.
 - Open Last Descriptor - Default. When the Content Connector Interface starts, it opens the last opened Descriptor. If no Descriptor was previously opened, it opens the Descriptor Setup dialog for creating a new Descriptor.
 - Open None - When the Content Connector Interface starts, it does not open a Descriptor.

CHAPTER 3

Maintaining Rhythmyx Enterprise Content Connector Descriptors

A Descriptor is a Rhythmyx Enterprise Content Connector configuration file. It includes elements that contain settings and rules for scanning and uploading. When you configure some of the elements, you also choose the plugins for the Enterprise Content Connector to use. Click on the element name in the table for more information about what it does.

Descriptor element	Plugins to choose	Settings configured
Content Selector (see "Configuring Content Selectors" on page 25)	A Content Selector plugin.	<ul style="list-style-type: none"> ▪ how to find source files; ▪ which source files to access or the source root to scan.
Static Item Extractors (see "Maintaining Content Extractors" on page 31)	Any number of Static Item Extractor plugins.	<ul style="list-style-type: none"> ▪ which Rhythmyx static items to extract; ▪ how to identify source files as matching each item type.
Item Extractors (see "Maintaining Content Extractors" on page 31)	Any number of Item Extractor plugins.	<ul style="list-style-type: none"> ▪ which Rhythmyx Content Types to extract; ▪ how to identify source files as matching each Content Type; ▪ Workflow Actions to apply to New content items.
Field and Item Transformers (see "Configuring Transformers" on page 49)	Any number of Field and Item Transformer plugins. Rhythmyx does not provide any Transformers.	<ul style="list-style-type: none"> ▪ how to modify data extracted by specific Item Extractors so that it conforms to Rhythmyx format.
Global Field and Global Item Transformers (see "Configuring Transformers" on page 49)	Any number of Field and Item Transformer plugins. Rhythmyx does not provide any Transformers.	<ul style="list-style-type: none"> ▪ how to modify data extracted by any Item Extractor so that it conforms to Rhythmyx format.
Content Loader (see "Configuring Content Loaders" on page 54)	A Content Loader plugin.	<ul style="list-style-type: none"> ▪ where to load items; ▪ user credentials for connecting to a server.

Descriptor element	Plugins to choose	Settings configured
Logging	none	<ul style="list-style-type: none">▪ how to log scan and upload information;▪ which information to log.
Error Handling (see "Configuring Descriptor Error Handling" on page 59)	none	<ul style="list-style-type: none">▪ what actions to take when an error occurs.

You can create new Descriptors for each upload, or modify existing ones, and once you have created a Descriptor, you can reuse it.

Through the Content Connector Interface, you can access the **Descriptor Setup dialog** (on page 19) for **creating new Descriptors** (see ["Creating a Rhythmyx Enterprise Content Connector Descriptor"](#) on page 21) and **editing existing Descriptors** (see ["Editing a Rhythmyx Enterprise Content Connector Descriptor"](#) on page 22).

Descriptor Setup Dialog

The Descriptor Setup dialog lets you create or edit a Descriptor. The dialog consists of a Descriptor Tree and an Editor pane. When you open the dialog, the Descriptor Tree displays a folder for each Descriptor element and the Editor pane displays a Descriptor **Name** and **Path** and credentials necessary for connecting to a Rhythmyx server. When you click a folder in the Descriptor Tree, the Editor pane displays a list of options or a configuration editor. Choose options and click [**Apply**] to add nodes for them below the folders in the Descriptor tree. Click the nodes to display their configuration editors in the Editor pane.

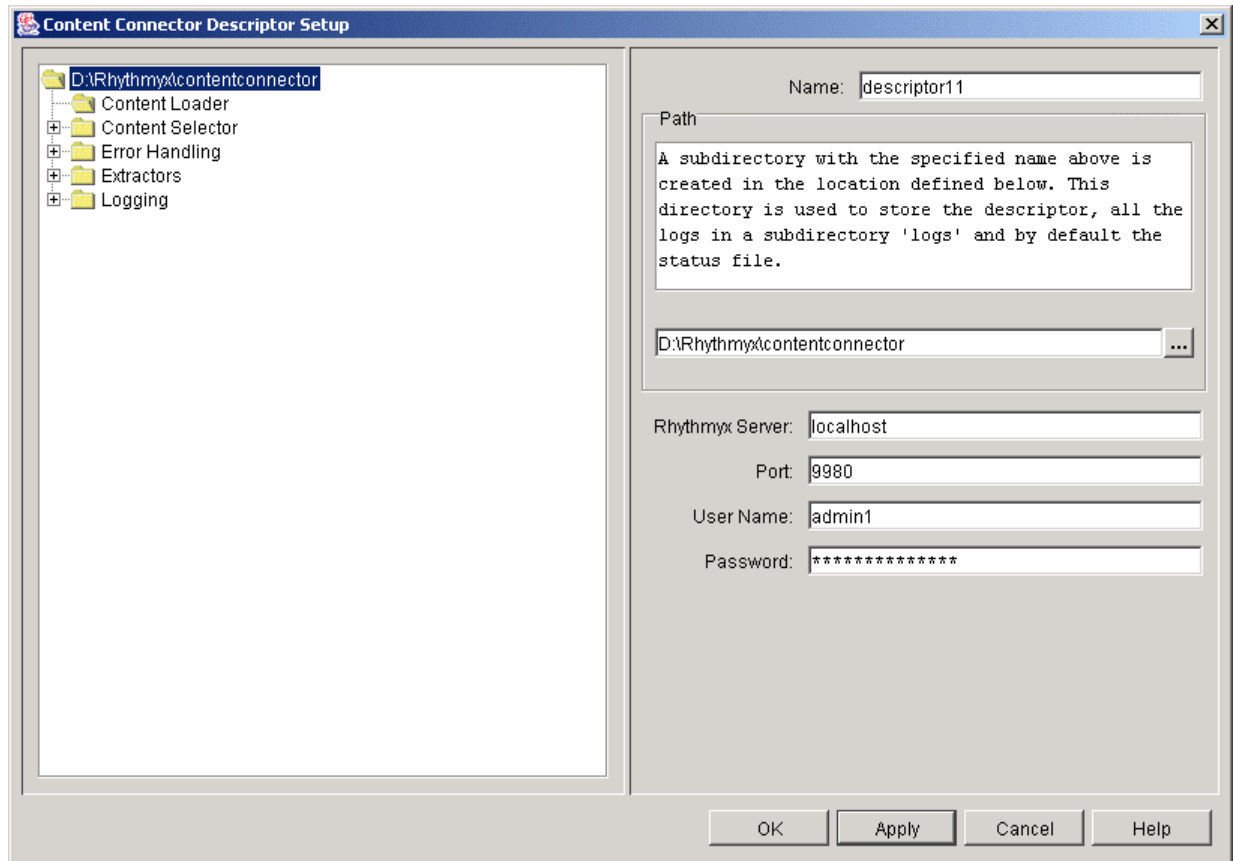


Figure 12: Descriptor Setup dialog

The Descriptor Setup dialog has four buttons:


Button	Action
[OK]	saves the current Descriptor settings and closes the Descriptor Setup dialog.
[Apply]	saves the current Descriptor settings and leaves the Descriptor Setup dialog open.
[Cancel]	closes the Descriptor Setup dialog without saving the current Descriptor settings.
[Help]	opens the online Help.

These buttons are accessible to you whenever the Descriptor Setup dialog is open.

Creating a Rhythmyx Enterprise Content Connector Descriptor

To create a new Descriptor:

- 1 Open the Content Connector Interface.

- 2 Choose *File > New* or click .

The Enterprise Content Connector opens the *Descriptor Setup dialog* (on page 19). The Descriptor tree displays folders for the Descriptor elements. The Editor pane displays the default Descriptor name in Name, and Enterprise Content Connector directory in Root. It also displays fields for required user credentials.


NOTE: The initial default Descriptor name is descriptor1. If you save that name, the next default Descriptor name is descriptor2, and so on.

- 3 Optionally, in Name, change the default name for the Descriptor. [Name] .xml is the file name of the Descriptor; Name is also the folder name that holds the Descriptor file, the log files, and the Status file in the Path.
- 4 Optionally, in Path, enter a path. Do not include <Name> at the end because Enterprise Content Connector appends it to path. The default path is <Rhythmyx root>\<contentconnector>\<Name>.
The Descriptor Tree displays the path as its top node .
- 5 Enter the Rhythmyx Server Name and Port, and enter the Username and Password that you will use to connect to it.
- 6 *Choose and configure a Content Selector* (see "[Configuring Content Selectors](#)" on page 25).
- 7 *Choose and configure Static Item Extractors* (see "[Static Item Extractor Editor](#)" on page 38).
- 8 *Choose and configure Item Extractors* (see "[Item Extractor Editor](#)" on page 33).
- 9 If you want to include field or item transformations for any Item Extractor, *configure field transformations* or *item transformations* (see "[Configuring Item Transformers](#)" on page 52).
- 10 *Choose and configure a Content Loader* (see "[Configuring Content Loaders](#)" on page 54).
- 11 If you want to include field or item transformations for all instances of a field or for all items, configure *global field transformations* and *global item transformations* (see "[Configuring Item Transformers](#)" on page 52).
- 12 Configure *logging* and *error handling* (see "[Configuring Descriptor Error Handling](#)" on page 59).
- 13 *Save the Descriptor* (see "[Saving Descriptors](#)" on page 61).

Editing a Rhythmyx Enterprise Content Connector Descriptor

To edit an existing Descriptor:

- 1 Open the Content Connector Interface.

- 2 Choose *File > Open* or click .

A File Open dialog opens and displays the Enterprise Content Connector directory.

- 3 Browse to the Descriptor file that you want to open and click [**Open**]. (The Descriptor file is an xml file with a file name that you assigned it in the first screen of the Descriptor Setup dialog.)

The Content Connector Interface opens the *Descriptor Setup dialog* (on page 19) for the open Descriptor. The Descriptor tree displays folders for the Descriptor elements. The Editor pane displays the Descriptor's Name and Path. Rhythmyx Server, Port, User, and Password also display any information that you have entered.

- 4 Optionally, change the Name, Path, Rhythmyx Server, Port, User, and Password. and click [**Apply**].

The Descriptor Tree displays the new path as the top node in the Descriptor Tree.

- 5 Optionally, *change your choice of Content Selector* (see "[Configuring Content Selectors](#)" on page 25) and configure it. Optionally, *change your choices for Static Item Extractors* (see "[Static Item Extractor Editor](#)" on page 38) and configure them. Optionally, reconfigure previously added Static Item Extractors.

- 6 Optionally, *change your choices for Item Extractors* (see "[Item Extractor Editor](#)" on page 33) and configure them or reconfigure previously added Item Extractors. Optionally, add or reconfigure *field transformations* or *item transformations* (see "[Configuring Item Transformers](#)" on page 52) in extractors.

- 7 Optionally, *change your choice of Content Loader and reconfigure it* (see "[Configuring Content Loaders](#)" on page 54).

- 8 Optionally, add or reconfigure *global field transformations* or *global item transformations* (see "[Configuring Item Transformers](#)" on page 52).

- 9 Optionally, *reconfigure logging* or *error handling* (see "[Configuring Descriptor Error Handling](#)" on page 59).

- 10 *Save the Descriptor* (see "[Saving Descriptors](#)" on page 61).

Creating a New Descriptor by Modifying an Existing Descriptor

You may save an existing Descriptor as another name to create a new Descriptor. The original Descriptor will keep its original settings and filename and the new Descriptor will have the original settings until you modify it.

To create a new Descriptor by modifying an existing Descriptor:

- 1 In the Enterprise Content Connector Interface, open the Descriptor that you want to copy. Save it to another file by choosing *File > SaveAs* and entering a name for the new Descriptor in the SaveAs dialog.

- 2 Click [**Save As**].

The Enterprise Content Connector saves the new Descriptor in a folder with the same name and opens the *Descriptor Setup dialog* (on page [19](#)) so you can modify it.

Adding a Descriptor from another Host

Someone working on another host machine may create a Descriptor file and send it to you for testing purposes or to allow you to use an upload configuration that has already been constructed.

To make sure that the Descriptor opens correctly in your Enterprise Content Connector interface:

- 1** Insert the Descriptor file into a folder with the same name as the Descriptor file. For example, if the Descriptor file is named `TestUpload.xml`, create a folder named `TestUpload` and insert it into that folder.
- 2** The initial time that you open the Descriptor file in the Enterprise Content Connector, immediately open it in the Descriptor Setup dialog and save it. This changes the log file specified in the Descriptor file to the current host and prevents problems opening and modifying the Descriptor file.

Configuring Content Selectors

A Content Selector plugin looks for files on an external site and creates a Content Tree showing the scanned files. The default Content Selectors are either "for migration" or "for syndication." The Selectors for syndication perform a *checksum* procedure to determine if files have changed. The checksum procedure, which calculates a value for the content in the previous and current version of the file and then compares the values to see if they are the same, is time consuming. In most cases, migration is a one-time activity performed when you initially install Rhythmyx. Therefore, the Selectors specified for migration do not perform the checksum procedure.

Syndication usually involves scheduled uploads of content from a specified location. Most likely you only want to upload content that you have not uploaded previously, so the Selectors specified for syndication do not upload content that was previously uploaded and is unchanged.

By default, the Enterprise Content Connector provides four Content Selector Options:

- File Selector for migration - Crawls files from specified external roots without calculating checksums of files. Uploads files as New if they were not previously uploaded and as Changed if they were previously uploaded.
- File Selector for syndication - Crawls files from specified external roots and calculates checksums of files. Uploads new files and files that have changed since previous upload; does not upload previously uploaded files that have not changed.
- List Selector for migration - Finds files using a list of filenames in an XML file created manually or by a third party tool. Does not calculate checksums of files. The List Selector file must conform to `ListContentSelector.dtd`, located in `<Rhythmyx root>\DTD`. Uploads files as New if they were not previously uploaded and as Changed if they were previously uploaded.
- List Selector for syndication - Finds files using a list of filenames in an XML file created manually or by a third party tool. Calculates checksums of files. The List Selector file must conform to `ListContentSelector.dtd`, located in `<Rhythmyx root>\DTD`. Uploads new files and files that have changed since previous upload; does not upload previously uploaded files that have not changed.

Choosing a Content Selector

To choose a Content Selector type:

- 1 In the *Descriptor Setup dialog* (on page 19), click the Content Selector folder in the Descriptor Tree.

The Editor pane displays unselected Content Selectors in the Content Selector List. If you have already selected a Content Selector, the Content Selector node in the Descriptor Tree includes a node for it.

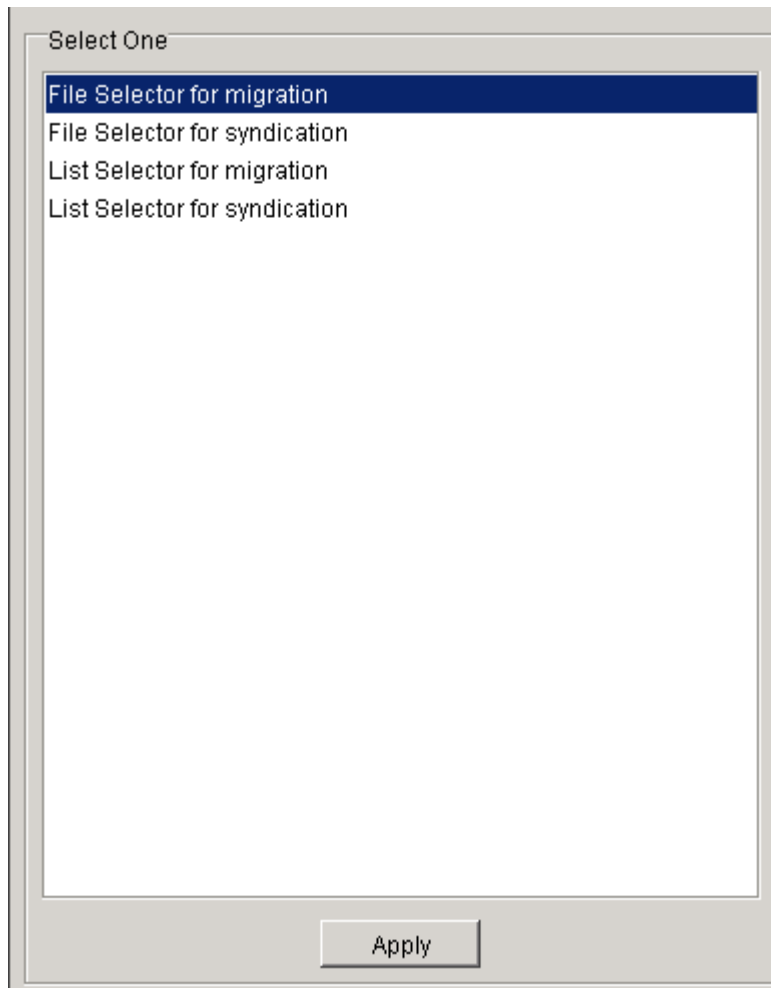


Figure 13: Content Selector List

- 2** Choose a Content Selector.
- 3** Click [**Apply**].

The Content Selector folder in the Descriptor Tree now includes a node for the Content Selector.

- 4** To change your choice, click a different Content Selector in the selection list and click [**Apply**].

The originally selected Content Selector is now listed in the selection list, and Content Selector folder in the Descriptor Tree includes a node for the newly selected Content Selector.

Configuring a List Content Selector

To configure a List Content Selector:

- 1 In the *Descriptor Setup dialog* (on page 19), click the List Content Selector node in the Descriptor Tree.

The Editor pane displays the List Content Selector Editor.

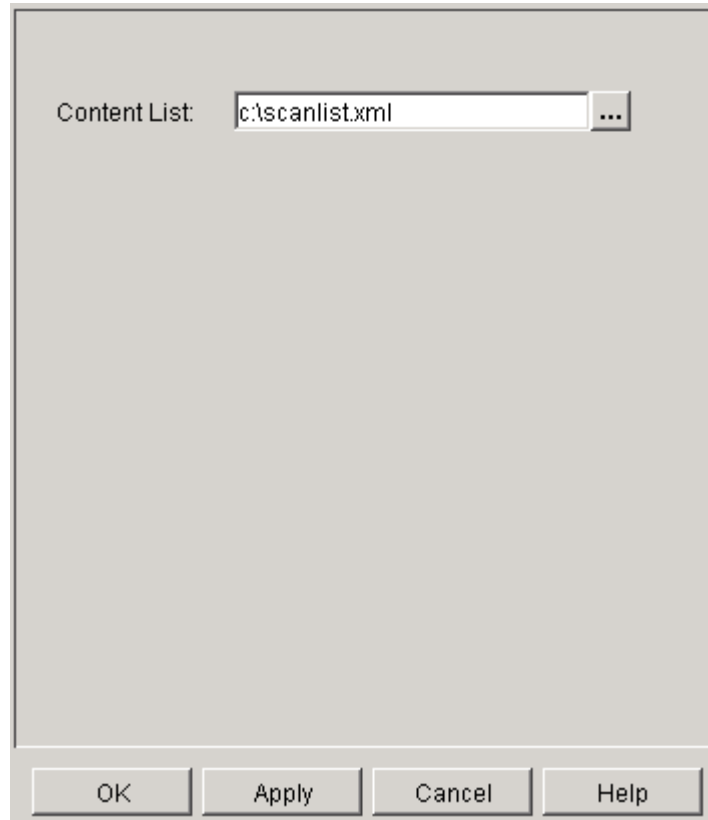


Figure 14: List Content Selector editor

- 2 In Content List enter the fully qualified name of the file containing the list of files to scan.

Configuring a File Content Selector

To configure a File Content Selector:

- 1 In the *Descriptor Setup dialog* (on page 19), click the File Content Selector node in the Descriptor Tree.

The Editor pane displays the File Content Selector Editor. When you initially open the File Content Selector Editor for a Descriptor, only the Default tab is present.

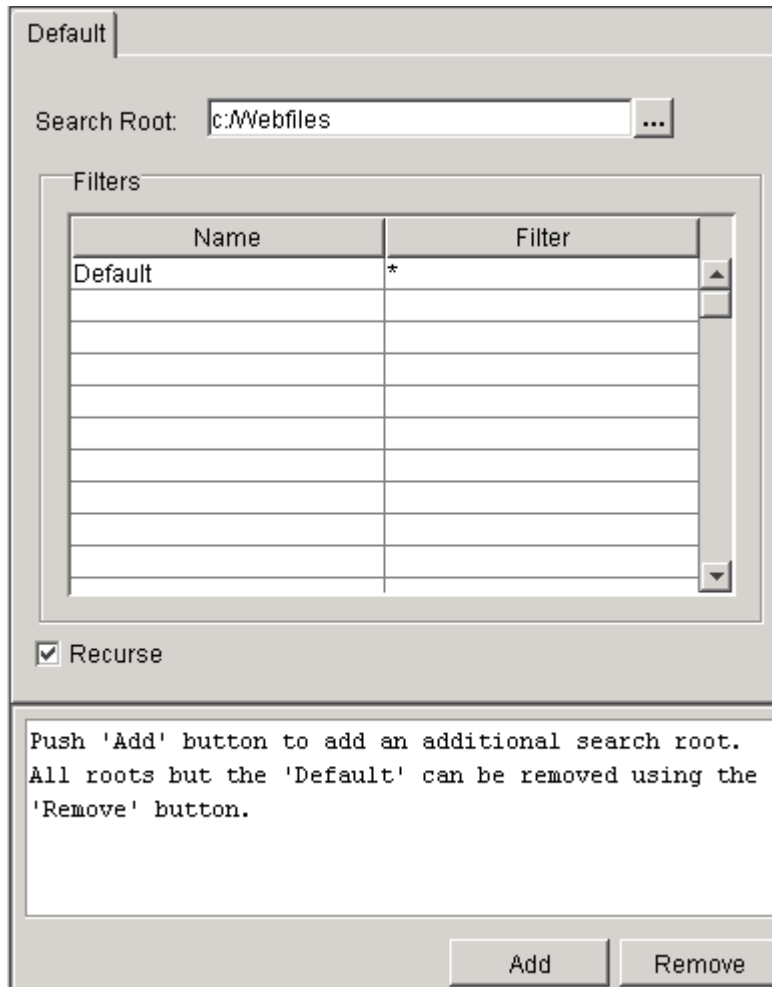


Figure 15: File Content Selector editor

- 2 In **Search Root** enter the root directory for beginning a content search. Required.
- 3 Under **Name**, enter names of filters for matching files.
- 4 Under **Filter**, enter the strings for files to match. The default is * (all files). *See **Configuring a File Content Selector Filter** (on page 29) for information about configuring filter strings.*
- 5 Check **Recurse** to search sub-directories of the Search Root. Checked by default.
- 6 Click [**Add**] to add an additional search root tab, and when prompted, enter a tab name.
The Editor adds the new tab, and by default, fills all fields except the Search Root with the settings in the Default tab. Change any of the settings.
- 7 Add any number of additional tabs.
- 8 Select a tab and click [**Remove**] to remove it.

Configuring a File Content Selector Filter

The filter for File Content Selectors uses the following wildcards for matching. All wildcard matches are case-insensitive:

Wildcard Expression	Function
?	Matches one character, and considers the path separator.
*	Matches from zero to any number of characters, and considers the path separator.
**	Used for subfolder matching. Matches from zero to any number of characters, and ignores the path separator.

Examples:

The **Search Root** is C:/Webfiles and the folder tree under it is:

```

\Merchandise
  +---Online
    +---Colors
      \---Primary
    +---Sizes
      \---typical
    +---Styles
    +---Years
    +---Categories
      \---Patterns
        \---Popular
          \---Sale
            \---Discontinued
    
```

Filter Pattern	Returns
**/s*.pdf	Any pdf file that starts with “s” under the Webfiles folder or any of its subfolders
Merchandise/Online/**/* .txt	Any txt file under any subfolder of the Online folder
Merchandise/Online/**/image?.*	Any file that starts with “image” and then has exactly one of any character followed by . [any extension] and is located under any subfolder of Online. Examples: image1.gif imageA.gif NOTE: image44.gif is not returned.
Merchandise/Online/**/* .html	Any HTML file under any first level subfolder of the Online folder.

Filter Pattern	Returns
Merchandise/Online/Styles/*	Any file under the Styles folder
Merchandise/Online/*.gif	Any gif file directly under the Online folder.

Maintaining Content Extractors

During scanning, Content Extractors search for files on the source site that match file extensions or other metadata that they specify. During uploading, Content Extractors retrieve content on the external site and if necessary, transform it to conform to Rhythmyx Content Types. Extractors also assign Workflow Actions to be applied to items after they enter Rhythmyx.

The Enterprise Content Connector provides two types of Content Extractors:

- **Static Item Extractor** - Finds and extracts source files that represent non-content items stored in Rhythmyx, such as images and stylesheets used as Web resources. The Enterprise Content Connector refers to non-content items that it uploads as static items.

The Enterprise Content Connector provides a default Static Item Extractor class and provides two default Static Item Extractors based on the class

- `$rximage` - finds and extracts Rhythmyx static image files;
- `$rxcss` - finds and extracts cascading style sheet files.

NOTE: The Enterprise Content Connector finds the context variable (publisher variable) for Static Item Extractors in the *Enterprise Content Connector configuration file* (see "[Changing Configuration File Default Values](#)" on page 101). The default Static Item Extractor publisher variables `$rximage` and `$rxcss` are registered by default in Rhythmyx, but you must register any that you add. See the section *Splitting Context-Dependent Resources* in the Rhythmyx Online Help for instructions on registering publisher variables.

- **Item Extractor** - Finds and extracts files that represent Rhythmyx Content Items. Item Extractors may contain Field or Item Transformers, which transform the extracted field or item. There is no default transformer in the default configuration file. If you add a Field or Item Transformer to an Item Extractor, it appears in a subfolder of the Extractor.

There are four default Item Extractor classes:

- **Page Extractor class** - Finds and extracts HTML files. By default, the Enterprise Content Connector provides the default Page Extractor based on the class. It assigns extracted content to the Page Content Type.
- **Binary Extractor class** - Finds and extracts binary files. By default, the Enterprise Content Connector provides two default Item Extractors based on the class:
- **Image Extractor** - Finds and extracts Image Content Types.
- **File Extractors** - Finds and extracts File Content Types.
- **XML Extractor class** - Finds and extracts XML files. By default, the Enterprise Content Connector provides the default XML Extractor based on the class. It assigns extracted content to the Page Content Type.
- **XSL Extractor class** - Finds and extracts XML files and formats them using an XSL file before assigning them to a Content Type. By default, the Enterprise Content Connector provides the default XSL Extractor based on the class. It assigns extracted content to the Page Content Type.

Managing the Extractors Assigned to a Descriptor

When expanded, the Extractors folder in the Descriptor Tree displays folders for Static Items and Items. The Static Items folder holds Static Item Extractors and the Items folder holds Item Extractors. When you select the Static Items folder or the Items folder, the Editor pane displays a window that lists the **Available** Extractors and **Used** Extractors. The **Available** Extractors are all of the Extractor types available for either Static Item Extractors or Items Extractors. The **Used** Extractors are the **Available** Extractors that you have chosen to use in the Descriptor.

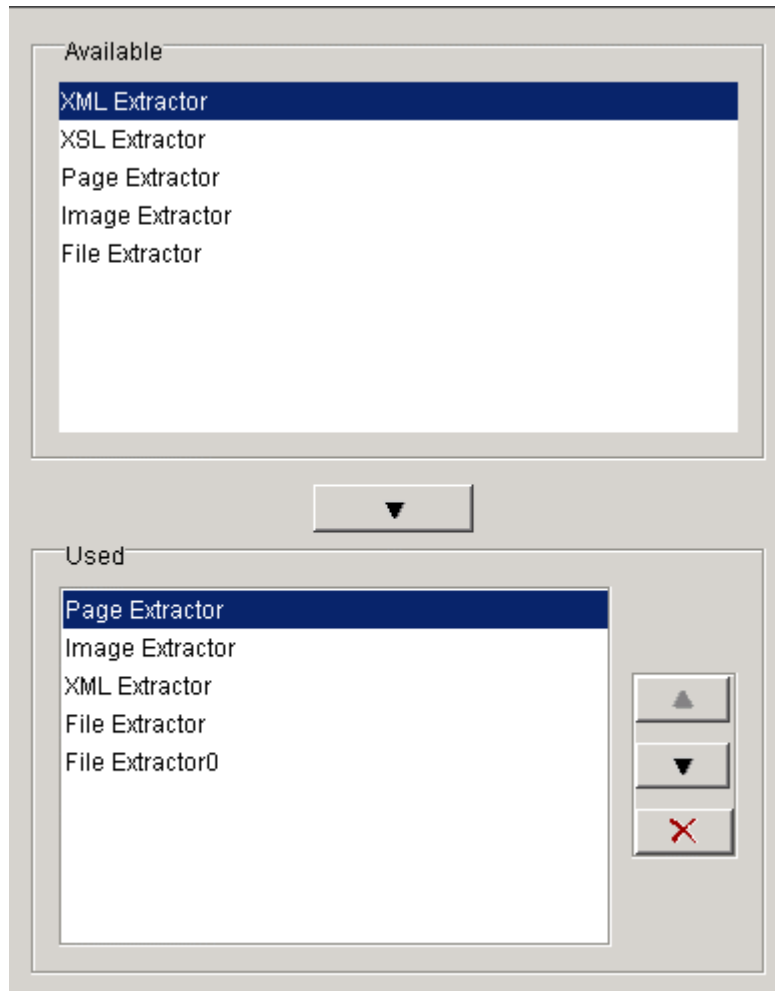


Figure 16: Available and Used Extractors window

To add an **Available** Extractor to the **Used** Extractor box and to the Descriptor Tree:

- 1 In the **Available** box, select each type of Extractor that you want to include and click



The Extractor is added to the **Used** box. It remains listed in the **Available** box. In the Descriptor Tree, the Extractor folder displays a node for the added Extractor.



To remove an Extractor from the **Used** Extractor box and the Descriptor Tree:

- 1 In the **Used** box, select each type of Extractor that you want to remove and click



The Extractor no longer appears in the **Used** box, and the Extractor folder in the Descriptor Tree no longer displays a node for the Extractor.

To change the order of an Extractor in the Descriptor Tree:

- 1 In the **Used** box, choose the Extractor whose order you want to change.
- 2 Click  to move the Extractor down.
- 3 Click  to move the Extractor up.

Extractor Editors

After you add Extractors to the Static Items or Items folder in the Navigation Tree, you can click on the specific Extractor nodes to display editors for configuring the Extractors in the Editor pane. The Static Item Extractor editor and the Item Extractor editor have different tabs and require different information.

Item Extractor Editor

Item Extractors extract data and create Rhythmyx Content Items from the data. The default Item Extractors extract HTML, binary, and XML data which they insert as the body of Content Items that they create.

When you click on a specific Item Extractor node in the Dependency Tree, the Edit pane displays the Item Extractor Editor, which has tabs for filling in information about the Item Extractor's Content Type, Fields, Filters, and Workflow.

System Field Value Map: `sys_title` is Required. Mappings of system fields used by the Item Extractor to value types and values.

- **Field** is a drop list including all valid system fields except `sys_workflowid` and `sys_communityid`, which are included in other locations on the Item Extractor Editor.
- **Value Type** is a drop list with the choices Variable, Literal, Number, and Date. The drop list also includes the choice Xpath for XML and XSL Extractors.
- **Value** either allows you to enter a literal or lets you choose from a drop list depending on the **Field**. If you choose a **Value Type** of *Variable*, **Value** displays a drop list of Extractor variables. If you choose a **Value Type** of *Date*, the preferred date format is `yyyy-mm-dd`. For other acceptable date formats, see the Rhythmyx Knowledge Base.

Fields Tab

The Fields tab specifies values of fields that are specific to an Item Extractor type. See the individual Item Extractor topics for descriptions of the Fields tabs.

XSL Extractor (on page [43](#))

XML Extractor (on page [44](#))

Binary Extractor (on page [45](#))

Page Extractor (on page [46](#))

Filters Tab

The Filters tab of the Item Extractor Editor includes a list of the filters used with the Item Extractor and a mapping of file mime types to file extensions.

The screenshot shows the 'Filters' tab of the Item Extractor Editor. It has four tabs: 'Content Type', 'Fields', 'Filters', and 'Workflow'. The 'Filters' tab is active. It contains two main sections:

- Filters:** A table with a header 'Fields' and one row containing '*.htm'.
- Mime Type Map:** A table with two columns: 'Name' and 'Extensions'. The first row shows 'text/plain' in the 'Name' column and 'htm, html' in the 'Extensions' column.

Figure 18: Item Extractor editor, Filters tab

Filters Tab Fields:

Filters - URL patterns that the Item Extractor uses to match the URL of scanned items. The URL pattern may be the directory where a Content Type extracted by the Item Extractor is located or a regular file expression.

For example, if you were configuring the Image Item Extractor, and the image files to be extracted were located in the resource/image and the art/image directories, then you would specify the **Filters** `*/resource/image/*` and `*/art/image/*`.

Mime Type Map - File extensions of items to be extracted. If you do not specify a file extension here, the Item Extractor will not extract files with that extension.

- **Name** - Mime type to assign to Content Item;
- **Extensions** - Mappings of extracted files' extensions to the mime types to be assigned to the Content Items that they create.

Workflow Tab

When a content item is uploaded from the Enterprise Content Connector to Rhythmyx, it acquires the Workflow that you assign it in the Workflow tab. You must specify the Workflow Transitions that are applied to New and Updated items of each item extractor type so they are transitioned to the appropriate State; you cannot simply specify a State. If you want the content item to have its initial Workflow State, do not enter any Workflow Transitions.

You enter the Workflow Transitions through the Workflow tab of the Item Extractor Editor. After a new or changed item enters Rhythmyx, the Workflow Actions are applied. The actor recorded as performing the Actions is the user supplied in the Descriptor's **Username** field. Therefore, after new and changed content items of this type are uploaded into Rhythmyx, they have the State following the last Workflow Transition that you specified.

When new items are uploaded, they are checked in; when changed items are uploaded, they are checked in if they were previously checked out.

The Workflow tab of the Item Extractor Editor includes a field for the Workflow to be assigned to the Content Item created by the Extractor and a list of the Workflow Transitions to be applied to the Content Item upon creation or update.

The screenshot shows the Workflow tab of the Item Extractor Editor. At the top, there are four tabs: 'Content Type', 'Fields', 'Filters', and 'Workflow', with 'Workflow' being the active tab. Below the tabs, there are two dropdown menus: 'Workflow' (set to 'Article') and 'Transition performed on' (set to 'Insert new item'). Below these is a table titled 'Transitions' with two columns. The first column contains the text 'Submit (Submit)' and 'Approve (Approve)', and the second column contains a small upward-pointing triangle. The table has several empty rows below it, and a vertical scrollbar is on the right side.

Transitions	
Submit (Submit)	▲
Approve (Approve)	

Figure 19: Item Extractor editor, Workflow tab

Workflow Tab Fields:

Workflow - Drop list of valid **Workflows** for the **Community** selected in the Content Type tab.

Transition Performed On - Case in which the specified Transitions are performed on Content Items.

Options are:

- *Insert New Item* - Perform transitions when inserting New Items
- *Pre-update Item* - Perform transitions on Update Items before updating them. Since Rhythmyx does not allow you to modify items in a Public State, this option is included so that you may apply an Ignore Publish Transition to a Public item, which moves it into a Quick Edit State. In a Quick Edit State, you may update the item.
- *Post-update Item* - Perform transitions on Update Items after updating them.

You may define transitions for each of these cases in a Workflow or leave any of the cases blank.

Transitions - Drop list specifying valid Transitions for the selected **Workflow**. Specify those Transitions that move the Content Item to the State in which you want it to be in Rhythmyx. NOTE: The first term is the Transition; the second term is the Transition Trigger.

Static Item Extractor Editor

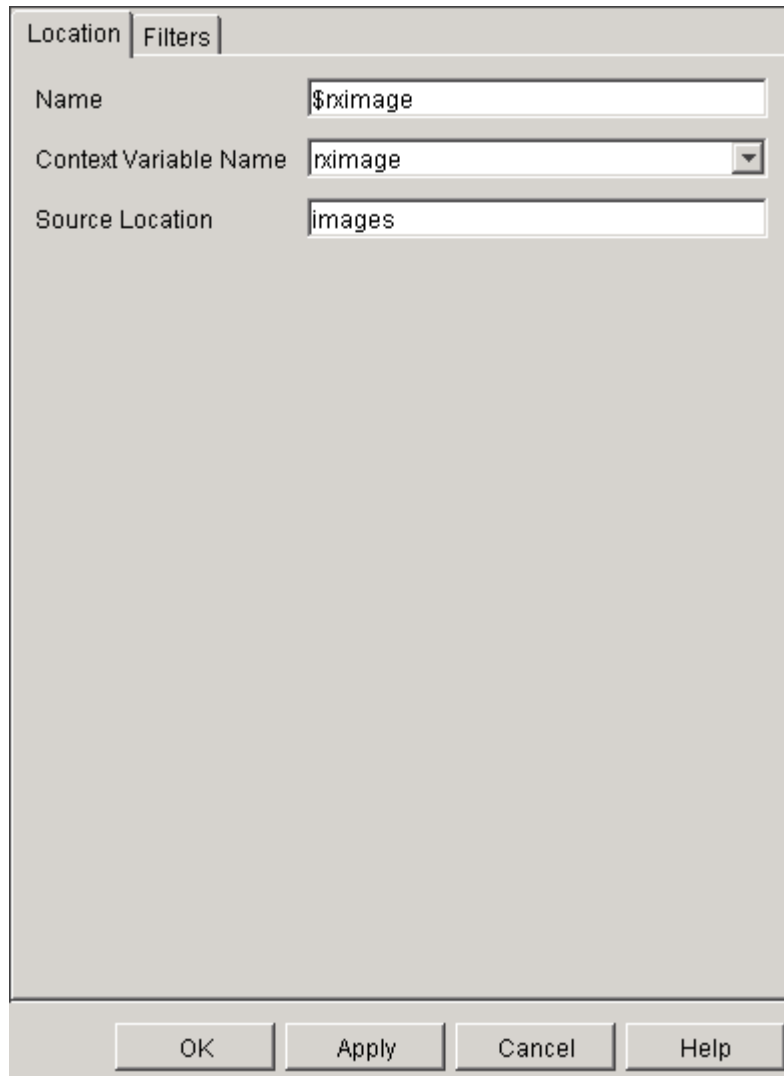
Static Item Extractors extract files and create static Rhythmyx items, such as stylesheets and logo images used on Web sites, from the data. The default Static Item Extractors extract static image and stylesheet files which they insert into Rhythmyx Web resource directories.

When you click on a specific Static Item Extractor node in the Dependency Tree, the Edit pane displays the Static Item Extractor Editor, which has tabs for filling in information about the Static Item Extractor's Location and Filters.

The Filters tab of the Static Item Extractor Editor functions in the same way as the Filters tab of the Item Extractor Editor. See the **Filters Tab** (on page 36) topic for the Item Extractor Editor for instructions on using it.

Location Tab

The Location tab of the Static Item Extractor Editor includes fields for the Extractor's **Name**, the **Context Variable Name**, and the **Source Location**.



The screenshot shows a dialog box titled "Static Item Extractor Editor" with two tabs: "Location" and "Filters". The "Location" tab is active. It contains three input fields: "Name" with the value "\$ximage", "Context Variable Name" with a dropdown menu showing "ximage", and "Source Location" with the value "images". At the bottom of the dialog are four buttons: "OK", "Apply", "Cancel", and "Help".

Figure 20: Static Item Extractor editor, Location tab

Location Tab Fields:

Name: The name of the Extractor.

Context Variable Name: A drop list of context variable names taken from the Rhythmyx server.

Source Location: Directory from which to retrieve source files. Directory is relative to the Search Root in the Selector.

Adding a New Extractor

You may create custom Item Extractors and Static Item Extractors based on the default Extractors by adding them to your Descriptor in the Descriptor Setup dialog. The following example demonstrates how to add a custom Item Extractor based on a default Item Extractor. You can use the same procedure for adding a custom Static Item Extractor.

In this example, we add a custom Item Extractor based on the Image Item Extractor, and begin with the following assumptions:

- the user must upload a group of image files into the Content Type CusImage;
- the fields to upload into each Content Item are `cus_imgbody`, `cus_imgbody_type`, `cus_imagebody_ext`, and `cus_imagebody_size`;
- the user has already added a default Image Item Extractor to the Descriptor, but an additional one, meeting the above specifications, is required.

To create the custom Image Item Extractor:

- 1 In the Navigation Tree of the Descriptor Setup Dialog, expand the Extractors folder and click the Items folder.

The editor pane displays a window that lists your **Available** and **Used** Extractors. Image Extractor appears in **Used** and in the Descriptor Tree.

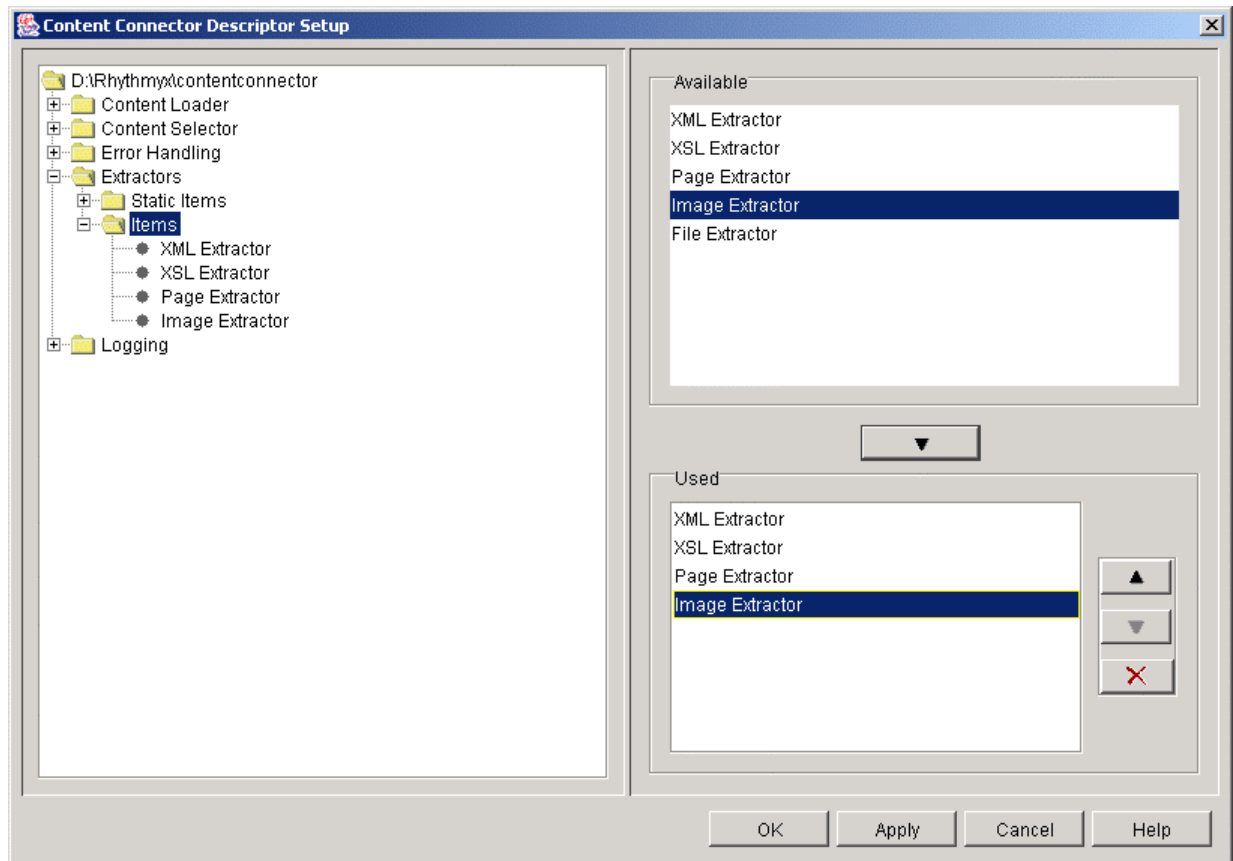



Figure 21: Adding a new Image Extractor

- 2 In the **Available** box, click Image Extractor and click .

The Editor adds Image Extractor0 to the **Used** box and the Descriptor Tree.

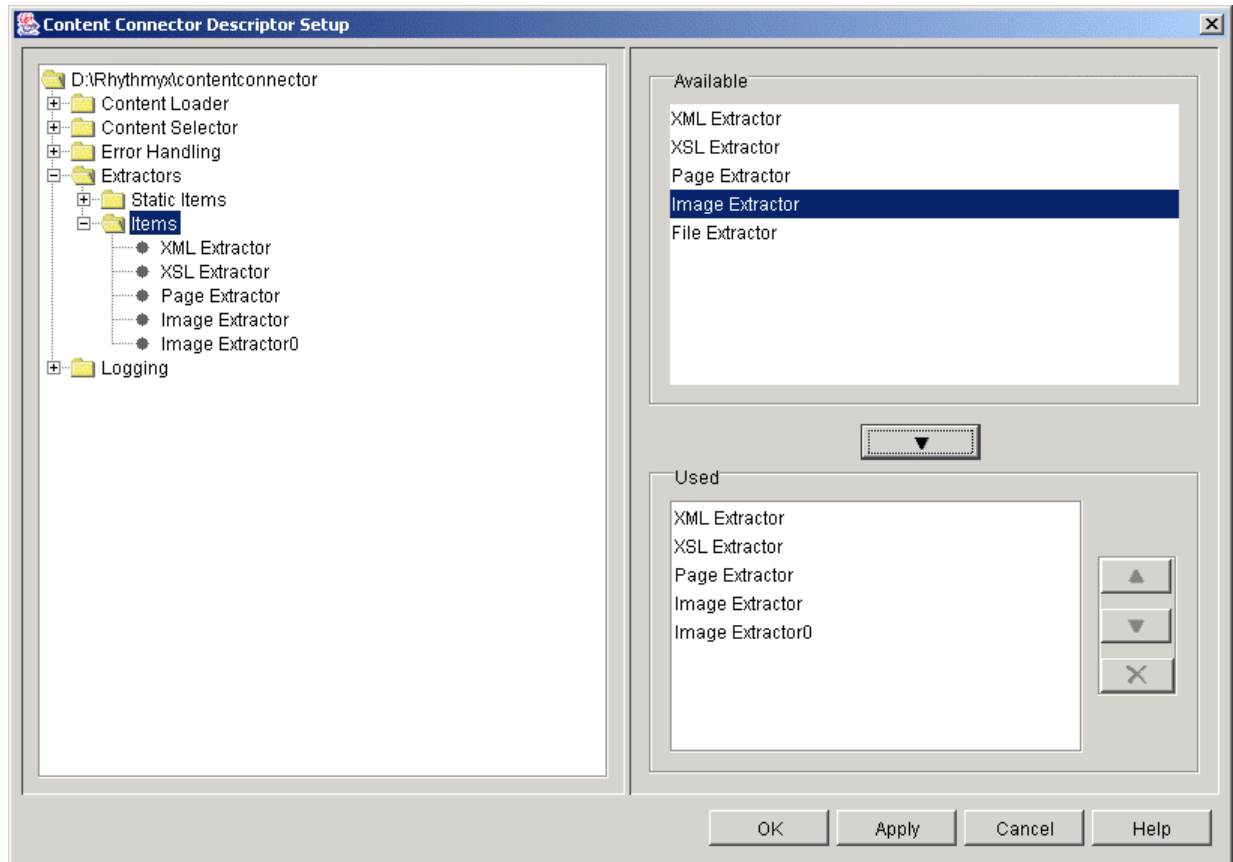


Figure 22: The new Image Extractor's name is Image Extractor0

- 3 In the Descriptor Tree, click the Image Extractor0 folder.

The Editor pane displays Item Extractor editor filled in with the fields from the original Image Extractor. We want to use different settings on the Content Type and Fields tabs.

- 4 On the Content Type tab, set the fields as follows:
 - Name - Customer Image Extractor
 - Community - Customer Community
 - Content Type - CusImage
 - System Field-value map - (leave sys_title mapped as a variable to \$filename.)

- 5 On the Fields tab, change the **Fields** as follows:

Field	Value Type	Value
cus_imgbody	Variable	\$content
cus_imgbody_type	Variable	\$mimetype
cus_imgbody_filename	Variable	\$filename
cus_imgbody_ext	Variable	\$fileextension
cus_imgbody_size	Variable	\$filesize

- 6 Click [**OK**] to save the Customer Image Item Extractor to the current Descriptor.

Configuring an Extractor

To configure or modify an existing Extractor, click on the node for the Extractor in the Descriptor Tree and change the settings in any of the tabs.

To configure an Item Extractor:

- 1 In the Descriptor Tree, expand the Extractors folder and then the Items folder.
- 2 Click on the node for the specific Extractor that you want to configure or modify.
- 3 In the Content Type tab, change the **Name**, **Community**, **Content Type** or **Field Mappings**. You may remove all field mappings except for `sys_title`.
- 4 In the Fields tab, change any of the field settings or mappings.
- 5 In the Filters tab, add, remove, or modify any **Filters** and **Mime Type** mappings.
- 6 In the Workflow tab, change the **Workflow** and/or the **Transitions**.

To configure a Static Item Extractor:

- 1 In the Descriptor Tree, expand the Extractors folder and then the Static Items folder.
- 2 Click the node for the specific Extractor that you want to configure or modify.
- 3 In the Location tab, you may change the **Name**, **Context Variable**, or **Source Location**.
- 4 In the Filters tab, add, remove, or modify any **Filters** and **Mime Type** mappings.

Default Extractors

XSL Extractor

The XSL Extractor finds and extracts XML files and formats them using an XSL file before inserting them into a Content Item. The XSL converts the XML to a standard item, which is specified by `sys_StandardItem.xsd` located in `<Rhythmyx root>\WebServices`.

Specific settings for the XSL Extractor in the Item Extractor Editor are as follows:

- In the Content Type tab, the default Content Type for XSL Extractor is Page. Fields set in the System Field-Value Map are only used if the transformed standard item does not set values for those fields.
- The Fields tab includes a single field for **XSL File**. The **XSL File** field specifies the path of the XSL file that will transform the extracted XML document. You may enter the path or select it using `...`.

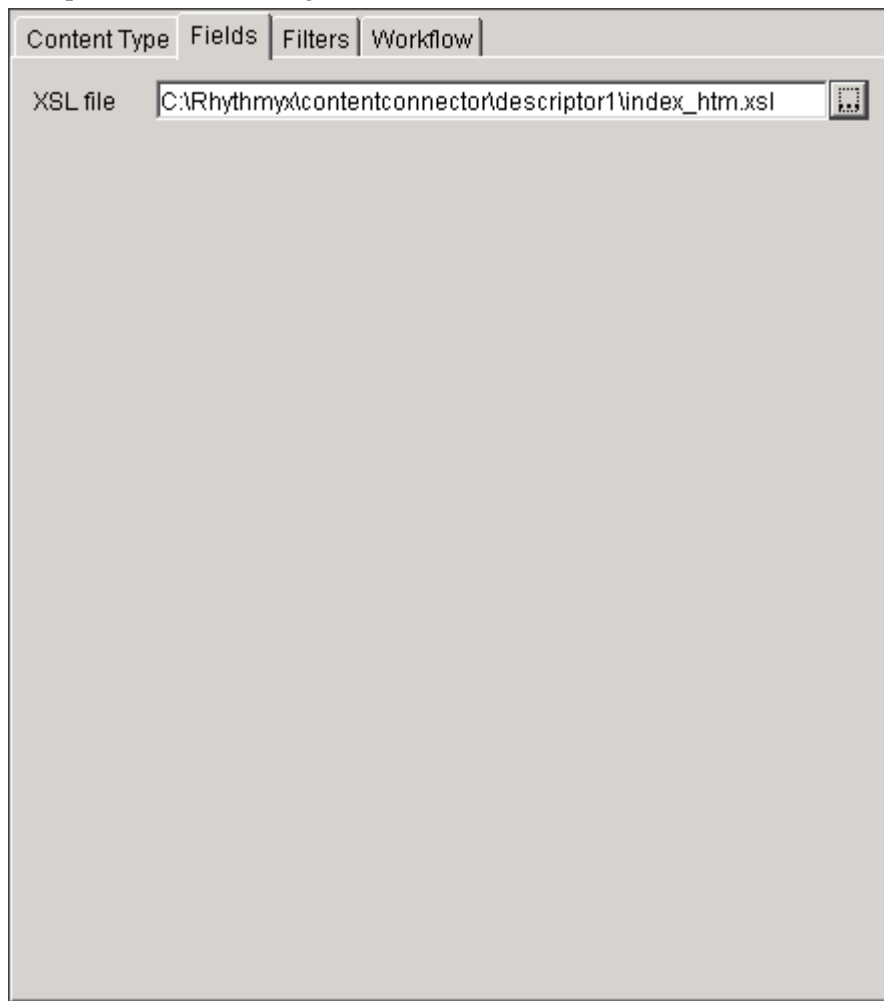


Figure 23: XSL Extractor, Fields tab

XML Extractor

The XML Extractor finds and extracts XML files and inserts values in XML elements into fields of the Content Item.

Specific settings for the XML Extractor in the Item Extractor Editor are as follows:

- In the Content Type tab, the default Content Type for XML Extractor is Page.
- The Fields tab includes a **Field-value map** that maps Content Type **Field** names to **Values** in the extracted XML. **Value Type** is a drop list with the choices Literal, Number, Variable, or Xpath. If you choose Variable, the **Value** field is a drop list of ECC Variables; if you choose any of the other **Value Types**, the **Value** field is a text entry field.

If you choose Xpath, and the Xpath **Value** includes more than one node, the Field is set to the first node and the other nodes are ignored. If an item is being updated, all old child entries are deleted and all newly extracted child entries are added.

The screenshot shows the 'Fields' tab of the XML Extractor configuration. It features a 'Field-value map' table with three columns: Field, Value Type, and Value. The table contains two rows of data.

Field	Value Type	Value
sourceurl	Variable	\$filepath
content	Variable	\$content

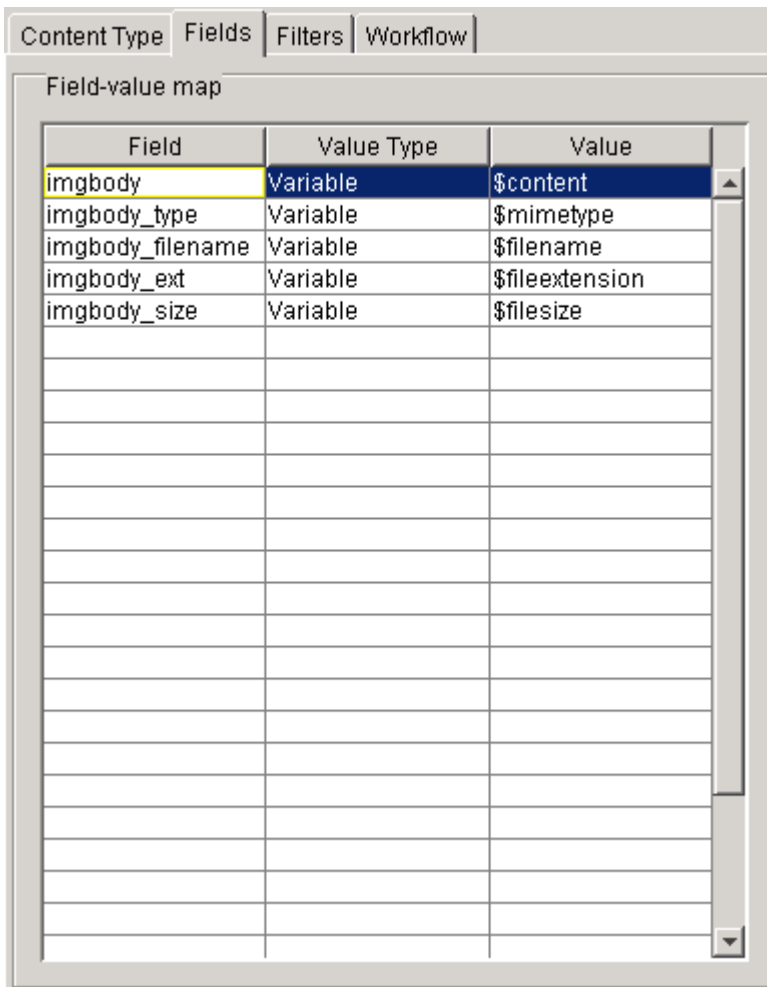
Figure 24: XML Extractor, Fields tab

Binary Extractor

The Binary Extractor finds and extracts binary files. There are two types of default Binary Extractors: Image Extractor and File Extractor.

Specific settings for the Binary Extractors in the Item Extractor Editor are as follows:

- In the Content Type tab, the default Content Type for the Image Extractor is Image, and the default Content Type for the File Extractor is File.
- The Fields tab includes a **Field-value map** that maps Content Type Field names to values. **Value Type** is a drop list with the choices Literal, Number, or Variable. If you choose Variable, the **Value** field is a drop list of ECC Variables; if you choose any of the other **Value Types**, the **Value** field is a text entry field.



The screenshot shows the 'Fields' tab of the Binary Extractor configuration. It features a 'Field-value map' table with three columns: 'Field', 'Value Type', and 'Value'. The table contains five rows of data, with the first row highlighted. The 'Field' column lists 'imgbody', 'imgbody_type', 'imgbody_filename', 'imgbody_ext', and 'imgbody_size'. The 'Value Type' column lists 'Variable' for all entries. The 'Value' column lists '\$content', '\$mimetype', '\$filename', '\$fileextension', and '\$filesize' respectively.

Field	Value Type	Value
imgbody	Variable	\$content
imgbody_type	Variable	\$mimetype
imgbody_filename	Variable	\$filename
imgbody_ext	Variable	\$fileextension
imgbody_size	Variable	\$filesize

Figure 25: Binary Extractor, Fields tab

Page Extractor

The Page Extractor finds and extracts HTML files.

Specific settings for the Page Extractor in the Item Extractor Editor are as follows:

- In the Content Type tab, the default **Content Type** for the Page Extractor is Page, and the default Field-Value mappings are:
 - sys_title is mapped to \$filename
 - sys_path is mapped to \$filepath
- The Fields tab includes an **Encoding** field that specifies the encoding to use to convert the content from bytes to string. By default it is your current JVM setting, but you may choose other valid encodings from the drop list.

The Fields tab also includes a **Field-value map** that maps Content Type Field names to values. **Value Type** is a drop list with the choices Literal, Number, or Variable. If you choose Variable, the **Value** field is a drop list of ECC Variables; if you choose any of the other **Value Types**, the **Value** field is a text entry field.

Field	Value Type	Value
sourceurl	Variable	\$filepath
content	Variable	\$content

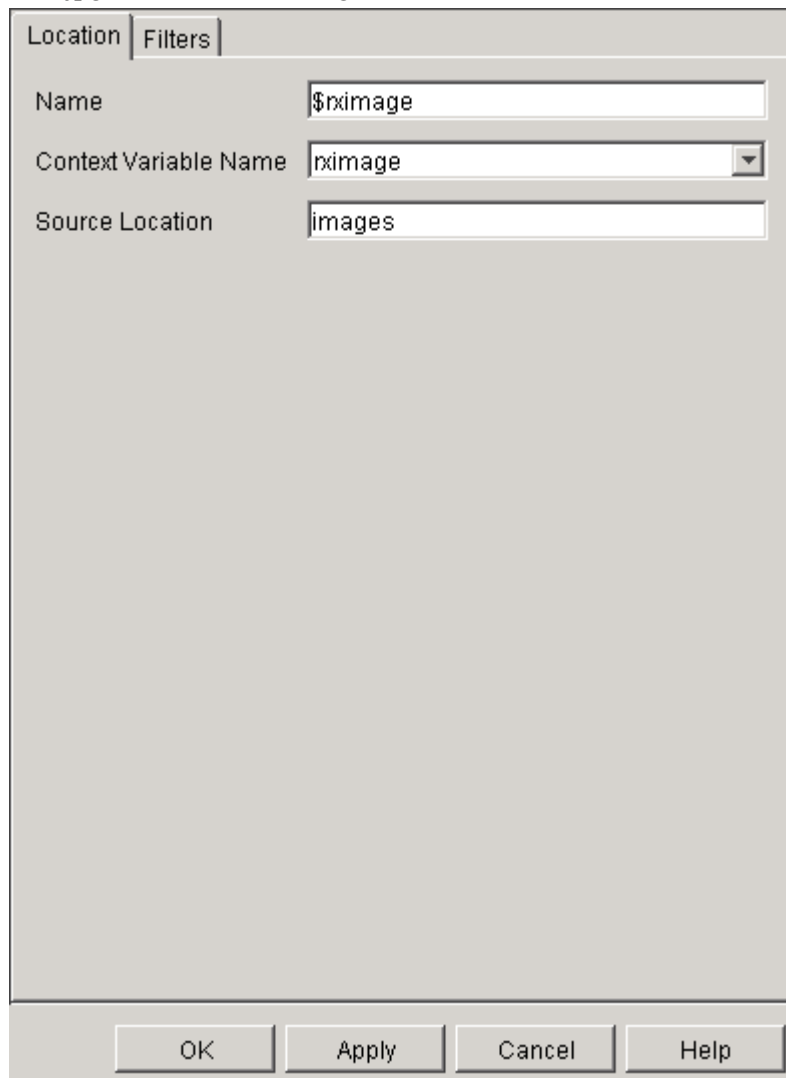
Figure 26: Page Extractor, Fields tab

Static Image Extractor

The Static Extractor finds and extracts non-content files used by Rhythmyx, such as style sheets and Web images. There are two types of default Static Extractors: \$rximage Extractor and \$rxcss Extractor.

Specific settings for the Static Extractors are as follows:

- In the Location tab, the default **Names** and **Context Variable Names** correspond to the files being extracted (for example \$rximage and rximage), and the Source Location is relative to the Search Root in the Selector (for example images).
- In the Filters tab, the **Filter** is * (select all), and the Mime Type Map maps image gif and jpg files for the \$rximage Extractor, and text css files for the \$rxcss Extractor.



The screenshot shows a dialog box titled "Static Item Extractor editor" with two tabs: "Location" and "Filters". The "Location" tab is active. It contains three input fields: "Name" with the value "\$rximage", "Context Variable Name" with a dropdown menu showing "rximage", and "Source Location" with the value "images". At the bottom of the dialog are four buttons: "OK", "Apply", "Cancel", and "Help".

Figure 27: Static Item Extractor editor, Location tab

Extractor Variables

Extractor variables specify some of the field values in an Extractor's definition in the Descriptor. The Extractor variables are predefined.

Variable	Description
\$filename	Name of source file or URL, not including file extension and parent directories.
\$filepath	Path of source file or URL relative to search root.
\$fileextension	File extension of source file or URL.
\$filesize	Size in bytes of source file or URL.
\$mimetype	Mime type of source file or URL.
\$content	Content type of source file or URL. Value is binary for binary Extractors and text/character for text Extractors.
\$currentdatetime	Value of the current date and time.

Configuring Transformers

Transformers are Java extensions that modify extracted items to conform to Rhythmyx formats before they are uploaded to Rhythmyx. Transformers are either specific to a Content Type (an Extractor type) or are global, and they may apply to fields or items. Rhythmyx does not include default transformers.

Configuring Field Transformers

To configure a global or content type-specific field transformer:

- 1 In the *Descriptor Setup dialog* (on page 19), click a Field Transformer folder in an Extractor folder or click the Global Field Transformers folder.

The Editor pane displays the Field Transformer List. The available Field Transformers are listed in the **Available** box. The Field Transformers that you are already using, if any, are listed in the **Used** box.

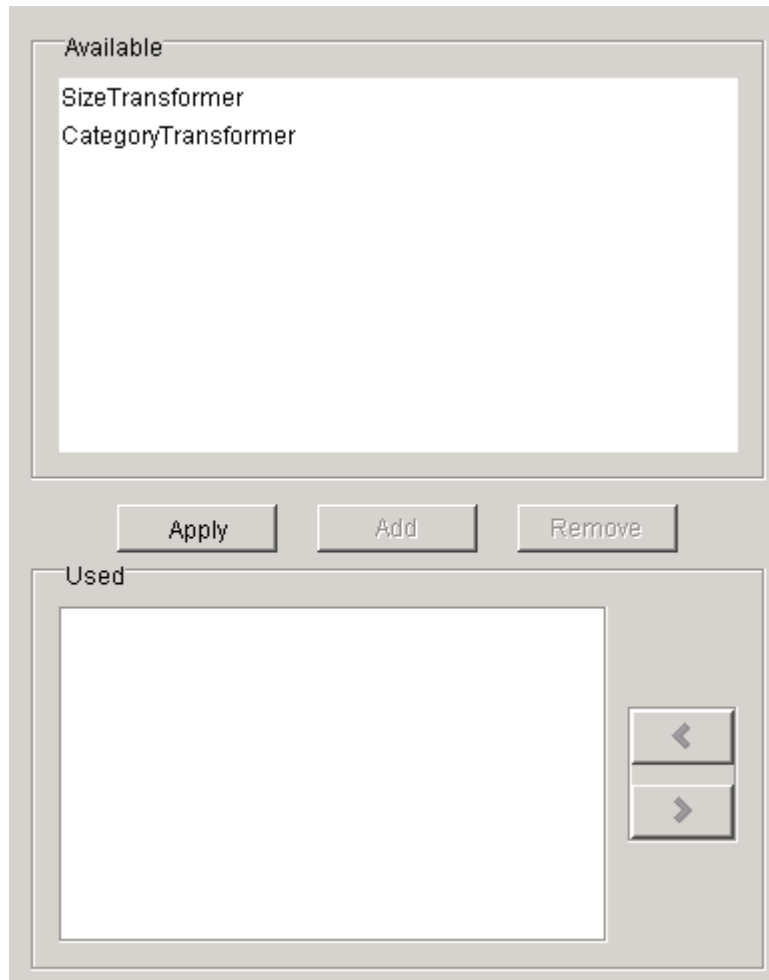


Figure 28: Field Transformer list

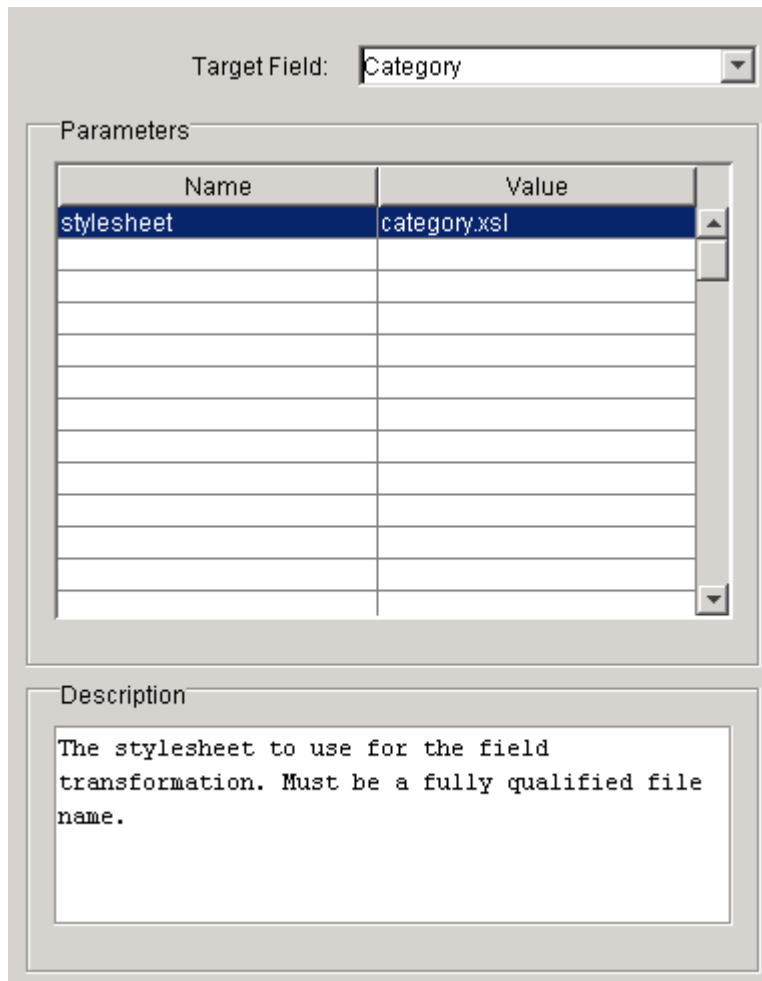
- 2 In the **Available** box, select each Field Transformer that you want to include and click [**Add**].
The Field Transformer moves to the **Used** box.
- 3 In the **Used** box, select each type of Field Transformer that you want to remove and click [**Remove**].
The Field Transformer moves to the **Available** box.
- 4 When you have finished selecting Field Transformers, click [**Apply**].

The Field Transformer folder includes a node for each Field Transformer. If you have removed any Field Transformers, they no longer appear under the Field Transformer folder.

- 5 Click on a Field Transformer node.

The Editor pane displays the Field Transformer editor. **Target Field** displays all fields for the Content Type.

The **Name** column lists all default parameters for the extension. The **Value** lists all default parameter values for the field.



Name	Value
stylesheet	category.xsl

Description

The stylesheet to use for the field transformation. Must be a fully qualified file name.

Figure 29: Field Transformer editor

- 6 In the **Target Field** drop list, choose a target field.
- 7 Modify or add parameter **Names** and **Values** for the transformer.
- 8 Repeat steps 5 through 7 for all Field Transformer nodes.

Configuring Item Transformers

To configure a global or Content Type-specific item transformer:

- 1 In the *Descriptor Setup dialog* (on page 19), click an Item Transformer folder in an Extractor Folder or click the Global Item Transformers folder.

The Editor pane displays the Item Transformer list. The available types of Item Transformers are listed in the **Available** box. The Item Transformers that you are already using, if any, are listed in the **Used** box.

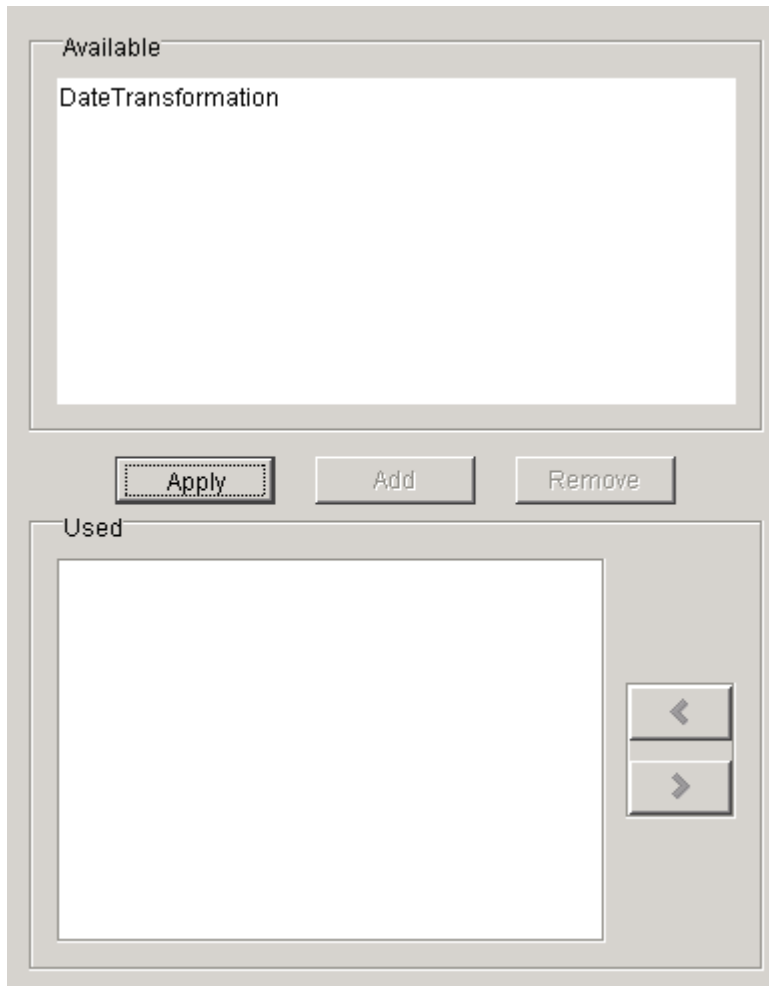


Figure 30: Item Transformer list

- 2 In the **Available** box, select each Item Transformer that you want to include and click [**Add**]. The Item Transformer moves to the **Used** box.
- 3 In the **Used** box, select each type of Item Transformer that you want to remove and click [**Remove**]. The Item Transformer moves to the **Available** box.
- 4 When you have finished selecting Item Transformers, click [**Apply**].

The Item Transformer folder includes a node for each Item Transformer. If you have removed any Item Transformers, they no longer appear under the Item Transformer folder.

- 5 Click on an Item Transformer node.

The Editor pane displays the Item Transformer editor. parameter **Name** and **Value** display default parameter names and values.

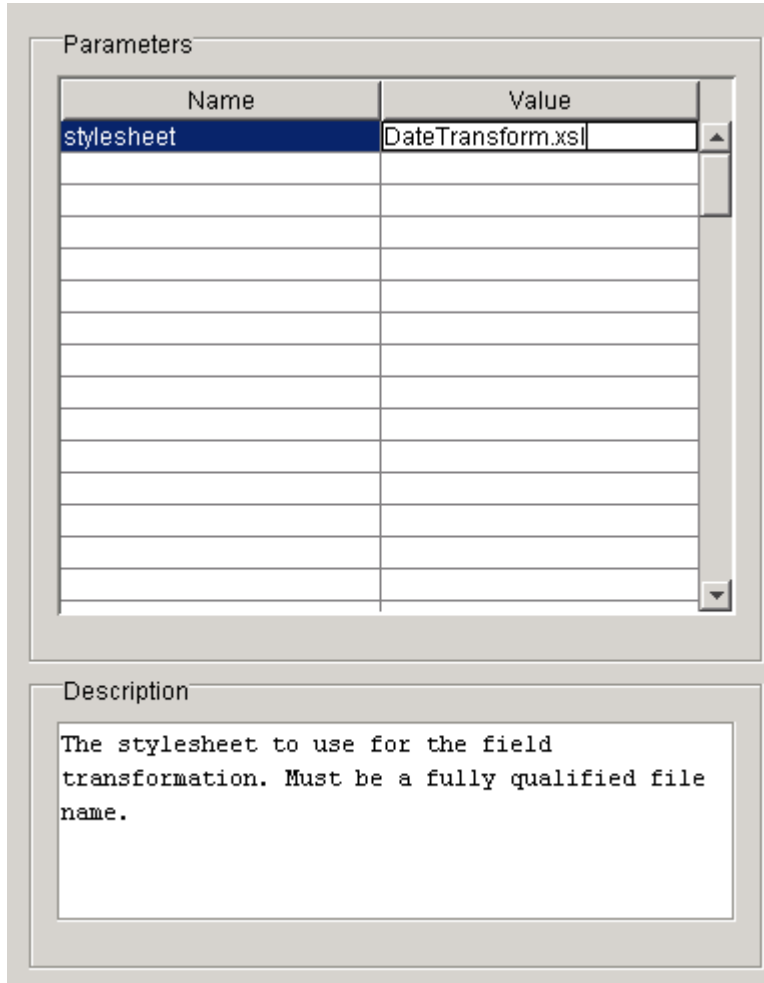


Figure 31: Item Transformer editor

- 6 Modify parameter **Names** and **Values** for the transformer.
- 7 Repeat steps 5 and 6 for all Item Transformer nodes.

Configuring Content Loaders

In a Descriptor, the Content Loader element specifies where items are uploaded and user credentials for accessing servers, if required. There are two default content loaders:

- Rhythmyx Loader - uploads items to the Rhythmyx server;
- Preview Loader - uploads content to a preview filesystem.

Choosing a Content Loader

To choose a content loader:

- 1 In the *Descriptor Setup dialog* (on page [19](#)), click the Content Loader folder in the Descriptor Tree.

The Editor pane displays unselected Content Loaders in the Content Loader List. If you have already selected a Content Loader, the Content Loader node in the Descriptor Tree includes a node for it and the Content List does not include it.

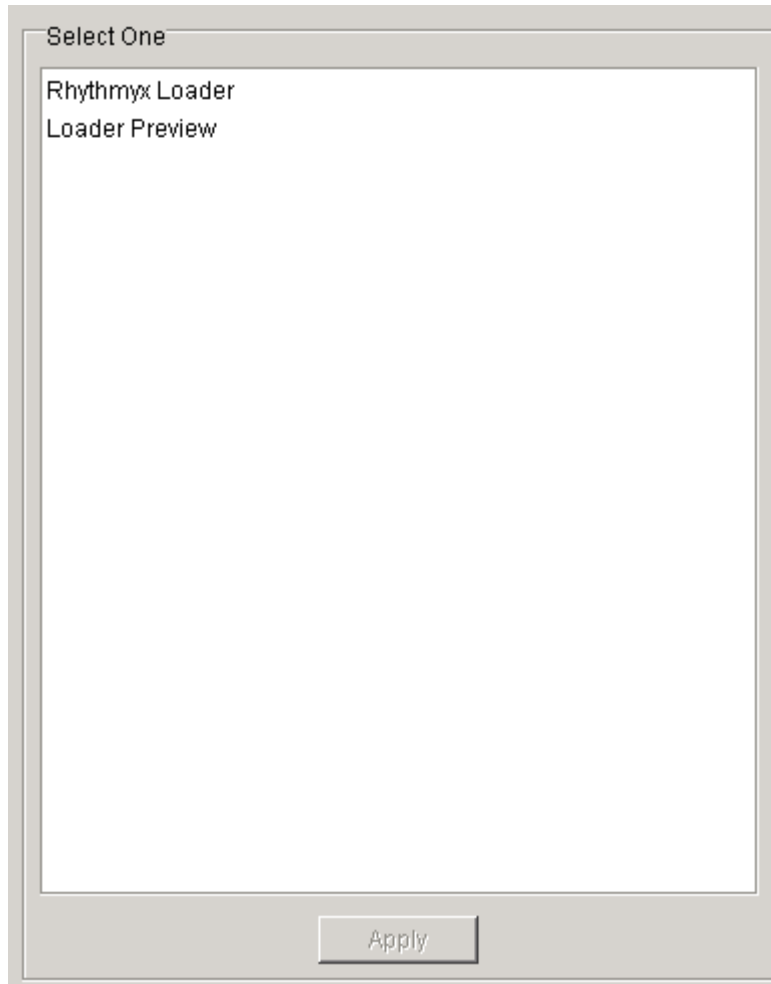


Figure 32: Content Loader list

- 2** Choose a Content Loader.
- 3** Click [**Apply**].
The dialog creates a node for the content loader under the Content Loader folder in the Descriptor Tree.
- 4** To change your choice, click a different content loader in the Content Loader List and click [**Apply**].
The originally selected Content Loader is now listed under **Select One**, and the Content Loader folder in the Descriptor Tree displays a node for the newly selected Content Loader.

Configuring the Rhythmyx Loader

When you open the Descriptor Setup dialog, The Editor pane displays fields for entering the **Rhythmyx Server**, **Rhythmyx Port**, and the **Username** and **Password** for accessing the server. No fields are available for editing when you click the Rhythmyx Loader node in the Content Tree.

Configuring the Preview Loader

To configure the Preview Loader:

- 1 In the *Descriptor Setup dialog* (on page 19), click the Preview Loader node in the Descriptor Tree.

The Editor pane displays the Preview Loader Editor:

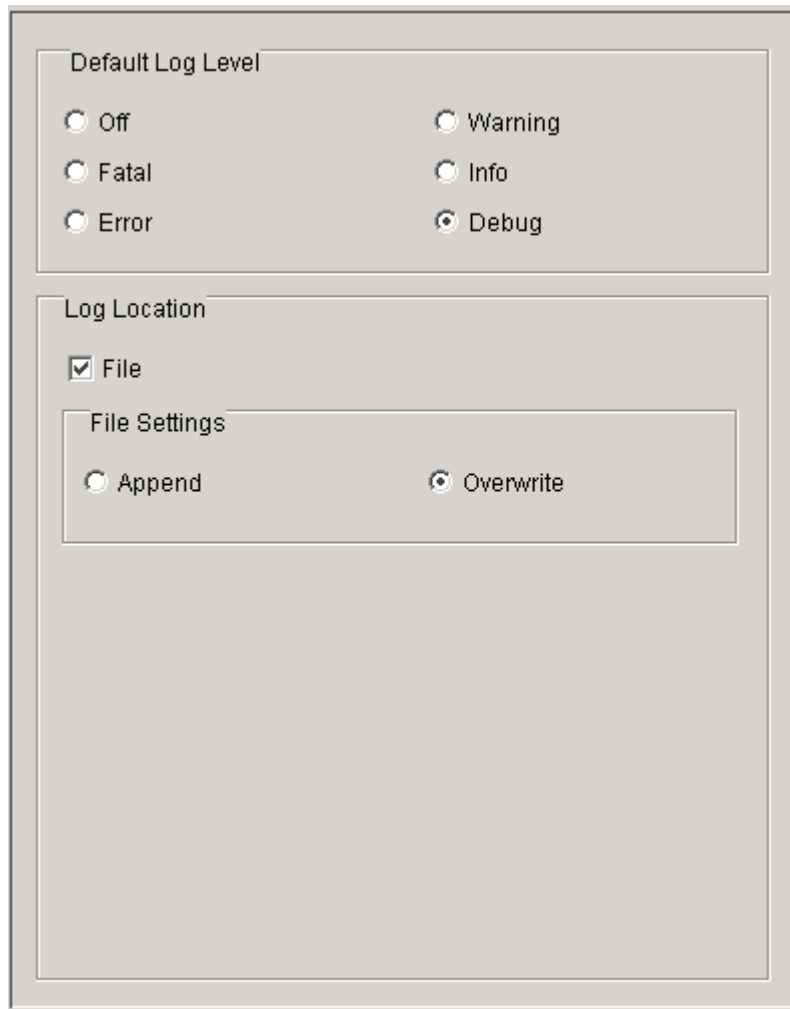


Figure 33: Preview Loader editor

- 2 In **Path**, enter the fully qualified path of the preview directory.

Configuring Logging

You can configure certain features of logging through the Descriptor Setup dialog:



The image shows a dialog box titled "Descriptor Setup" with two main sections. The first section, "Default Log Level", contains six radio buttons arranged in two columns: Off, Fatal, Error, Warning, Info, and Debug. The "Debug" option is selected. The second section, "Log Location", contains a checked checkbox for "File". Below this is a sub-section titled "File Settings" with two radio buttons: "Append" and "Overwrite", with "Overwrite" selected.

Figure 34: Logging options

- Default Log Level - Lowest severity level of information to log. For each scan and upload, the Enterprise Content Connector logs events of this severity level and higher. Options, in order of highest to lowest severity are:
 - *Fatal* - error events that cause Content Connector processing to abort;
 - *Error* - default; error events that do not cause Content Connector processing to abort;
 - *Warnings* - events that may be errors;
 - *Info* - general events that record the progress of Content Connector processing;
 - *Debug* - detailed events that are useful for debugging; and
 - *Off* - do not log any information.

- Log Location - location for printing log information;
 - *File* - Unchecked by default. Check to print log information to a file in the Descriptor path. If you uncheck File, all options under File Settings are disabled. NOTE: The Content Connector always prints log information to the Log View on the console.
- File Settings
 - *Append* - Unselected by default. Select to append all log entries to the current log file; deselect to create separate log files.
 - *Overwrite* - Selected by default. Select to overwrite the log file each time you re-upload.

NOTE: The Content Connector uses the Java Apache Project's Log4j package for logging. You may manually add additional features of Log4j in the Enterprise Content Connector *configuration file* (see "[Changing Configuration File Default Values](#)" on page 101) or in individual Descriptors for custom logging. See the documentation in for more information.

To configure logging:

- 1 In the *Descriptor Setup dialog* (on page 19), click the Logging node in the Descriptor Tree. The Editor pane displays the Log Editor.
- 2 Choose a **Default Log Level**.
- 3 If you want to print the log to a file, check **File** under **Log Location**. If you check **File**, the dialog displays **File Settings**.
- 4 Choose a **File Setting**.

Configuring Descriptor Error Handling

You can configure certain features of error handling through the Descriptor Setup dialog. The Enterprise Content Connector sends email notifications using SMTP protocol. In the *configuration file* (see "[Changing Configuration File Default Values](#)" on page 101), replace the default SMTP host with your system's SMTP host before creating Descriptors that send notifications.

To configure error handling:

- 1 In the *Descriptor Setup dialog* (on page 19), click the Error Handling node in the Descriptor Tree.

The Editor pane displays the Error Handling Editor:

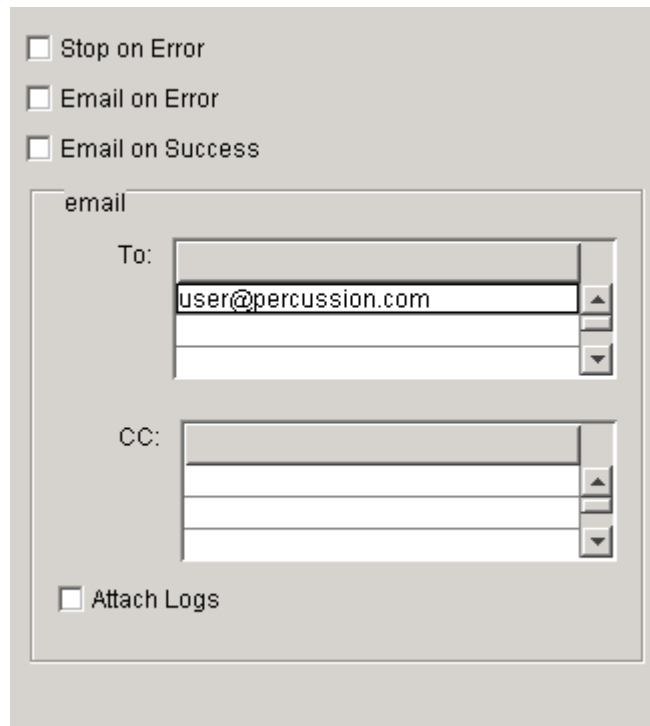


Figure 35: Error Handling editor

- 2 Check **Stop on Error** to stop running the Enterprise Content Connector when a Fatal or Error event occurs. Default is unchecked.
- 3 Check **Email on Error** to send an email to the addresses specified when an error occurs. Default is unchecked.
- 4 Check **Email on Success** to send an email to the addresses specified if an upload is successful. A successful upload results in no Fatal or Error events.
- 5 In **To**, enter one or more email addresses to receive error or success messages. If **Email on Error** or **Email on Success** is checked, this list must contain at least one email address. Default is none.

- 6 In **CC**, enter one or more email addresses to receive cc error or success messages. Default is none.
- 7 Check **Attach Logs** to attach the error log to each email. Default is unchecked.

Saving Descriptors

At any time that you are configuring a Descriptor in the Descriptor Setup dialog, save your settings to file by clicking [**Apply**]. To save your settings to file and close the dialog, click [**OK**].

When you click [**Apply**] or [**OK**], the Enterprise Content Connector validates all of your changes and displays messages if you have any errors.

If you are creating a new Descriptor, the Enterprise Content Connector saves it as an XML file with the Name you specified when you first opened the Descriptor Setup dialog. The directory for the Descriptor file has the same Name and is located in the Path that you specified. Status and log files for the Descriptor are stored in the same directory. When the Content Connector saves a new Descriptor, it also creates a logs subdirectory.

For example, in the following graphic, the Descriptor file is `descriptor1.xml` and it is stored in the directory `D:\Rhythmyx\contentconnector\descriptor1`. The log directory is `D:\Rhythmyx\contentconnector\descriptor1\logs`.

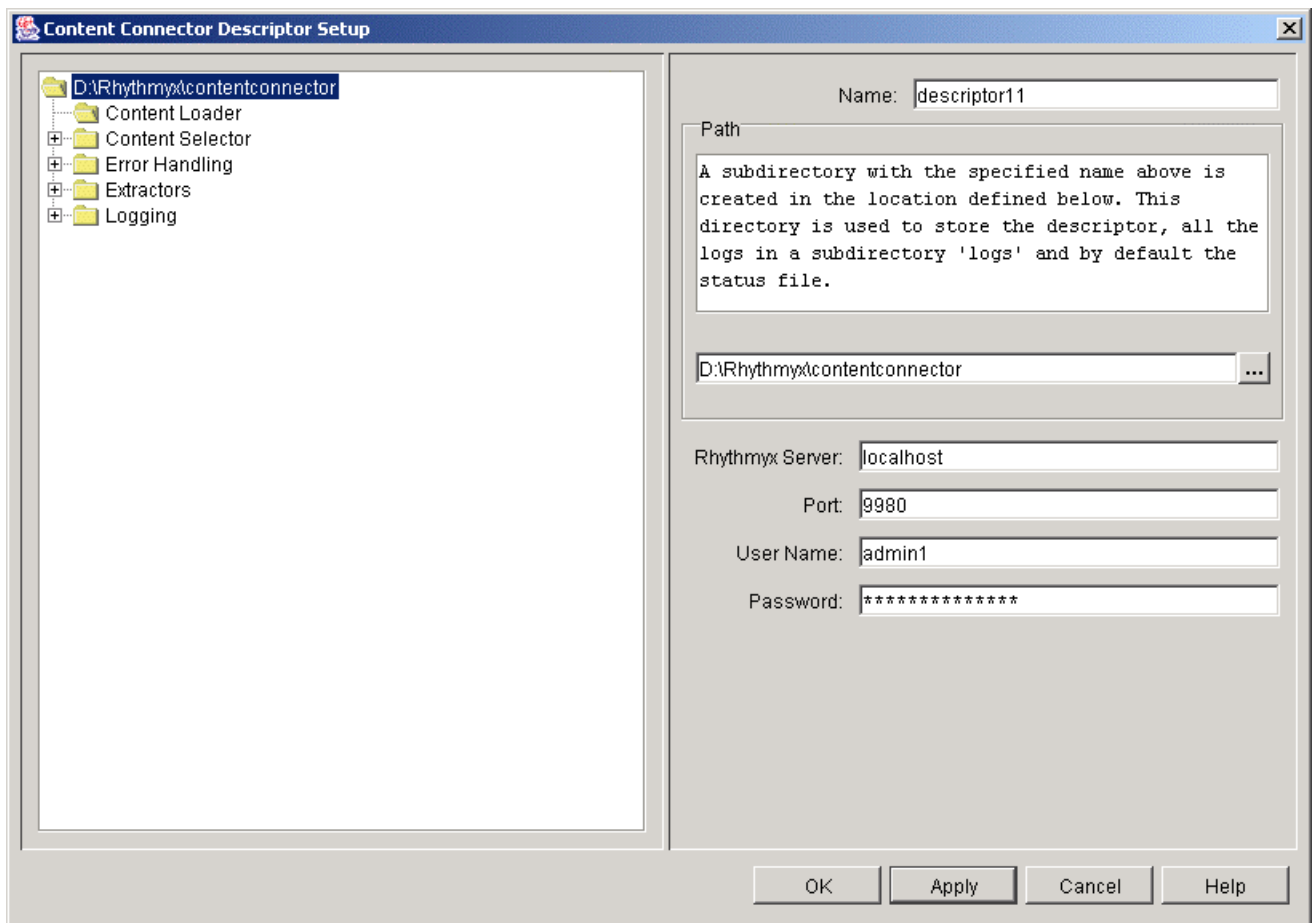


Figure 36: `descriptor1.xml`

Running the Rhythmyx Content Connector

You can run the Rhythmyx Content Connector interactively through the Content Connector Interface, or you can run it automatically using a batch file.

Running interactively with the Content Connector Interface gives you more control over the process. You can modify the Content Types assigned to files before you upload them, mark files for exclusion from uploading, view errors in each step as they occur, and choose to upload only portions of the scanned content. However, interactive running requires that you stand by; since uploading large amounts of data is time consuming, you should only run the Enterprise Content Connector interactively when you are uploading small amounts of data. Running with the Content Connector Interface is best for:

- migrating small Web sites with 200 pages or fewer;
- testing or debugging a Descriptor before you use it ;
- initially running the Descriptor for a scheduled batch run.

Running the Enterprise Content Connector with a batch file is useful for:

- migrating large Web sites with 200 pages or more;
- Running scheduled, incremental uploads of data stored in an external location.

Running The Enterprise Content Connector Interactively

To run the Enterprise Content Connector Interactively:

- 1 Open the Content Connector Interface.
- 2 Select *File > New* to create a new Descriptor or *File > Open* to load an existing Descriptor.
If you open a Descriptor that you used in a previous upload, the **Status file** (see "[Scanning for New and Changed Content](#)" on page 70) for the Descriptor automatically opens so the Enterprise Content Connector can determine if scanned files are New, Changed, or Unchanged
- 3 **Create and save the new Descriptor** (see "[Creating a Rhythmyx Enterprise Content Connector Descriptor](#)" on page 21) or **make any necessary modifications to the existing Descriptor**.
- 4 Select *Actions > Scan* to initiate item scanning.

During scanning:

- Status View indicates that Scanning Content is in progress and displays progress messages;
- Log View displays a Scanning tab and records errors for the task on the tab;
- the Content Selector finds specified files on the source site;
- the Extractors match with files.

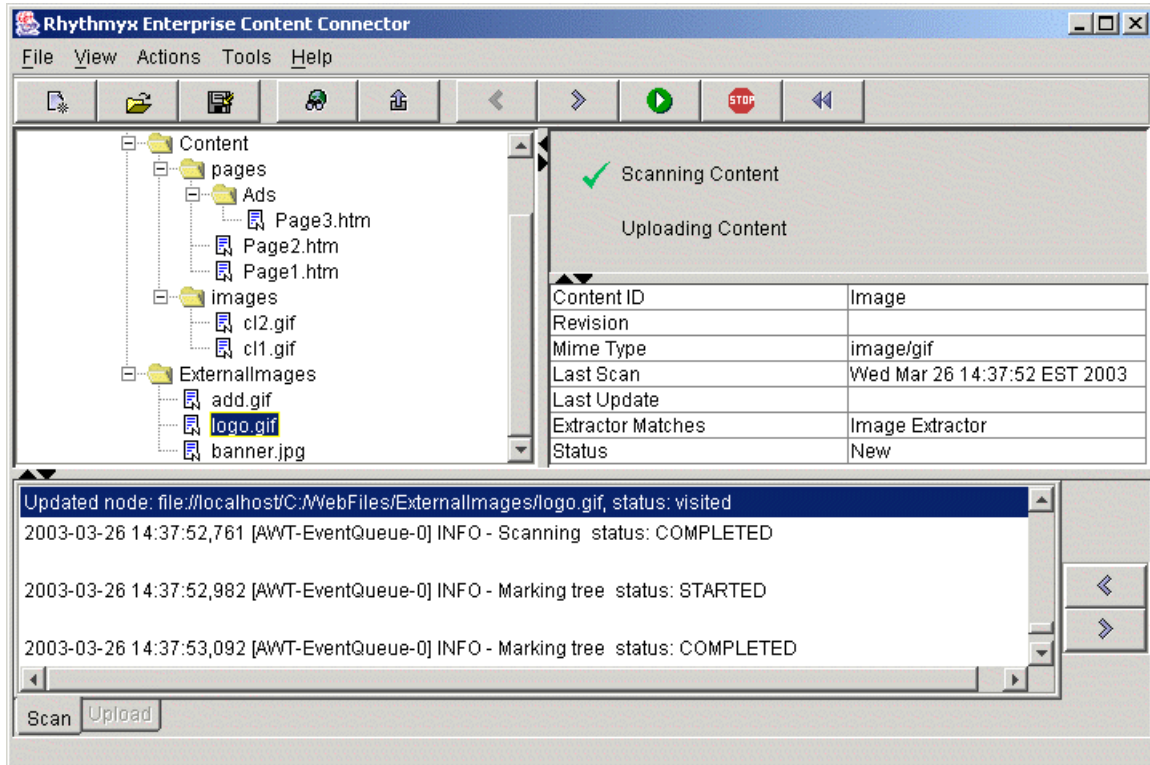



Figure 37: Scanning content

When scanning is complete:

- Scanning Content is checked in the Status View;
- the Content Tree displays a tree of the files scanned. The tree has the Source site directory structure. Extractors have assigned guessed Content Types to each file. If none of the Extractors matched with a file, it is included in the tree with the  icon, which indicates that it will not be **uploaded**. *If you are running* an incremental upload (see "[Running The Enterprise Content Connector Interactively](#)" on page 64), the **file icons** (see "[Content Tree](#)" on page 11) indicate whether previously scanned files are Changed or Unchanged.

5 *Manually review and modify the Content Tree* (see "[Modifying the Content Tree](#)" on page 67).

6 Select *Actions > Upload* to initiate item uploading.

During uploading:

- Status View displays an arrow indicating that Uploading Content is in progress and displays progress messages;
- Log View displays an Upload tab and records errors for the task on the tab;
- *The Enterprise Content Connector uploads New and Changed items associated with Item Extractors to the configured Rhythmyx server or a filesystem* (see ["Configuring The Rhythmyx Enterprise Content Connector"](#) on page 100).

If the Enterprise Content Connector uploads the items to Rhythmyx (with the Rhythmyx Loader):

- in Rhythmyx, New content items are added and given new Content Ids; changed content items replace existing content items and take the IDs of the items that they modify;
- Rhythmyx inserts metadata as it would for any new or modified content items; for example, it assigns the system date as the Modify Date or Create Date;
- Rhythmyx applies Workflow Actions specified in Extractors to content items; the acting user is the **Username** in the Descriptor;
- Rhythmyx checks in all uploaded (New and Changed) content items.

If the Enterprise Content Connector uploads the items to a filesystem (with the Preview Loader):

- prior to uploading, the Enterprise Content Connector gives all content items a Content Id of 0 and a Revision Id of 1;
- the Enterprise Content Connector uploads content items in the folder structure of the source site (also the folder structure of the Content Tree); by default, it uploads external items into the folder `RhythmyxServer`.

When uploading is complete:

- *Uploading Content* is checked in the Status View.

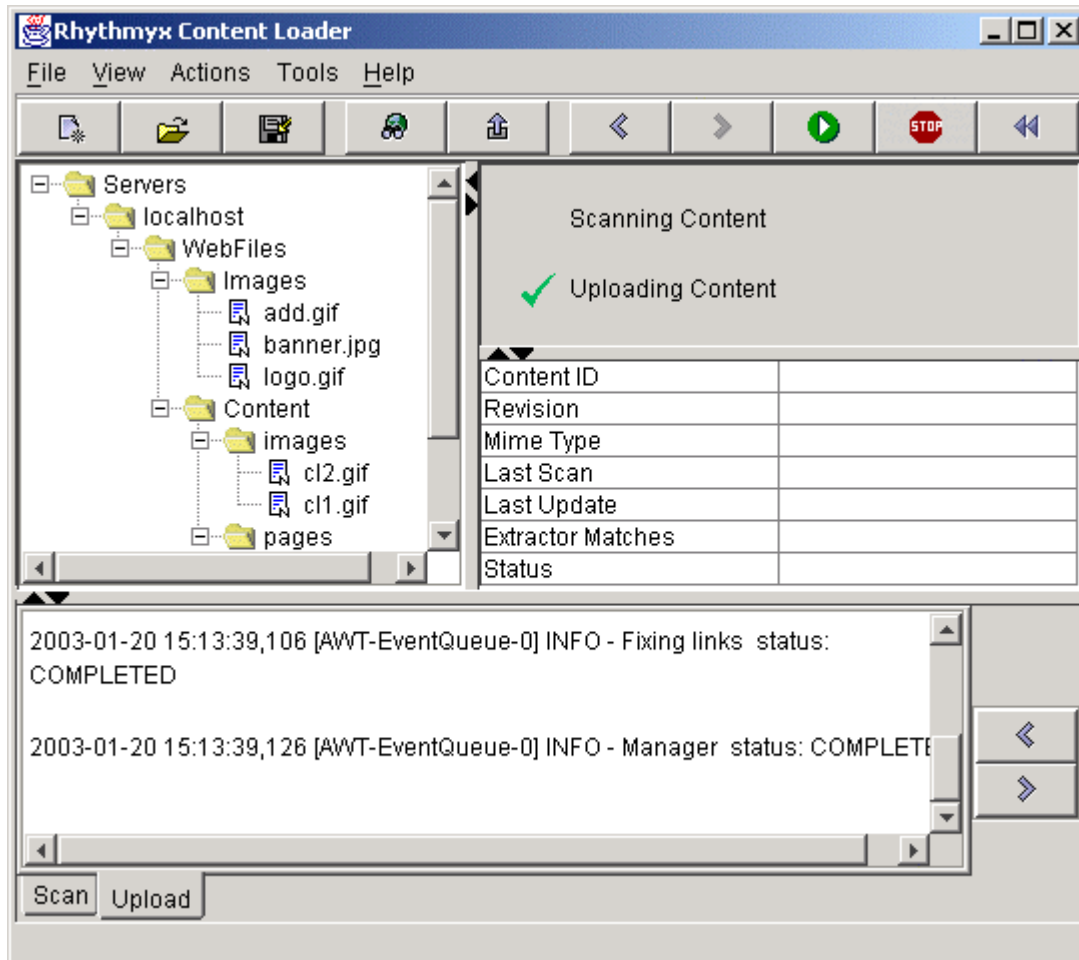


Figure 38: Content upload complete

NOTE: You can also **upload specific items** (see "Uploading Specific Items" on page 68) or **upload the contents of a folder** (see "Uploading the Contents of a Content Tree Folder" on page 69) in the Content Tree.

- 7 Review the tabs in Log View for errors. When you click an error entry, if a file in the Content Tree corresponds to it, it becomes highlighted. If necessary, correct errors **and redo the scan and/or the upload** (see "Scanning and Uploading Again to Correct Errors" on page 71).

Modifying the Content Tree

After the Content Selector in your Descriptor scans files and displays them in the Content Tree, you can change the Content Types it has assigned them and mark any files that you want to exclude from uploading.

Changing Content Types

To change file Content Types in the Content Tree:


- 1 Open the drop menu for the files that you want to change:
 - right-click a file (to change the Content Type of a single file);
 - select the individual files that you want to change, and right-click anywhere in the Content Tree window (to change the Content Type of multiple selected files);
 - right-click a folder (to change the Content Type of all files in the folder).
- 2 Choose *Assign* > <*Content Type*>, where <Content Type> is the content type that you want the contents of the file(s) to have.

The Content Connector changes the Content Type of the file or files.

Excluding Items

To mark a file in the Content Tree for exclusion:

- 1 Open the drop menu for the files that you want to change:
 - right-click a file (to exclude a single file);
 - select the individual files that you want to exclude and right-click anywhere in the Content Tree window (to exclude multiple selected files);
 - right-click a folder (to exclude all files in the folder).
- 2 Choose *Exclude*.

The Content Connector changes the files' icons to  and will not upload them.

Uploading Selected Items and Folder Contents

Uploading Specific Items

To upload a single item:

- 1 Complete steps 1 through 4 in *Running the Enterprise Content Connector Interactively* (on page 64).
- 2 Right-click the file that you want to upload, or select several files that you want to upload and right-click in the Content Tree area.
- 3 Choose *Upload*.

The Content Connector extracts the files, performs all transformations specified in the Descriptor, and uploads the items.

NOTE: To upload the item's dependents, you must perform a full upload with *Actions* > *Upload*.

Uploading the Contents of a Content Tree Folder

To upload the content of a Content Tree folder:

- 1 Complete steps 1 through 4 in *Running the Enterprise Content Connector Interactively* (on page 64).
- 2 Right-click the folder.
- 3 Choose *Upload All*.

The Content Connector extracts all files in the folder (except those marked for exclusion), performs all transformations specified in the Descriptor, and uploads the items.

NOTE: To upload the items' dependents, you must perform a full upload with *Actions > Upload*.

Scanning for New and Changed Content

If you run incremental uploads from the same external site, the Enterprise Content Connector can differentiate between New, Changed, and Unchanged (previously uploaded) files so you do not create duplicate items in Rhythmyx. For each upload, the Enterprise Content Connector creates a Status file that contains records of the content of each uploaded file. The Enterprise Content Connector compares the content of each file that it scans with the Status file record of the file's contents to determine if the file is New, Changed, or Unchanged.

The Enterprise Content Connector saves the Status file automatically after an upload. After scanning, you may save a status file if you want it to replace the current status file. If a Descriptor has a Status file associated with it, the Enterprise Content Connector automatically loads the Status file when you open the Descriptor and compares scanned files to its records during Scanning. The Status file for a Descriptor is always `[Descriptor Name].cls` and is stored in the Descriptor folder. When the Enterprise Content Connector saves a Status file after an upload, it overwrites the existing Status file.

Saving a Status File After Scanning

To save a Status file after scanning:

- 1 Run a scan..
- 2 Select *File > Save*.



The Status file is saved as `[Descriptor Name].cls` in the Descriptor folder.

Scanning and Uploading Again to Correct Errors

After you have scanned and uploaded, you can correct errors and scan or upload the same files again.

To scan again after uploading:

- 1 Leave the Descriptor open in the Content Connector Interface.

- 2 In the Toolbar, click  or .

OR

In the Menu Bar, choose *Action > Scan*.

To upload again after uploading:

- 1 If the previous upload action successfully uploaded any files, click *File > Save* to save the Status file and avoid reuploading the same files.

- 2 In the Toolbar, click .


OR

In the Menu Bar, select *Action > Upload*.

To scan again after scanning:

- 1 Click  to reset the Descriptor as unscanned.

The Content Tree view clears.

- 2 Click  or select *Actions > Scan*.

Running The Rhythmyx Enterprise Content Connector Automatically

Rhythmyx supplies the executable file `RhythmyxContentConnector.exe` for scanning and uploading content automatically. Run `RhythmyxContentConnector.exe` on a machine that has access to the content to be uploaded and the Rhythmyx server. The batch command scans and uploads as the Content Connector Interface does, but it does not give you the option of modifying the Content Tree. The Enterprise Content Connector automatically loads the Descriptor's Status file, if it exists, when it runs `RhythmyxContentConnector.exe`, and automatically saves the Status file when `RhythmyxContentConnector.exe` finishes uploading content.

The Windows command format is:

```
c:\>RhythmyxContentConnector [-descriptor <Descriptor path>] [-server <name>] [-port <#>] [-noui] [-d] [-p <path>]
```

The Unix command format is:

```
c:\>RhythmyxContentConnector.bin [-descriptor <Descriptor path>] [-server <name>] [-port <#>] [-noui] [-d] [-p <path>]
```

where:

`<Descriptor>` is the location of an existing Descriptor (include only the relative directory, not the Descriptor file). Required. If a Status file exists for the Descriptor, the Enterprise Content Connector automatically opens the Status file when it runs the `RhythmyxContentConnector.exe`.

`<name>` is the name of the target Rhythmyx server. Optional. If supplied, it overrides the Rhythmyx server specified in the Descriptor.

`<#>` is the Rhythmyx port. Optional. If supplied, it overrides the port specified in the Descriptor.

`-noui` turns off interactive mode. By default, interactive mode is turned on.

`-d` turns on Debug level of error logging. Debug level is the highest level of error logging and includes all events that may be useful for debugging. Turned off by default.

`-p` overrides the Content Loader specified in the Descriptor with the Preview Loader. Turned off by default.

For example:

```
c:\>RhythmyxContentConnector -server localhost -port 9992 -descriptor C:\contentconnector
```

For details about processing, see *Running the Enterprise Content Connector Interactively* (on page 64).

CHAPTER 5

Examples of Rhythmyx Enterprise Content Connector Operations

The examples provided in this section illustrate the typical usage scenarios for Site Migration and Content Syndication. For the Site Migration, the usual scenario is a one time migration of a Web site into Rhythmyx. For Content Syndication, the usual scenario is an incremental upload from a site that stores content. In the one time upload, the user must first create the Descriptor, configuring the Content Selector, Extractors, Transformers, and Content Loader, then review the Content Tree, make any necessary corrections, and upload the content. The user completes this procedure interactively, using the Content Connector Interface. In the incremental upload, the user initially completes the procedure described for the one time upload, and saves the Status file. After the initial upload, the user performs incremental uploads by batch file.

NOTE: The Rhythmyx Workflows and Content Types used in this example are stored in the archive <Rhythmyx root>/Samples/DeprecatedSamples.pda. For instructions on installing this archive, see the *Installing Rhythmyx* document.

Implementing Site Migration with Enterprise Content Connector: One Time Migration Example

Migration Example: Introduction

The One Time Migration example shows you how to use Enterprise Content Connector to upload the content from a Web site.

In this example, the user has just installed Rhythmyx and wants to install the contents of its current Web site, which is not currently handled by Rhythmyx, into the Rhythmyx server directory. In the future the Web site files will be managed by Rhythmyx.

The information required for uploading is as follows:

Rhythmyx server - the local server

Rhythmyx port - 9992

Rhythmyx Username - Uploader1 (see *Setting Up Rhythmyx Roles and Workflows* (see "[Migration Example: Setting up Rhythmyx Roles and Workflows](#)" on page 75))

Rhythmyx password - demo

Source site - C:/Webfiles

Types of items stored in Webfiles:

- static images used on the Web site; stored in the sub-folder Images;
- HTML pages; stored in the sub-folder content/pages;
- image content; stored in the sub-folder content/images.

Format of items in the source site:

- all image files have .gif or .jpg extensions;
- HTML files;
- Content Types extracted and their default Workflows:
 - Pages- Article Workflow
 - Images - Images Workflow

File type in source	Content Type in Rhythmyx	Sample file name	Sample file contents
HTML file	Page	Page2.htm	<pre> <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"> <html> <head> <title>Lorem ipsum </title> </head> <body> <table border="0" width="100%"> <tr> <td>Example I Lorem ipsum dolor sit amet, consectetuer adipiscing elit </td> </tr> <tr> <td>LiEuropean lingues es membres del sam familie. </td> </tr> </table> </body> </html> </pre>
Image file	Image	cl1.gif	
Image file	None. Becomes an image Web resource.	add.gif	

Migration Example: Setting up Rhythmyx Roles and Workflows

Before uploading:

- 1 Create a Rhythmyx Uploader role with the access required to upload content. Add a new user, Uploader1, to the role and give the user access to the source site and your Rhythmyx server. Do not assign Uploader1 to any other roles.

In the Rhythmyx Online Help see the section Security in Rhythmyx/Maintaining Roles for information about creating roles and users.

- 2 Since much of the uploaded content was already edited and approved on the source site, create transitions in the existing Workflows that allow the Uploader to transition uploaded content from a draft state to a Publish state. The draft state is the Workflow State that Uploaded content will have when it enters Rhythmyx. In the Article Workflow the draft state is Draft; in the Image Workflow the draft state is Artwork.
- 3 Assign the new Publish Transitions to the Uploader role only. After uploading new content items, the Uploader can review them and transition them directly to Public.

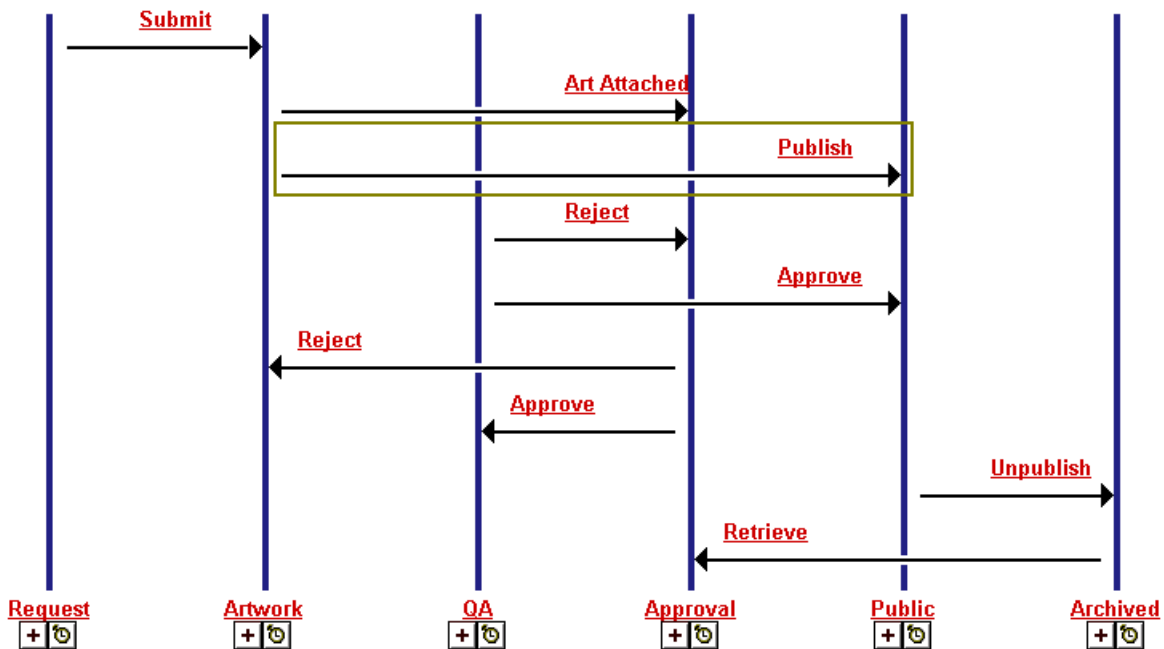


Image Workflow with Publish transition added to Artwork State.

Migration Example: Creating a Descriptor

You may create the Descriptor for a one time upload by modifying an existing Descriptor or creating a new one. In this example, you create a new Descriptor.

To create the Descriptor:

- 1 Open the Content Connector Interface.
- 2 Choose *File > New* to open the Descriptor Setup dialog for a new Descriptor.

The Descriptor Tree displays a folder for each configurable element of the Descriptor.

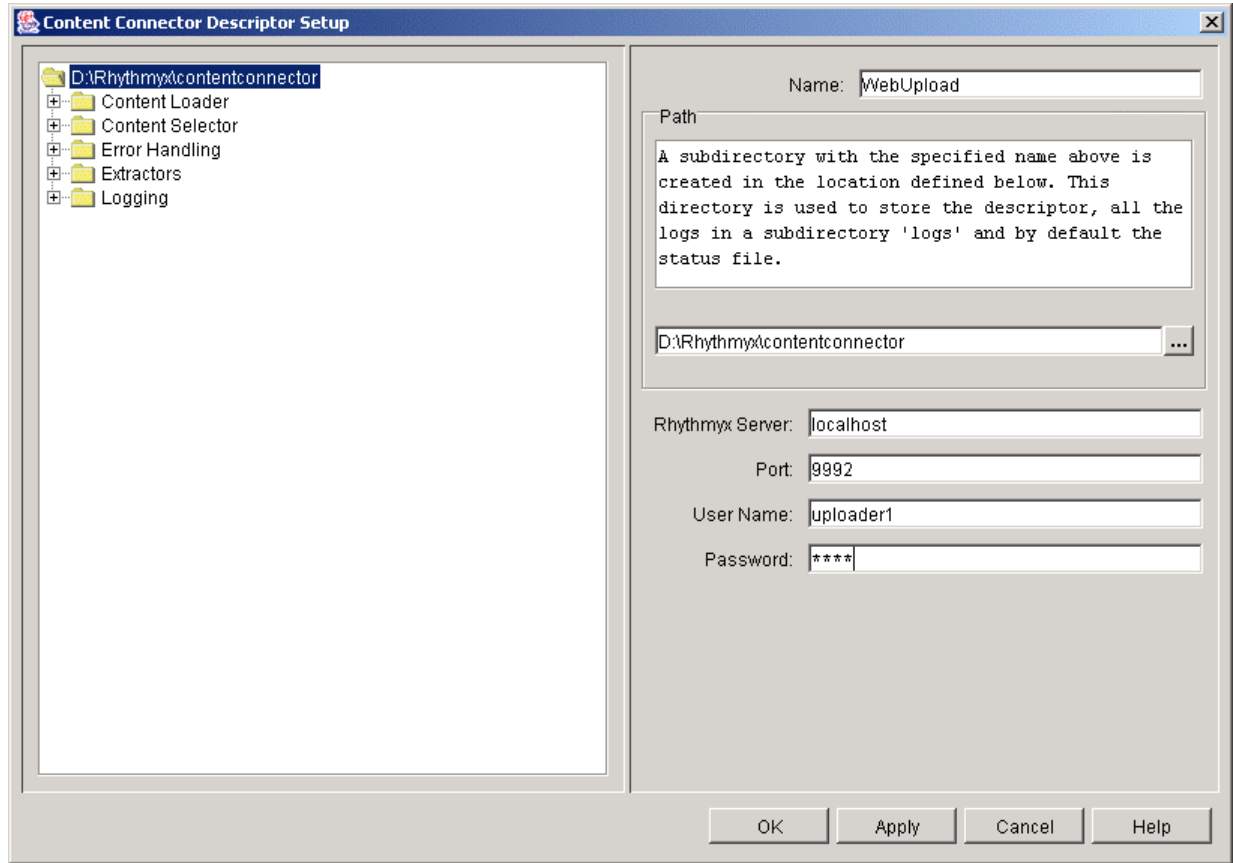


Figure 39: Preparing to create a new descriptor

- 3 Fill in the fields of the first Descriptor Setup window:
 - In **Name**, enter *WebUpload* as the Descriptor name.
 - In **Path** use the default path, in this case, *D:/Rhythmyx/contentconnector*. The Enterprise Content Connector will create a folder named *WebUpload* in the path to store the Descriptor file (*WebUpload.xml*)
 - In User Name, enter *Uploader1*.
- 4 **Choose the Content Loader** (see "[Choosing a Content Loader](#)" on page 54).
 - a) Click the Content Loader folder in the Descriptor Tree.

The Editor pane displays the Content Loader List.

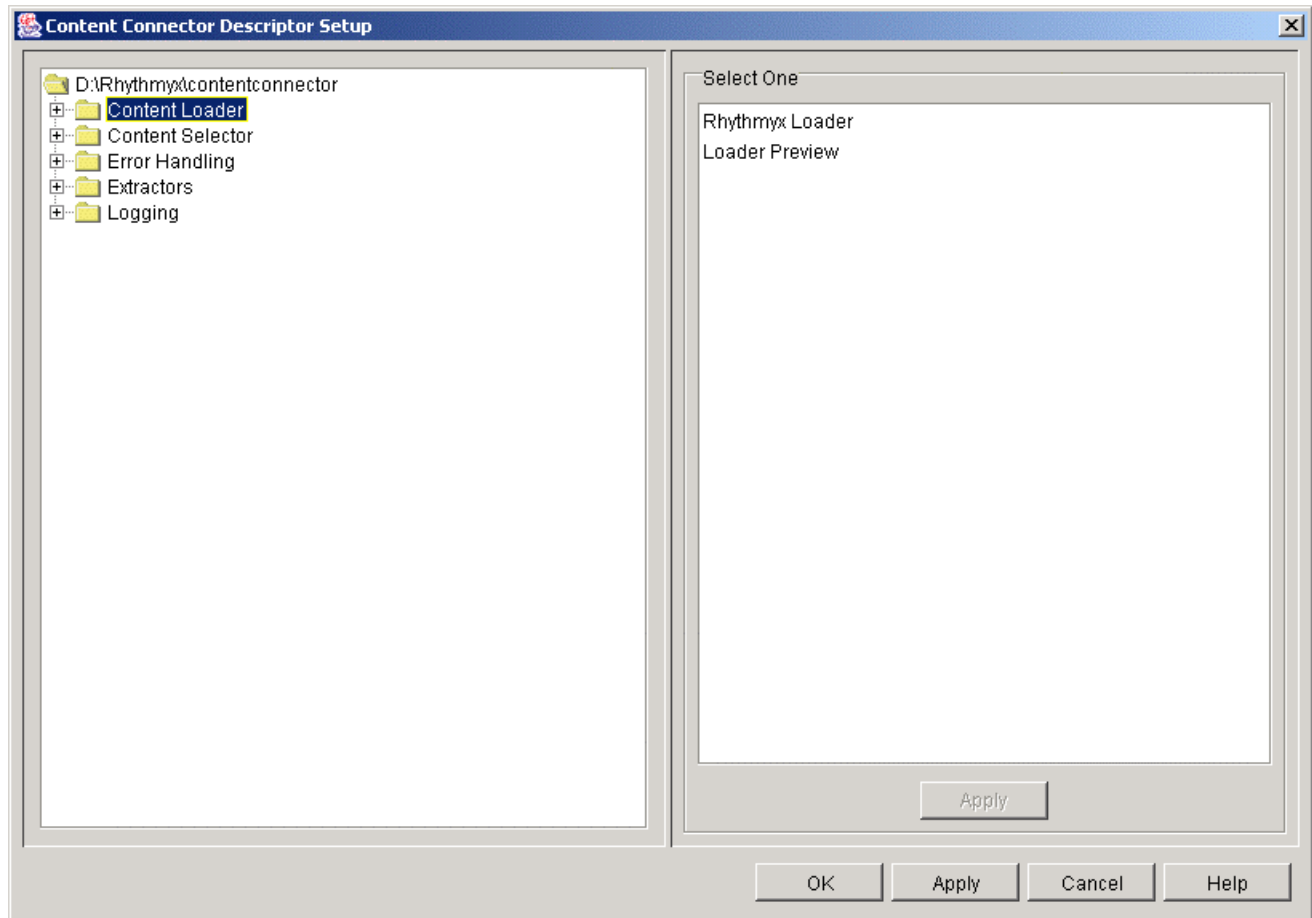


Figure 40: Choosing the descriptor's Content Loader

- b) Since you want to upload the Web content into your Rhythmyx directory, choose Rhythmyx Loader and click [**Apply**].

The Content Loader folder now contains a Rhythmyx Loader node.

- 5** Choose and configure a Content Selector.
 - a) Click the Content Selector folder in the Descriptor Tree.

The Editor pane displays the Content Selector List.

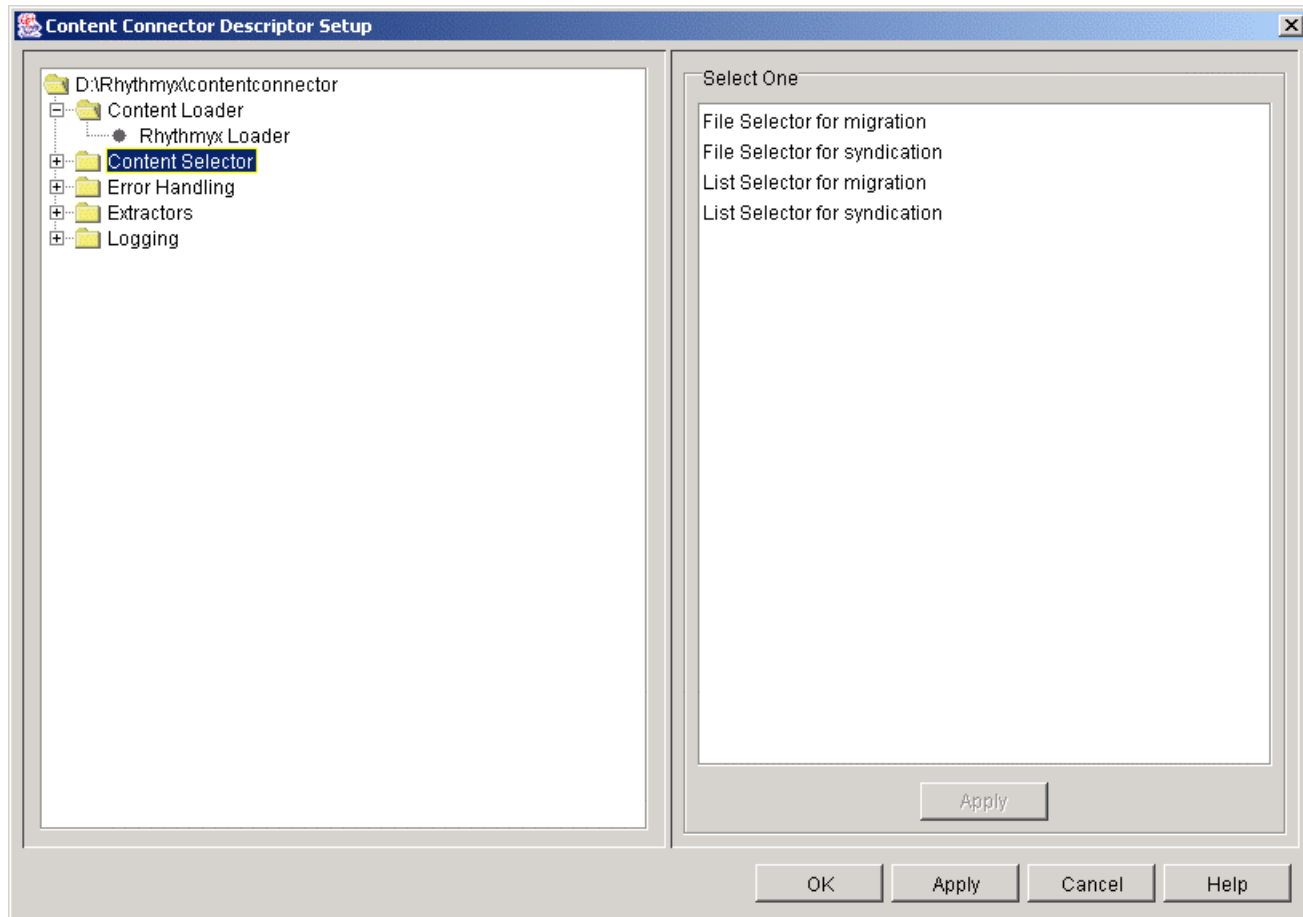
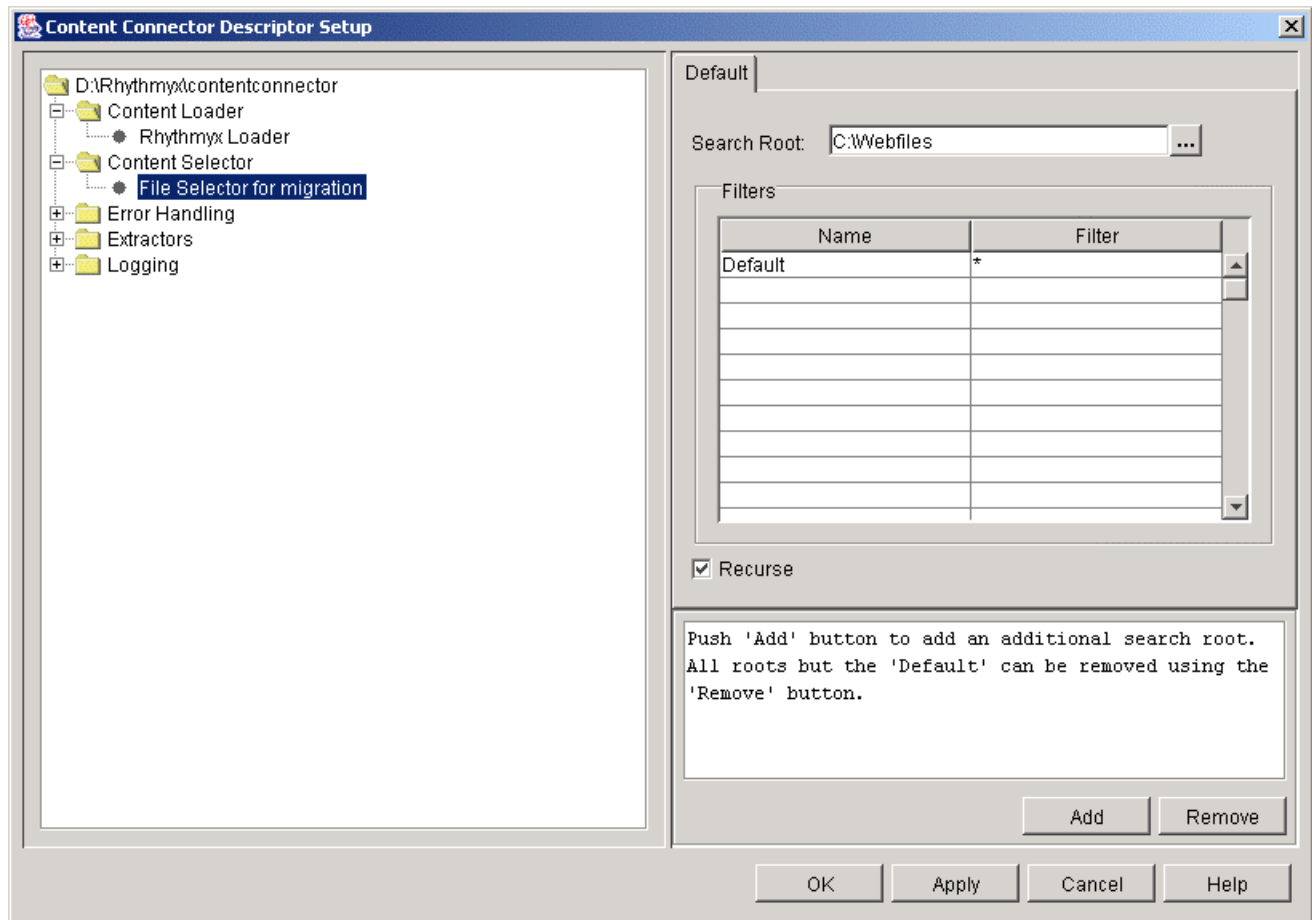


Figure 41: Choosing a Content Selector

- b) In the Content Selector List, choose *File Selector for migration* and click [**Apply**].
The Descriptor tree displays a node for the File Selector below the Content Selector Folder.
- c) Click the File Selector for migration node.
The Editor pane displays the File Selector Editor.



r.

- f) In the *File Selector Editor*, in **Search Root**, enter the location where the source content is currently stored, *c:/WebFiles*. Leave the default filter as * to scan all files. Check **Recurse** to scan subfolders in the search root.

- 6 Use the default error handling settings.
 - do not stop processing on errors of any level;
 - do not send notifications on errors or success.

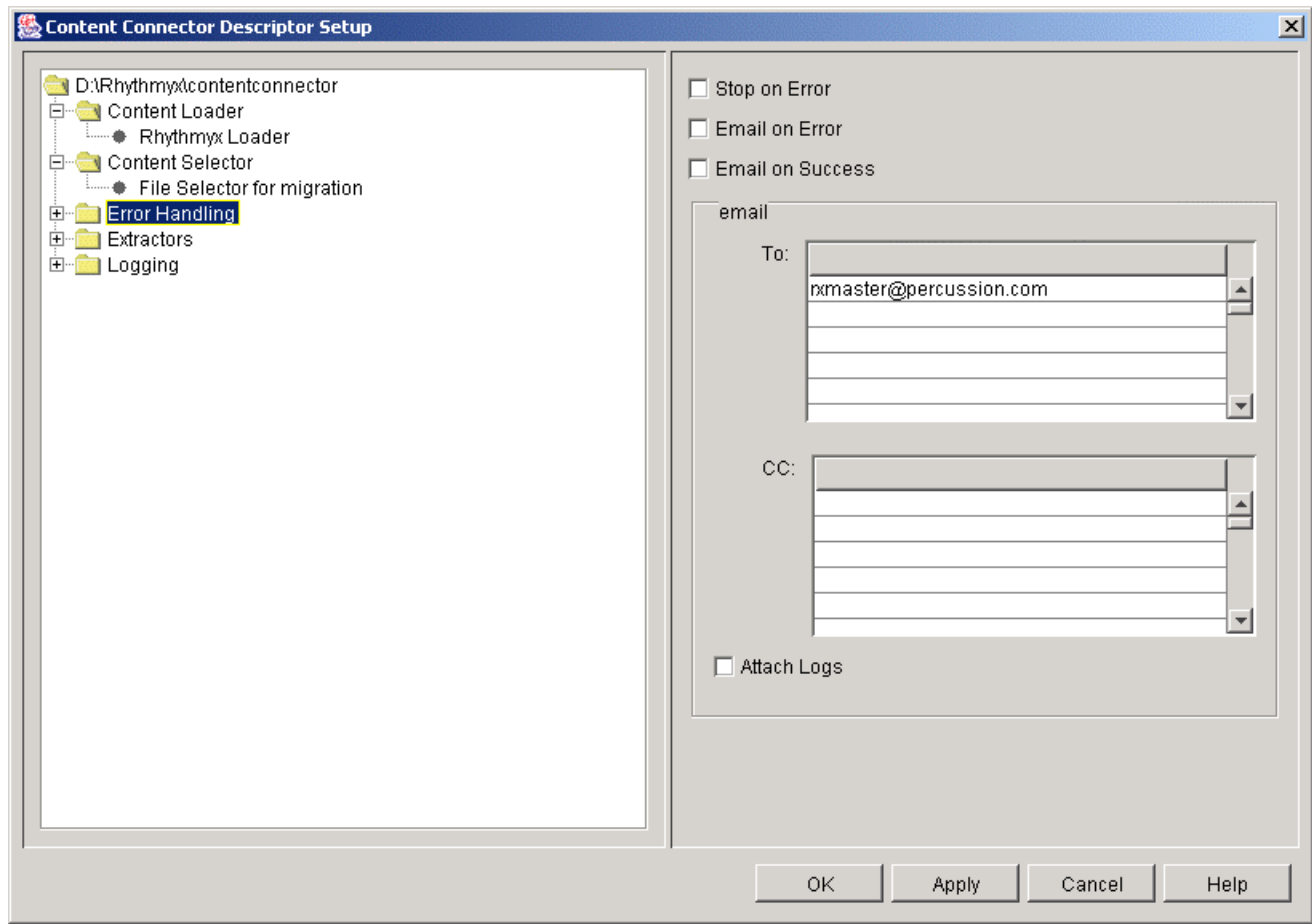


Figure 42: Configuring error handling

- 7 Choose and configure Static Item Extractors.
 - a) In the Descriptor tree, expand the Extractors folder and click Static Items.

The Editor pane displays the **Available** and **Used** Static Item Extractors.

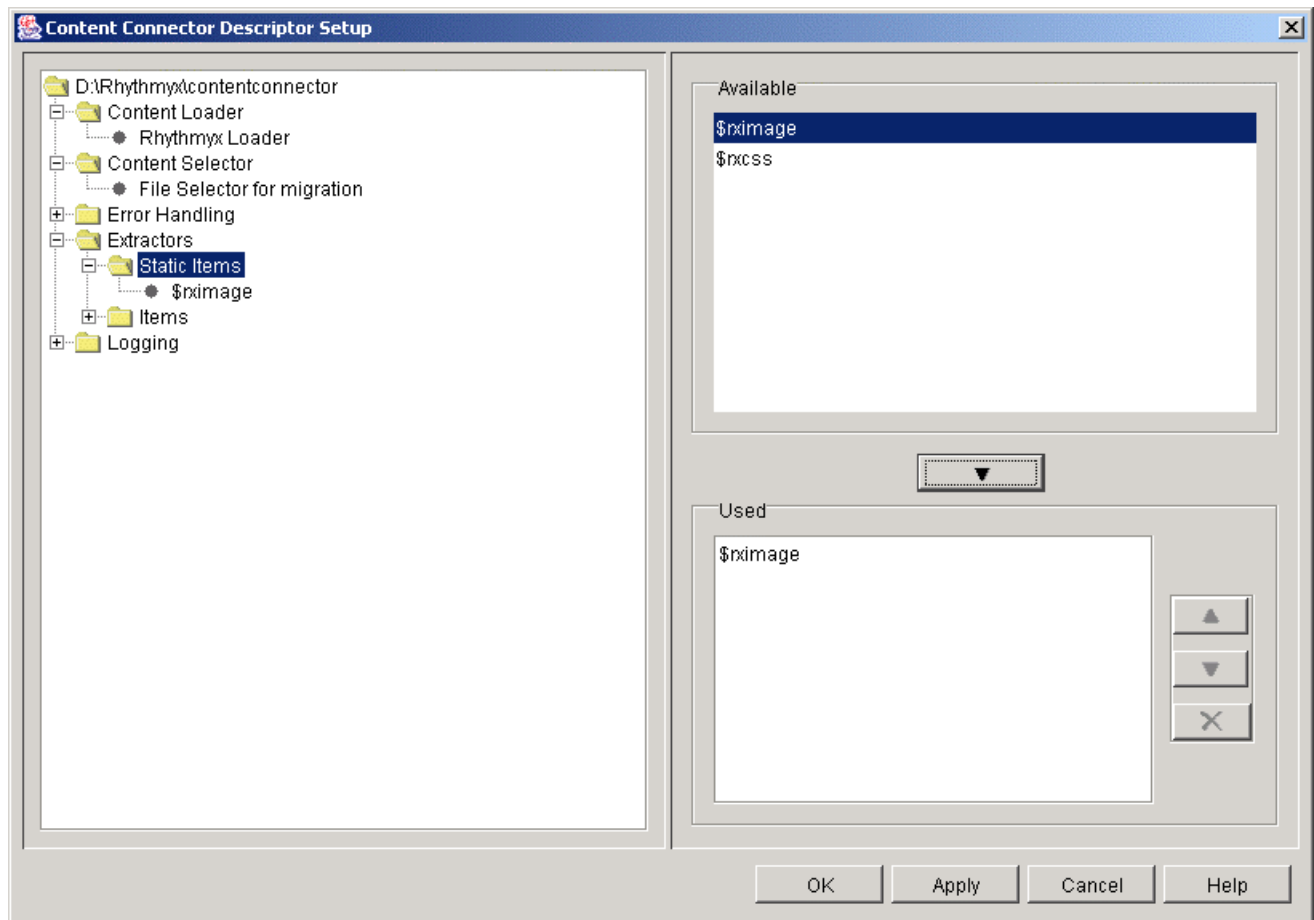


Figure 43: Choosing Static Item Extractors

- b) Since the *Webfiles* folder includes static images that are used on the Web site, add \$rximages.

The Used box and the Static Items folder displays \$rximage.

- c) Click the \$rximage node.

The Editor pane displays the Static Item Extractor Editor.

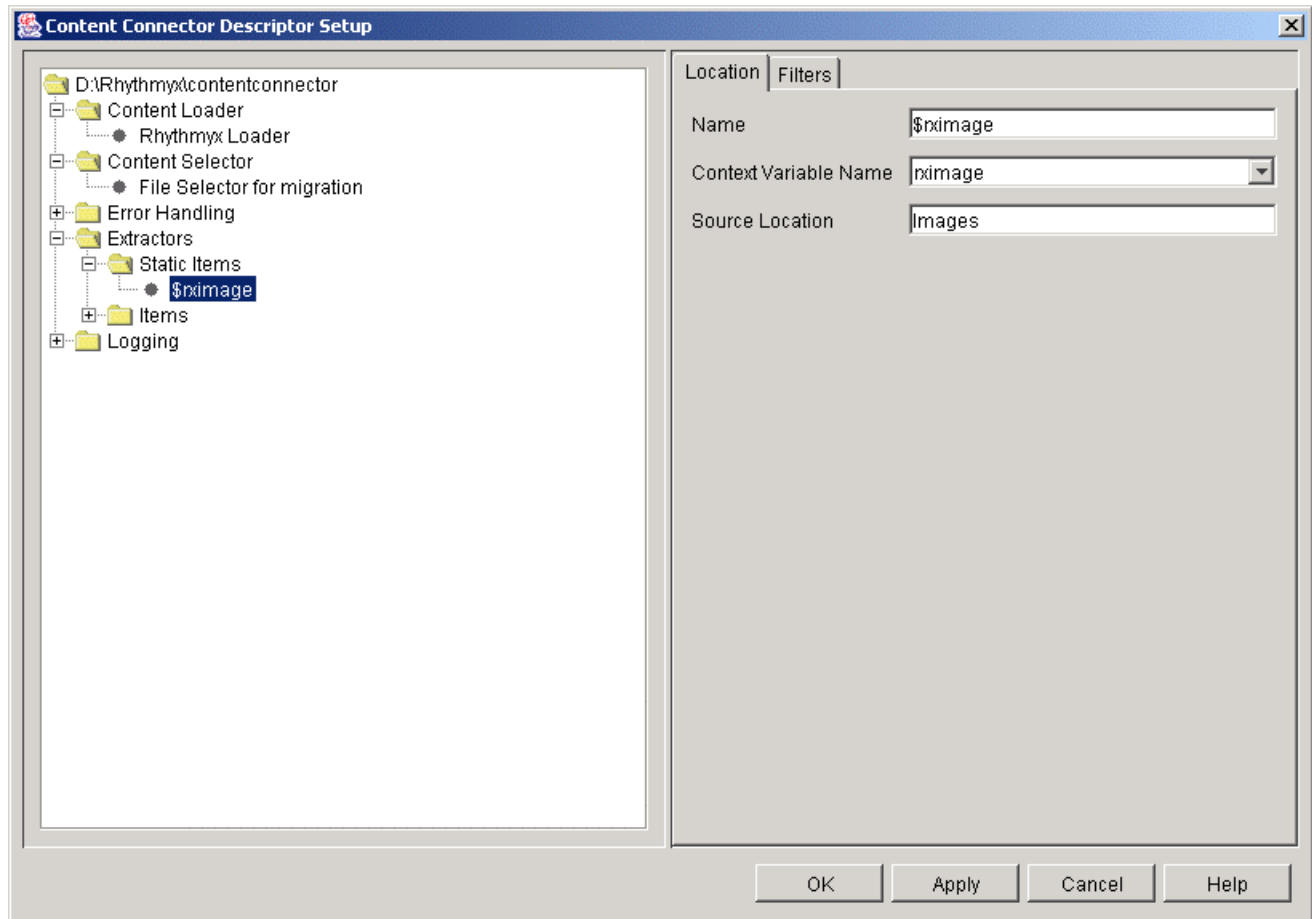
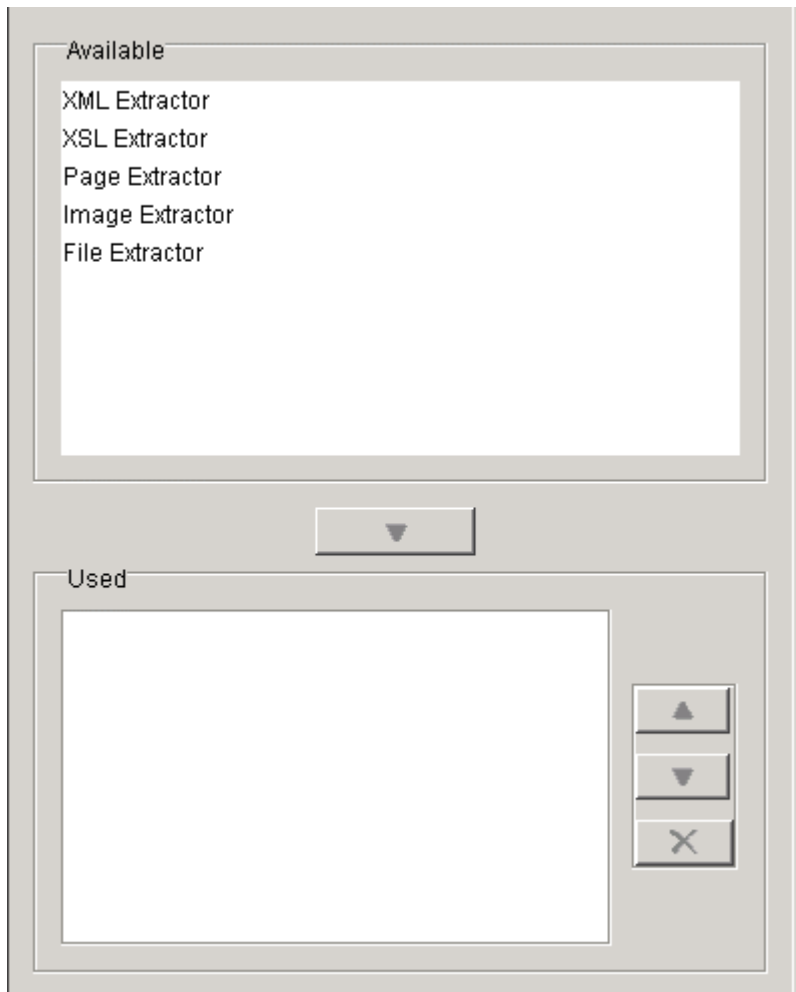


Figure 44: Editing the \$rximage Extractor

- d) In the Location tab leave the Name and Context Variable Name fields at their default settings. . Since Web images are stored in *c:/WebFiles/Images*, enter *Images* in *Source Location*. The filenames are case-sensitive, so you must change the default from images.
 - e) It is not necessary to modify information on the Filters tab; you want to extract all files in the source images folder.
- 8** Choose and configure Item Extractors.
- a) In the Extractors folder, click Items.

The Editor pane displays Available and Used Item Extractors.



- b) Since the content files in the current Web site are either image or HTML files, in the Items List, add Page Extractor, which extracts HTML files, and Image Extractor, which extracts image files and click [**Apply**].

The Items folder displays nodes for Page Extractor and Image Extractor. Both folders have nodes for Actions.



Figure 45: Page and Image Extractors in the Items folder

9 Configure Page Extractors.

- a) Click the Page Extractor node.

The Editor pane displays an Item Extractor Editor for the Page Extractor.

Field	Value Type	Value
sys_title	Variable	\$filename
sys_suffix	Variable	\$fileextension

Figure 46: Editing the Page Extractor

- b) Leave the default information on all of the tabs. You do not have to enter any Transitions on the Workflow tab because you want each Page Content Item to enter Rhythmyx in the first state for the Article Workflow (draft).
- 10** Configure the Image Extractor.
- a) Click the Image Extractor node.

The Editor pane displays an Item Extractor Editor for the Image Extractor.

The screenshot shows the 'Item Extractor Editor' for an 'Image Extractor'. At the top, there are four tabs: 'Content Type', 'Fields', 'Filters', and 'Workflow'. Below the tabs are three input fields: 'Name' with the value 'Image Extractor', 'Community' with a dropdown menu showing 'Default', and 'Content Type' with a dropdown menu showing 'Image'. Below these fields is a section titled 'System field-value map' which contains a table with three columns: 'Field', 'Value Type', and 'Value'. The table has one row with the following data: Field: 'sys_title', Value Type: 'Variable', Value: '\$filename'. The table has 14 rows in total, with the first row containing data and the rest being empty.

Field	Value Type	Value
sys_title	Variable	\$filename

Figure 47: Editing the Image Extractor

- b) Leave the default information on the Content Type, Fields, and Filters tabs. The image files in your source site have .gif and .jpg extensions. Leave the default filters since they locate files with these extensions.

c) Click the Workflow tab.

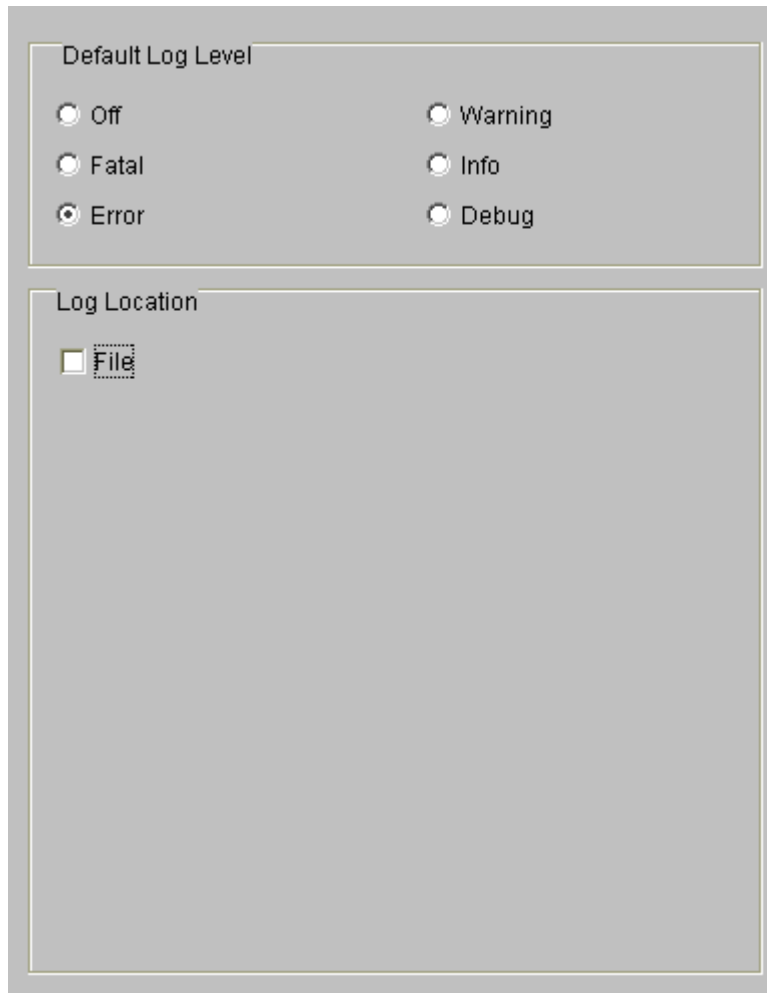
The screenshot shows a configuration window with four tabs: 'Content Type', 'Fields', 'Filters', and 'Workflow'. The 'Workflow' tab is active. Below the tabs are two dropdown menus: 'Workflow' set to 'Image' and 'Transition performed on' set to 'Insert new item'. Below these is a table titled 'Transitions' with a scrollable area. The first row of the table is highlighted in yellow and contains the text 'Submit (Submit)'. The table has approximately 15 rows in total.

Figure 48: Configuring the Workflow Transition

- d) Change Workflow to Image and Transition performed on as Insert new item.
- e) Since you want Image content items to enter Rhythmyx in the second Workflow State, Artwork, enter Submit in the first line of the Workflow Actions Editor.

11 Click the Logging folder and Configure the log settings:

- Log information levels of Error and higher.
- Send the log file to the console only.



The screenshot shows a configuration window with two sections. The top section, titled "Default Log Level", contains six radio button options arranged in two columns. The first column has "Off", "Fatal", and "Error" (which is selected). The second column has "Warning", "Info", and "Debug". The bottom section, titled "Log Location", contains a single checkbox labeled "File" which is currently unchecked.

Figure 49: Configuring logging settings

- 12 You have completed your Descriptor. Click [**Apply**] and review the Descriptor Tree.

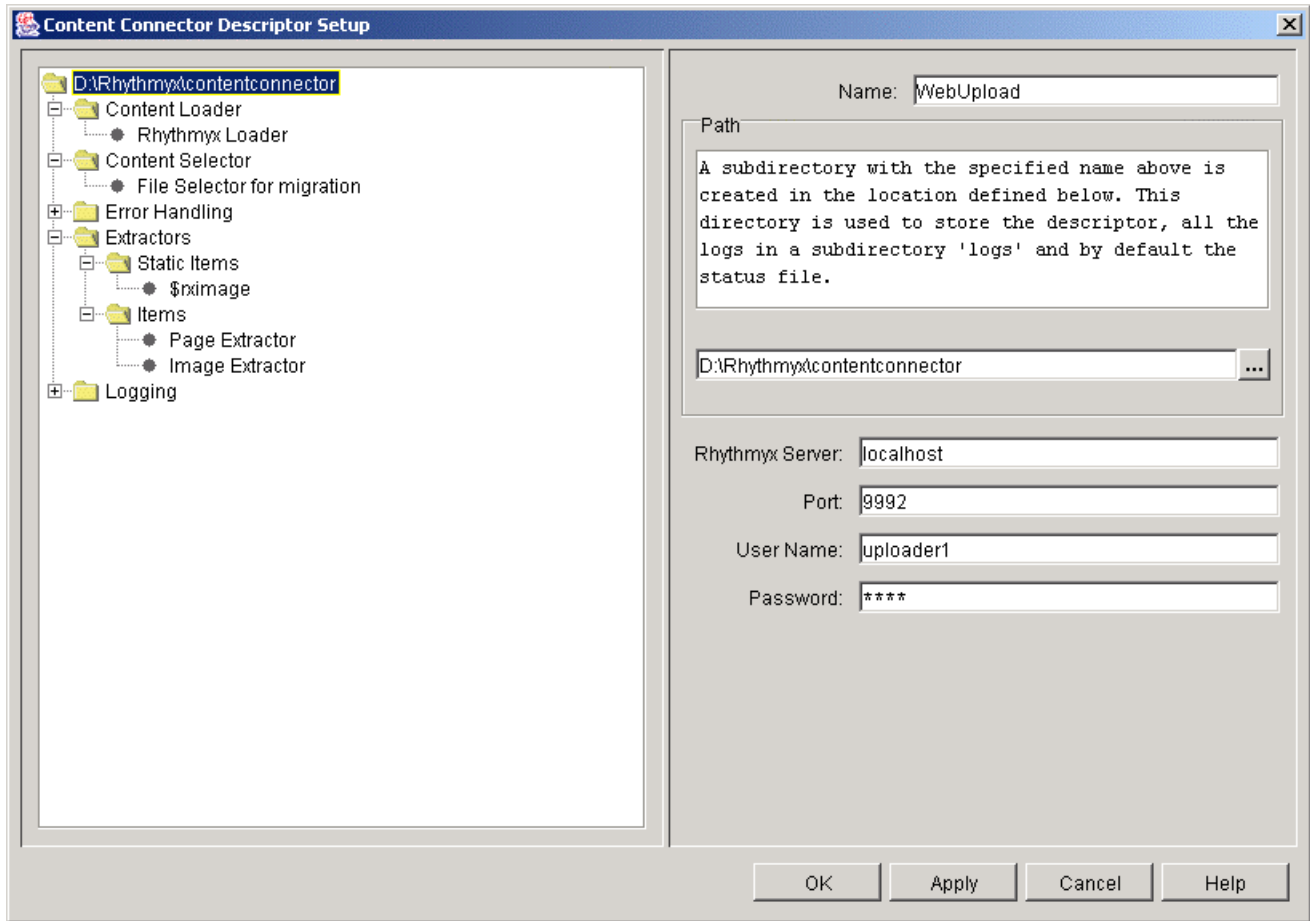


Figure 50: Reviewing the Descriptor Tree

- 13 Make any necessary changes to the Descriptor configuration and click [**OK**] to save it. The Descriptor Setup dialog closes. The WebUpload Descriptor file remains open for processing.

Migration Example: Scanning and Uploading

- 1 Scan the files in the WebUpload Descriptor by choosing *Action > Scan*.

The Content Selector scans the root that you specified in the configuration, `c:/WebFiles`. It scans for all files, because you specified a filter of `*`, and it searches all subdirectories, because you checked **Recurse**.

During scanning:

- progress messages in Status View display the file that the Content Selector is currently scanning;
- an arrow in Status View points to Scanning Content;
- Log View displays a Scan tab. Since your log configuration specifies sending errors to the console, the Scan tab displays any errors.

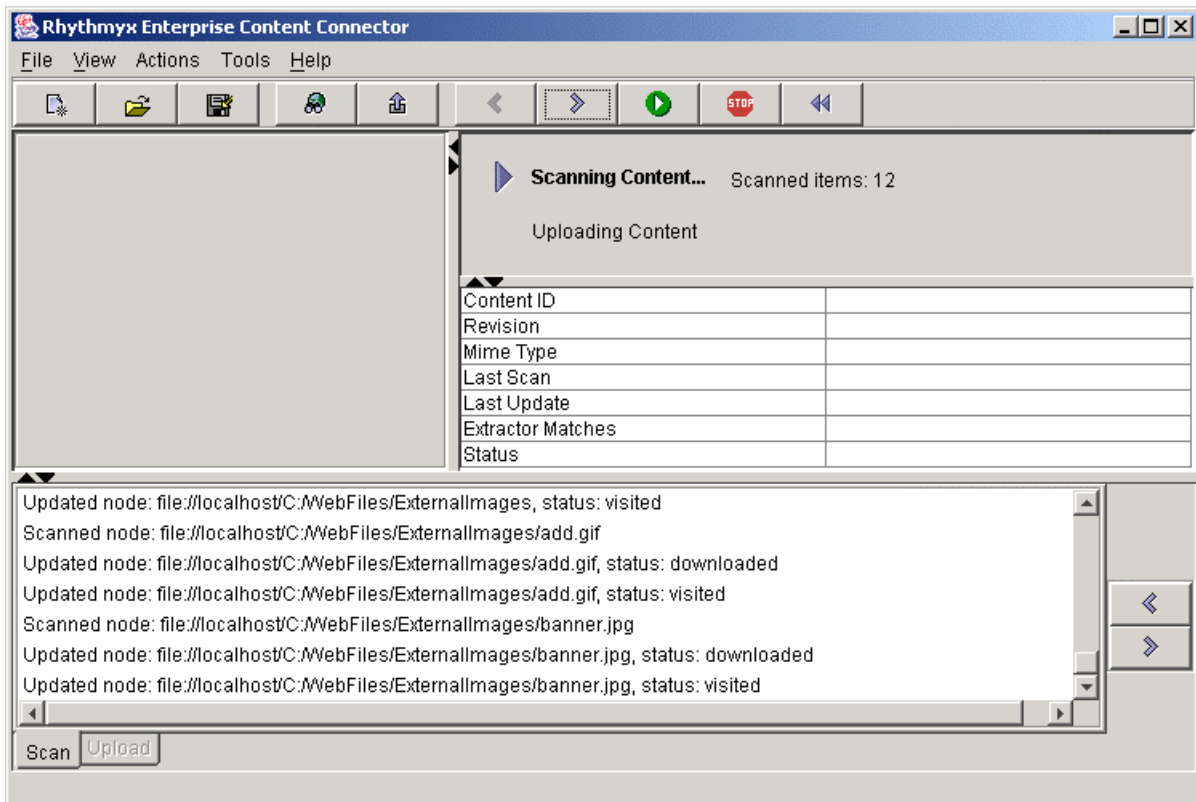


Figure 51: Scanning content

When scanning is complete:

- Scanning Content is checked in the Status View;
- a Content Tree shows files scanned.

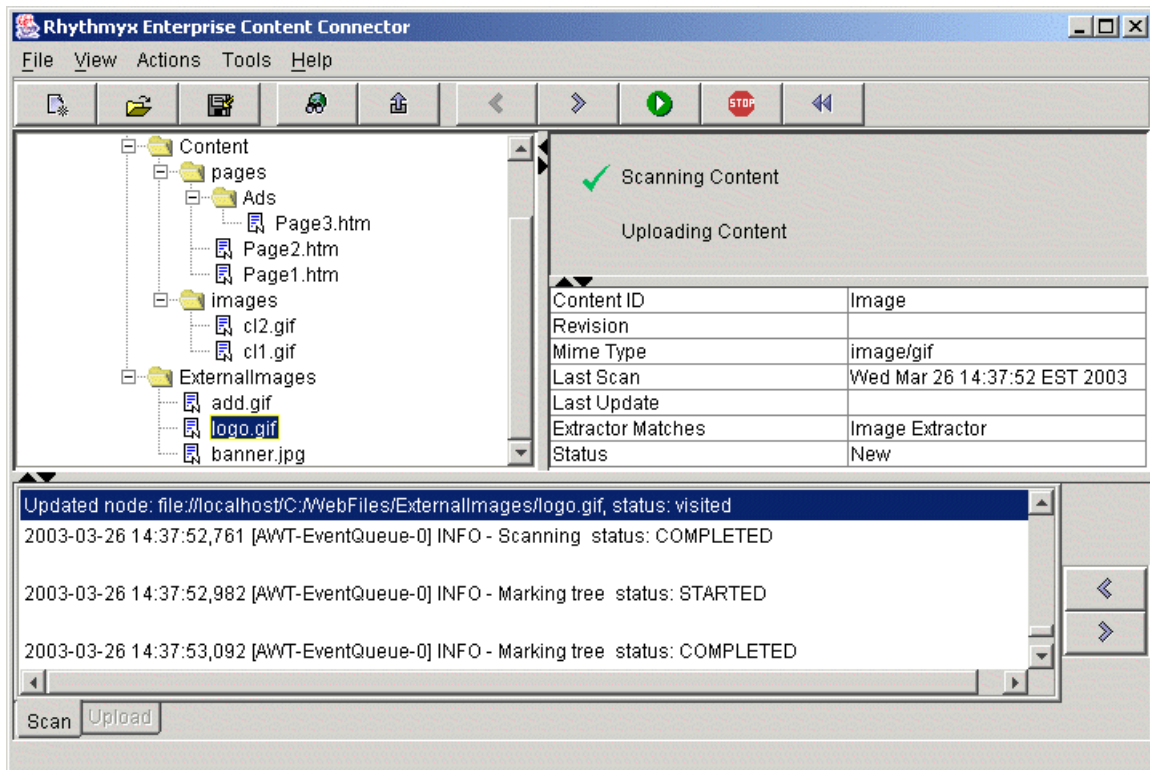



Figure 52: Content scan completed

- 2 Review the Scan.
 - a) Click files in the Content Tree and check the Content Type in Metadata *View*. **Reassign incorrect Content Types** (see "[Changing Content Types](#)" on page 68).
 - b) All files in the Content Tree should have the New icon  because this is the first time you are uploading from the source site. Mark any items that should not be uploaded as **Exclude** (see "[Content Tree](#)" on page 11).
 - c) Check for errors in the Scan Log. Click on an error to see the associated item, if any, highlighted in the Content Tree.
 - d) If necessary, correct errors and rescan.
- 3 To upload the scanned content into Rhythmyx, choose Action > Upload. During uploading, the Extractors that you have configured extract files and transform them, then upload them to Rhythmyx. Rhythmyx assigns the specified Workflow Actions to them. Since you are using a new Descriptor, all uploaded items are New and are added to Rhythmyx.

During uploading:

- progress messages in Status View display the upload location and the file currently being uploaded;
- in Status View, an arrow points to Uploading Content;
- Log View displays an Upload tab next to the Scan tab. This tab displays messages for errors that occur during uploading.

When uploading is complete, Uploading Content appears checked in Status View.

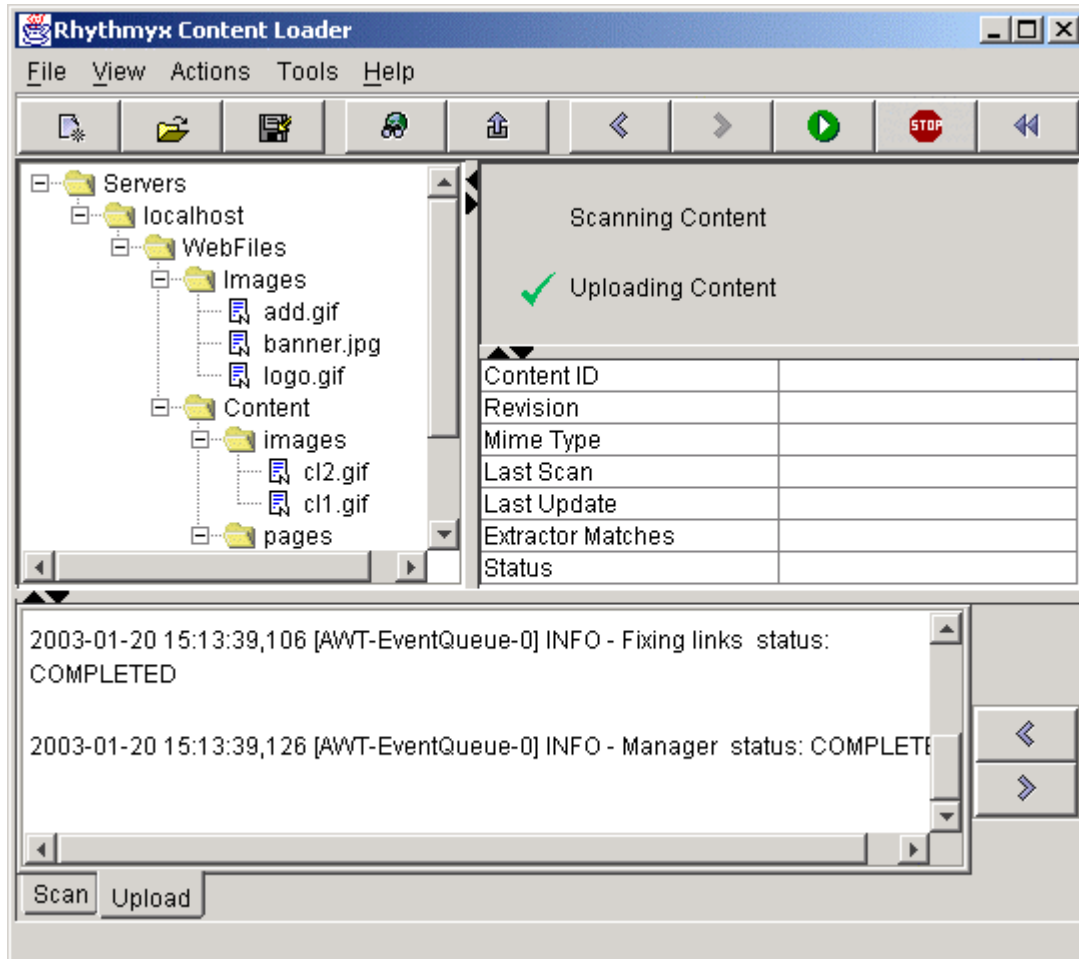


Figure 53: Content upload completed

Migration Example: Reviewing an Upload

To review an upload:

- 1 Review the Upload logs for errors. Click on an error to highlight the associated item, if any, in the Content Tree. If necessary, correct any errors. Rescan and/or reupload the items again.
- 2 Review the uploaded files in Rhythmyx.
 - a) Open Rhythmyx as the user Uploader1 and display your inbox.

- b) Check that the uploaded content items are listed with their original filenames under the States you intended.
- c) Open a few imported Image and Page content items and confirm that their contents are correct.

For example, after being uploaded, Page2.htm should have become a Page content item with the System Title Page2.htm. The content of the original file fills the Page content item's Content field.

Navigate to the Rhythmyx directory where you uploaded non-content items, rx_resources/images and confirm that the image files that you uploaded are present.

Incremental Upload Example

Incremental Upload Example: Introduction

The Incremental Upload example shows you how to use Enterprise Content Connector to upload syndicated content.

This example describes how set up and run a Descriptor to regularly upload syndicated information created and maintained outside of Rhythmyx. The non-Rhythmyx application stores new content in a repository along with older content. Each time you upload, you want to retrieve only new information and existing information that has been modified to avoid creating duplicate content items in Rhythmyx.

Incremental Upload Example: Procedure

Before uploading syndicated content, create a Content Type in Rhythmyx for storing the uploaded syndicated content. The Content Type can be similar to the Page Content Type in that it may insert all or most of the contents of each syndicated file into a content field. However, you may want to add additional fields to hold some of the information in specific HTML or XML tags (for example, dates or display titles).

After creating the Content Type, create a default Workflow for the syndicated Content Type. The Workflow for syndicated content will probably have fewer review States than other Workflows because syndicated content has already been edited. For a model of this Workflow, see *Creating the Syndicated Content Workflow*.

After you create the syndicated Content Type and Workflow in Rhythmyx, create a syndicated content Extractor in the Enterprise Content Connector. This Extractor may be similar to the Page Extractor, but should include code and Transformers for filling additional content fields that you may have added. The syndicated content Extractor must create content items of the syndicated Content Type. These content items will automatically acquire the default syndicated content Workflow when they enter Rhythmyx.

After you have created the Content Type, Workflow, and Extractor for the syndicated content upload, create a syndicated content upload Descriptor through the Enterprise Content Connector. Initially, run the Descriptor interactively to confirm that it works correctly. After running it, save the Status File so that during the next run, the Descriptor can determine which files are new or modified. Then modify the Descriptor's logging and error handling for batch processing, and create a batch file for running the Descriptor. Reconfigure error handling to send you a message whenever the batch file runs to remind you to enter Rhythmyx and publish or archive the uploaded content items. Finally, set up a scheduler program to run the batch program regularly.

Incremental Upload Example: Creating the Syndicated Workflow

Content items in a syndication repository are usually ready for publication or only require minor edits before you can publish them. The Workflow for these items does not require as many review states as Workflows for items created in Rhythmyx. The Workflow should allow the Uploader role to publish or archive each uploaded item and possibly return it to an edit state from which an Approver role can edit it. Once the item has entered the Publish or Archive state, an Admin role should have the ability to Archive or Retrieve it as in other Workflows.

The following is a simple model for the syndicated Workflow. You could include additional review States depending on the format of your uploaded content:

- **Uploaded State** - The State of content after an upload. Transitions from this state are Publish and Archive. Only the Uploader Role can transition content in this state.
- **Public State** - Transition from this state is Unpublish, which sends the content to the Archive state. Transition Role or Roles are those that you normally assign to the Unpublish transition, such as Admin.
- **Archive** - Archive State for content items archived from the Publish state and for uploaded content items that the Uploader does not want to publish. Include a Transition for retrieving items back to the Uploaded state. Transition Role or Roles are those that you normally assign to the Retrieve transition, such as Admin.

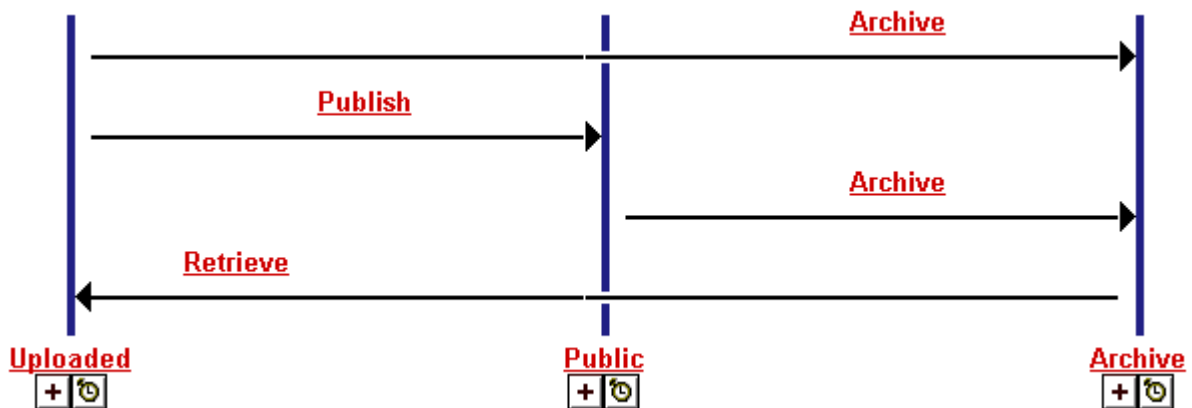


Figure 54: Syndicated Workflow

In Rhythmyx, make the syndicated content Workflow the default Workflow of the syndicated Content Type.

Incremental Upload Example: Creating the Descriptor and Performing the Initial Upload

To initially create the Descriptor for the incremental upload, follow the steps in the *Migration Example* (see "[Migration Example: Creating a Descriptor](#)" on page 76) but in Content Selector, choose *File Selector for syndication*, and in Extractors, configure only the Extractor that you created for syndicated content uploads.

Since you chose *File Selector for syndication*, in subsequent runs, the Enterprise Content Connector will compare scanned content items with previously loaded items to determine which items are New and Changed.

Do all necessary checks and debugging after the initial upload. The process should run smoothly in the interface before you begin performing batch uploads.

After you determine that the upload is running correctly, *modify the Descriptor for batch processing* (see "[Incremental Upload Example: Batch Processing](#)" on page 97).

Incremental Upload Example: Batch Processing

Before you perform batch processing, modify the Logging and Error Settings in your Descriptor.

The screenshot shows a configuration window with two main sections: "Default Log Level" and "Log Location".

Default Log Level: This section contains six radio button options arranged in two columns. The "Error" option in the left column is selected.

- Off
- Fatal
- Error**
- Warning
- Info
- Debug

Log Location: This section contains a checked checkbox for "File".

File Settings: This sub-section contains two radio button options, "Append" and "Overwrite", with "Overwrite" selected. Below these is a checked checkbox for "Backup".

Below the "Backup" checkbox is a text input field containing the number "10", followed by the text "Maximum Backups".

Figure 55: Configuring logging for an incremental upload

Configure the following Logging settings:

- Under **Log Location**, check *File* so that log data is saved to file.
- Under **File Settings**, click *Overwrite*. Since loading will run repeatedly, the file may become excessively long if you choose the default option, *Append*. Check *Backup* and enter *10* in the *Maximum Backups* field. Since you are performing incremental loading, you may need to check for errors that occurred in previous uploads.

The screenshot shows a configuration window with the following elements:

- Stop on Error
- Email on Error
- Email on Success
- email
 - To:
 - CC:
- Attach Logs

Figure 56: Configuring error handling for an incremental upload

Configure the following Error Handling Settings:

- Check **Email on Error** and **Email on Success** to be informed of the status of scheduled batch runs.
- Under **To**, enter the email address of the Uploader.
- Check **Attach Logs** to receive error logs with notifications.

To set up batch processing:

- 1 In your `c:\Rhythmyx\contentconnector` directory, create a batch file with the command line:

```
RhythmyxContentConnector -descriptor <Descriptor path> -noui
```

This command executes your Descriptor.

- 2 Set up a scheduler program to run the batch file regularly.

When you run the batch file, it locates New and Changed items to upload. It adds the New items to Rhythmyx and replaces existing items with Changed items. When the items are uploaded, the Uploader receives a notification and can then enter Rhythmyx to review the items and send them to Publish or Archive.

CHAPTER 6

Configuring The Rhythmyx Enterprise Content Connector

When you initially prepare to use the Rhythmyx Enterprise Content Connector, change its default settings to correspond with your system. Default settings are stored in the configuration file: `<Rhythmyx root>/rxconfig/contentconnector/contentconnector.xml`. This file contains the plugin selections available in the Descriptor Setup dialog, the default plugin options, and other settings. Open `contentconnector.xml` in an xml editor and manually add new plugins, remove plugins that you want to make unavailable, and change default values.

Changing Configuration File Default Values

Change the values stored in `<Rhythmyx root>/rxconfig/contentconnector/contentconnector.xml` to correspond with your system and to display custom options for Descriptors in the Descriptor Setup dialog.

NOTE: Users can also change configuration settings for specific uploads in Descriptors.

The following table indicates default `contentconnector.xml` element and attribute values that you may need to change when you initially begin using The Rhythmyx Enterprise Content Connector. Some of the elements and attributes listed are not used in the default `contentconnector.xml`; see `contentconnector.dtd` for their complete definitions.

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
Rhythmyx username, password, server or port	<pre> <Connection> <Properties> <PSXProperty name="Rhythmyx Server Protocol"> <Value>http</Value> </PSXProperty> <PSXProperty name="Rhythmyx Server"> <Value>localhost</Value> </PSXProperty> <PSXProperty name="Port"> <Value>9992</Value> </PSXProperty> </Properties> <User>admin1</User> <Password encrypted="no">demo</Password> </Connection> </pre>	<p>Rhythmyx Server Protocol - protocol used by your Rhythmyx server, such as http or https;</p> <p>Rhythmyx Server - target server name;</p> <p>Port - target server port;</p> <p>User - Rhythmyx user performing the upload;</p> <p>Password - Rhythmyx password;</p> <p>Encrypted - password encryption.</p>

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
File Selector settings	<pre> <Selectors> <PSXContentSelectorDef name="File Selector for migration" class="com.percussion.loader.selector.PSFileSelector"> <Properties> <PSXProperty name="Calculate checksum"> <Value>no</Value> </PSXProperty> </Properties> <PSXSearchRoots> <PSXSearchRoot name="Default"> <Properties> <PSXProperty name="Search Root"> <Value>C:\inetpub\wwwroot </Value> </PSXProperty> <PSXProperty name="Recurse"> <Value>yes</Value> </PSXProperty> </Properties> <Filters> <PSXFilter name="Default"> <Value>*</Value> </PSXFilter> </Filters> </PSXSearchRoot> </PSXSearchRoots> </PSXContentSelectorDef> </pre>	<p>Calculate Checksum - whether or not to check if files have been previously uploaded or have changed;</p> <ul style="list-style-type: none"> ▪ No - do not check if files have been previously uploaded or have changed; upload all files as new. ▪ Yes - check if files have been previously uploaded or have changed and do not upload previously uploaded, unchanged files. <p>Search Root - location where Content Selector searches on source;</p> <p>Recurse - true/false indicates search through subfolders/do not search through subfolders;</p> <p>Filter Value - search expression that indicates which files to select.</p>

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
List Selector Settings	<pre> <Selectors> . . . <PSXContentSelectorDef name="List Selector for syndication" class="com.percussion.loader.selector.PSListContentSelec tor"> <Properties> <PSXProperty name="ContentList"> <Value>c:\ListSelector.xml</Value> </PSXProperty> <PSXProperty name="Calculate checksum"> <Value>yes</Value> </PSXProperty> </Properties> </PSXContentSelectorDef> </Selectors> </pre>	<p>ContentList - the file holding addresses of files to select;</p> <p>Calculate Checksum - whether or not to check if files have been previously uploaded or have changed;</p> <ul style="list-style-type: none"> ▪ No - do not check if files have been previously uploaded or have changed; upload all files as new; ▪ Yes - check if files have been previously uploaded or have changed and do not upload previously uploaded, unchanged files.

To modify the default ...	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
<p>Static Item Extractor settings</p>	<p>In the <Extractors> element, find <PSXExtractorDef> elements with type="staticItem" such as:</p> <pre data-bbox="394 407 1373 1377"> <Extractors> <PSXExtractorDef name="\$rximage" type="staticItem" class="com.percussion.loader.extractor.PSStaticItemExtractor"> <Properties> <PSXProperty name="ContextVariableName"> <Value>rximage</Value> </PSXProperty> <PSXProperty name="Source Location"> <Value>images</Value> </PSXProperty> <PSXProperty name="Target Location"> <Value>web_resources/xroads/resources/images</Value> </PSXProperty> </Properties> <MimeTypes> <PSXMimeTypeDef name="image/gif"> <PSXExtensionDef name="gif" /> </PSXMimeTypeDef> <PSXMimeTypeDef name="image/jpeg"> <PSXExtensionDef name="jpeg" /> </PSXMimeTypeDef> <PSXMimeTypeDef name="image/jpg"> <PSXExtensionDef name="jpg" /> </PSXMimeTypeDef> </MimeTypes> <Filters> <PSXFilter name="All"> <Value>*</Value> </PSXFilter> </Filters> </PSXExtractorDef><Value>*</Value> </PSXFilter> </Filters> </PSXExtractorDef> </pre>	<p>ContextVariableName - context Variable used in Rhythmyx link generation schemes to locate the item;</p> <p>Source Location - folder on source site containing these items;</p> <p>Target Location - folder in Rhythmyx root or filesystem where these items are uploaded;</p> <p>PSXMimeTypeDef - Mime type of file to extract;</p> <p>PSXExtensionDef - file extension that matches Mime type and indicates files to extract;</p> <p>PSXFilter Value - search expression that indicates which files to extract.</p>

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
Item Extractor settings	<p>In the <Extractors> element, find <PSXExtractorDef> elements with type="Item" such as:</p> <pre> <PSXExtractorDef class="com.percussion.loader.extractor.PSPageExtractor" name="Page Extractor" type="item"> <FieldProperties> <PSXFieldProperty name="sys_title" valueType="Variable"> <Value>\$filename</Value> </PSXFieldProperty> <PSXFieldProperty name="sys_suffix" valueType="Variable"> <Value>\$fileextension</Value> </PSXFieldProperty> <PSXFieldProperty name="sourceurl" valueType="Variable"> <Value>\$filepath</Value> </PSXFieldProperty> <PSXFieldProperty name="content" valueType="Variable"> <Value>\$content</Value> </PSXFieldProperty> </FieldProperties> <Properties> <PSXProperty name="ContentTypeName"> <Value>Page</Value> </PSXProperty> <MimeTypes> <PSXMimeTypeDef name="text/plain"> <PSXExtensionDef name="htm"/> <PSXExtensionDef name="html"/> </PSXMimeTypeDef> </MimeTypes> <Filters> <PSXFilter name="All"> <Value>*.htm</Value> </PSXFilter> </Filters> <PSXWorkflowDef name="Default"> <TransitionSet name="insert"> <PSXTransitionDef name="Submit" trigger="Submit"/> <PSXTransitionDef name="Approve" trigger="Approve"/> </TransitionSet> </PSXWorkflowDef> </pre>	<p>Definitions of bold values:</p> <p>PSXFieldProperty - Default Content Type field mappings.</p> <ul style="list-style-type: none"> ▪ Name - Content Item name; ▪ ValueType - type of value; options are Literal, Number, or Variable for all Extractors and also Xpath for XML and XSL Extractors; ▪ Value - property value; <p>ContentTypeName - Rhythmyx Content Type;</p> <p>PSXMimeTypeDef - Mime type of file to extract;</p> <p>PSXExtensionDef - file extension that matches Mime type and indicates files to extract;</p> <p>PSXFilter - search expression that indicates which files to extract;</p> <p>Name - filter name;</p> <p>PSXFieldTransformation Def or PSXTransformation Def (not present by default, see DTD);</p> <p>Name - name of field transformation Java extension;</p> <p>Class - Java class for extension;</p> <p>PSXParamDef</p> <p>Name - extension parameter name;</p> <p>Type - parameter type;</p> <p>Description - parameter description;</p> <p>Value - default parameter value;</p>

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
	<pre> </PSXExtractorDef> <PSXTransitionDef name="Approve" /> </PSXWorkflowDef> </Workflows> </PSXActionDef> </Actions> </PSXExtractorDef> </pre>	<p>PSXTransitionDef - Workflow transition to include in drop list;</p> <p>Name - Transition name.</p>
Rhythmyx Loader settings	<pre> <Loaders> <PSXLoaderDef name="Rhythmyx Loader" class="com.percussion.loader.PSContentLoader" /> . . . </pre>	<p>Settings for the Rhythmyx Loader are in the <Properties> element.</p>
Preview Loader Settings	<pre> <Loaders> . . . <PSXLoaderDef name="Rhythmyx Loader Preview" class="com.percussion.loader.PSContentLoaderPreview"> <Properties> <PSXProperty name="PreviewPath"> <Value>C:\temp</Value> </PSXProperty> <PSXProperty name="StaticRoot"> <Value>RhythmyxServer</Value> </PSXProperty> </Properties> </PSXLoaderDef> </Loaders> </pre>	<p>PreviewPath - location of filesystem to which content is uploaded;</p> <p>StaticRoot - the folder in PreviewPath that holds external items.</p>

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
Logging properties	<pre> <log4j:configuration xmlns:log4j="http://jakarta.apache.org/log4j/"> <appender class="org.apache.log4j.FileAppender" name="fileAppender"> <layout class="org.apache.log4j.PatternLayout"> <param name="ConversionPattern" value="%d [%t] %p - %m%n" /> </layout> <param name="Append" value="yes" /> <param name="Backup" value="no" /> <param name="MaxBackups" value="0" /> <param name="File" value="D:\\rx50\\contentconnector\\WebUpload\\WebUpload.log" /> </appender> <appender class="com.percussion.loader.PSLogDispatcher" name="logDispatcher"> <layout class="org.apache.log4j.PatternLayout"> <param name="ConversionPattern" value="%d [%t] %p - %m%n" /> </layout> </appender> </root> <priority value="error" /> <appender-ref ref="logDispatcher" /> </root> </log4j:configuration> </pre>	See http://jakarta.apache.org/log4j/docs (http://jakarta.apache.org/log4j/docs) for information about changing log4j configurations.

To modify the default . . .	In contentconnector.xml find the following element and modify any of the highlighted values.	Definitions of bold values:
Error Handling settings	<pre> <PSXErrorHandlingDef> <Properties> <PSXProperty name="stoponerror" > <Value>no</Value> </PSXProperty> <PSXProperty name="emailonerror" > <Value>no</Value> </PSXProperty> <PSXProperty name="emailonsuccess" > <Value>no</Value> </PSXProperty> <PSXProperty name="attachlog" > <Value>False</Value> </PSXProperty> <PSXProperty name="smtphost" > <Value>pan</Value> </PSXProperty> </Properties> <PSXEmailDef attachlogs="no" fromAddr="rxserver@percussion.com" smtpHost="pan" > <Recipients> <Recipient>rxmaster@percussion.com</Recipient> </Recipients> </PSXEmailDef> </PSXErrorHandlingDef> </pre>	<p>Stoponerror - true/false to stop/not stop processing on error and fatal events;</p> <p>Emailonerror - true/false to email/not email recipients on error and fatal events;</p> <p>Emailon success - true/false to email/not email recipients when upload is successful;</p> <p>Attachlog - true/false to attach error log/not attach error log to emails;</p> <p>smtphost - name of SMTP host;</p> <p>fromAddr - address marked as sender on success or error emails;</p> <p>Recipients</p> <p>Recipient - address to receive success or error emails;</p> <p>CCRecipients - address to receive success or error emails.</p>

Adding New Plugins

You can create custom plugins for Content Selectors, Extractors, Field Transformers, Item Transformers and Content Loaders and add them to `<Rhythmyx`

`root>/rxconfig/contentconnector/contentconnector.xml`. After you add them, the selection lists in the Descriptor Setup dialog include them as options, and the Content Tree includes any that you select. The Content Tree only displays folders for Item and Field Transformers if you add Transformer plugins.

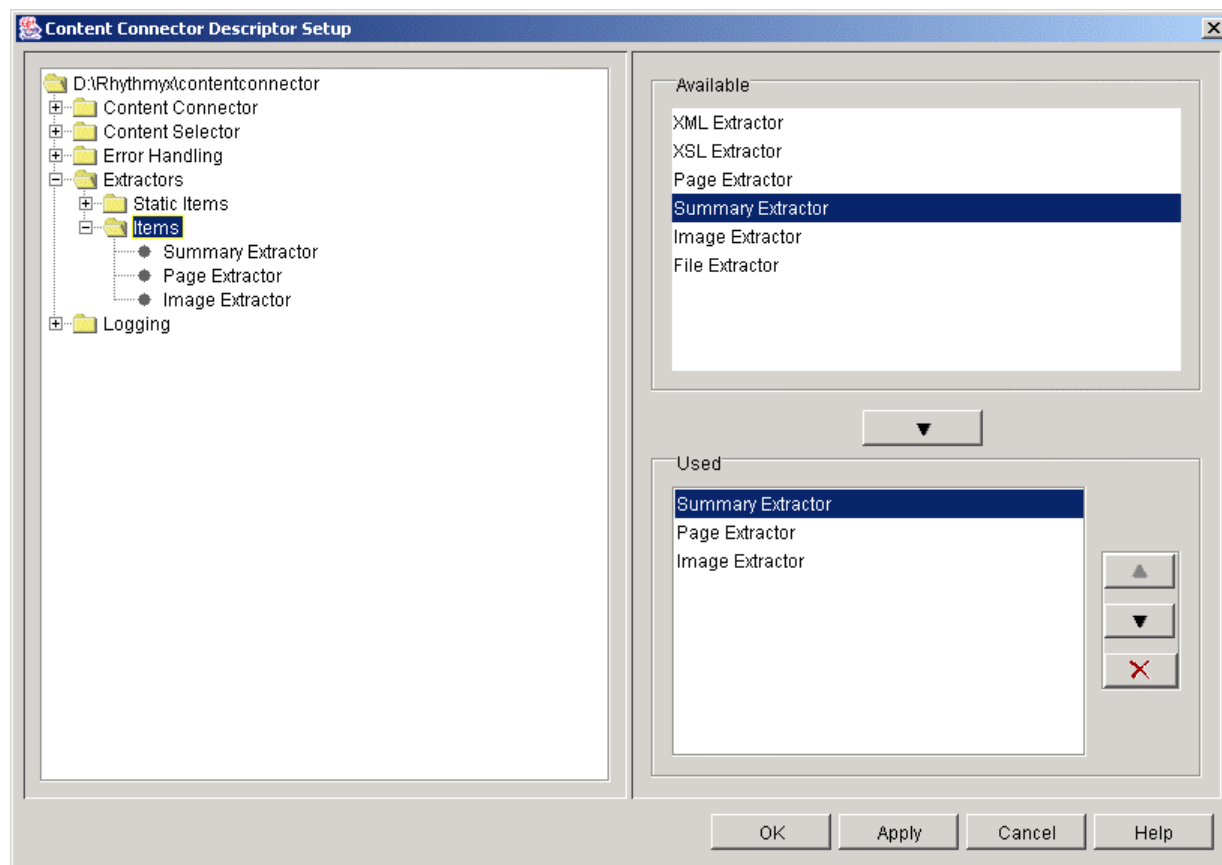


Figure 57: Descriptor Setup dialog displaying custom Summary Extractor

The following table shows which element type to modify and add to `contentconnector.xml` when you write a custom plugin. For help modifying the element, see ***Changing Configuration File Default Values*** (on page 101) and `contentconnector.dtd`. For information about writing the plugin, click the plugin type in the first column of the table.

To add a
custom
...
Content
Selector

In
contentconnector
.xml find ...
<Selectors>

And add the following element. Change the element values to match your plugin.

```
<Selectors>
  <PSXContentSelectorDef name="File Selector for migration"
class="com.percussion.loader.selector.PSFileSelector">
    <Properties>
      <PSXProperty name="Calculate checksum">
        <Value>no</Value>
      </PSXProperty>
    </Properties>
    <PSXSearchRoots>
      <PSXSearchRoot name="Default">
        <Properties>
          <PSXProperty name="Search Root">
            <Value>C:\Inetpub\wwwroot</Value>
          </PSXProperty>
          <PSXProperty name="Recurse">
            <Value>yes</Value>
          </PSXProperty>
        </Properties>
      </PSXSearchRoot>
    </PSXSearchRoots>
    <Filters>
      <PSXFilter name="Default">
        <Value>*</Value>
      </PSXFilter>
    </Filters>
  </PSXContentSelectorDef>
  . . .
```

To add a
custom
...
Static
Item
Extractor

In
contentconnector
.xml find ...
<Extractors>

And add the following element. Change the element values to match your plugin.

```
<PSXExtractorDef name="$rximage" type="staticItem"
class="com.percussion.loader.extractor.PSStaticItemExtractor">
  <Properties>
    <PSXProperty name="ContextVariableName">
      <Value>rximage</Value>
    </PSXProperty>
    <PSXProperty name="Source Location">
      <Value>images</Value>
    </PSXProperty>
    <PSXProperty name="Target Location">
      <Value>web_resources/xroads/resources/images</Value>
    </PSXProperty>
  </Properties>
  <MimeTypes>
    <PSXMimeTypeDef name="image/gif">
      <PSXExtensionDef name="gif"/>
    </PSXMimeTypeDef>
    <PSXMimeTypeDef name="image/jpeg">
      <PSXExtensionDef name="jpeg"/>
    </PSXMimeTypeDef>
    <PSXMimeTypeDef name="image/jpg">
      <PSXExtensionDef name="jpg"/>
    </PSXMimeTypeDef>
  </MimeTypes>
  <Filters>
    <PSXFilter name="All">
      <Value>*</Value>
    </PSXFilter>
  </Filters>
</PSXExtractorDef>
```

To add a
custom
...
Item
Extractor

In
contentconnector
.xml find ...
<Extractors>

And add the following element. Change the element values to match your plugin.

```

    <Extractors>
      <PSXExtractorDef
class="com.percussion.loader.extractor.PSPageExtractor" name="Page Extractor"
type="item">
        <FieldProperties>
          <PSXFieldProperty name="sys_title" valueType="Variable">
            <Value>$filename</Value>
          </PSXFieldProperty>
          <PSXFieldProperty name="sys_suffix" valueType="Variable">
            <Value>$fileextension</Value>
          </PSXFieldProperty>
          <PSXFieldProperty name="sourceurl" valueType="Variable">
            <Value>$filepath</Value>
          </PSXFieldProperty>
          <PSXFieldProperty name="content" valueType="Variable">
            <Value>$content</Value>
          </PSXFieldProperty>
        </FieldProperties>
        <Properties>
          <PSXProperty name="ContentTypeName">
            <Value>Page</Value>
          </PSXProperty>
        <MimeTypeTypes>
          <PSXMimeTypeDef name="text/plain">
            <PSXExtensionDef name="htm"/>
            <PSXExtensionDef name="html"/>
          </PSXMimeTypeDef>
        </MimeTypeTypes>
        <Filters>
          <PSXFilter name="All">
            <Value>*.htm</Value>
          </PSXFilter>
        </Filters>
        <PSXWorkflowDef name="Default">
          <TransitionSet name="insert">
            <PSXTransitionDef name="Submit" trigger="Submit"/>
            <PSXTransitionDef name="Approve" trigger="Approve"/>
          </TransitionSet>
        </PSXWorkflowDef>
      </PSXExtractorDef>

```

To add a
custom
...
Field
Transformer

In
contentconnector
.xml find ...
<FieldTransform
ations>

And add the following element. Change the element values to match your plugin.

By default, there are no Field Transformer elements in contentconnector.xml. Add the an element with the following format within a <PSXExtractorDef> element to add an Extractor-specific Field Transformer or within the <PSXContentLoaderConfig> element to a global Field Transformer:

```
<PSXFieldTransformationsDef>
  <PSXFieldTransformationDef name="sys_PsXslFieldTransformation1"
class="com.percussion.loader.PSFieldTransformerTest"
targetField="bodycontent">
  <ParamDefs>
    <PSXParamDef name="stylesheet" type="java.lang.String">
      <Description>The stylesheet to use for the field
transformation. Must be a fully qualified file name.</Description>
    </PSXParamDef>
  </ParamDefs>
</PSXFieldTransformationDef>
</PSXFieldTransformationsDef>
```


To add a custom ... Item Transformer	In contentconnector.xml find ... <ItemTransformations>	<p>And add the following element. Change the element values to match your plugin.</p> <p>By default, there are no Item Transformer elements in contentconnector.xml. Add the an element with the following format within a <PSXExtractorDef> element to add an Extractor-specific Item Transformer or within the <PSXContentLoaderConfig> element to a global Item Transformer:</p> <pre data-bbox="609 487 1680 812"> <PSXItemTransformationsDef> <PSXTransformationDef name="sys_PSXslItemTransformation1" class="com.percussion.loader.PSItemTransformerTest"> <ParamDefs> <PSXParamDef name="stylesheet" type="java.lang.String"> <Description>The stylesheet to use for the field transformation. Must be a fully qualified file name.</Description> </PSXParamDef> </ParamDefs> </PSXTransformationDef> </PSXItemTransformationsDef> </pre>
Content Loader	<Loaders>	<pre data-bbox="609 844 1449 1292"> <Loaders> <PSXLoaderDef name="Rhythmyx Loader" class="com.percussion.loader.PSContentLoader"/> <PSXLoaderDef name="Loader Preview" class="com.percussion.loader.PSContentLoaderPreview"> <Properties> <PSXProperty name="PreviewPath"> <Value>C:\temp</Value> </PSXProperty> <PSXProperty name="StaticRoot"> <Value>RhythmyxServer</Value> </PSXProperty> </Properties> </PSXLoaderDef> </Loaders> </pre>

Writing Content Selectors

Content Selectors select content for uploading. The Enterprise Content Connector supplies File Selector and List Selector Content Selectors for migration and syndication uploads, and you may create custom Content Selectors. All Content Selectors must implement the IPSContentSelector interface. See the IPSContentSelector Javadoc

(`<rxroot>\docs\Rhythmyx\javadocs\com\percussion\loader\IPSContentSelector.html`) for more information.

After you create a Content Selector plugin using IPSContentSelector, create a `<Selector>` element for the plugin in `<Rhythmyx root>/rxconfig/contentconnector/contentconnector.xml` in the `<Selectors>` element.

Writing Extractors

Extractors attempt to match content in source objects with the Content Types (or static item types) that they represent, and extract the files from the source site. The Enterprise Content Connector provides Page, Binary, XML and XSL Item Extractor classes and a Static Item Extractor class. You may create custom Extractors that implement the IPSItemExtractor interface and are derived from the PSExtractor class. You must define the extractItem method for your new class.

For example:

You can create a custom static extractor that is derived from PSStaticItemExtractor.

```
public class CustomizedStaticExtractor extends PSStaticItemExtractor
{
    // specify override methods . . . etc
}
```

You can create a custom item extractor with an empty Editor pane that is derived from PSItemExtractor

```
public class PSTestExtractor extends PSItemExtractor
{
    // Implements IPSItemExtractor.extractItems(PSItemContext,
    // InputStream)
    public PSItemContext[] extractItems(PSItemContext resource,
    InputStream in)
        throws java.io.IOException
    {
        // process fields, one at a time
        PSClientItem clientItem = resource.getItem();
        Iterator fieldProperties = getExtractorDef().getFieldProperties();
        while (fieldProperties.hasNext())
        {
            PSFieldProperty property = (PSFieldProperty)
fieldProperties.next();
            PSItemField itemField = getParentField(clientItem,
property.getName());

            // set your value to this field ...
        }

        PSItemContext[] items = new PSItemContext[1];
        items[0] = resource;

        return items;
    }
}
```

You can create a custom item extractor with some UI capability that is derived from `PSItemExtractor` and uses `PSItemExtractorConfigPanel` to define its Editor pane.

The Item Extractor Editor will have four tabs with the same fields as it has for all Item Extractors, but the Fields tab will be empty.

See the `PSItemExtractor` Javadoc

(`<rxroot>\docs\Rhythmyx\javadocs\com\percussion\loader\PSItemExtractor.html`) for more information.

Example:

```
public class PSTestExtractor extends PSItemExtractor
{
    // Implements PSItemExtractor.extractItems(PSItemContext,
    // InputStream)
    public PSItemContext[] extractItems(PSItemContext resource,
    InputStream in)
        throws java.io.IOException
    {
        // Implements the IPSUIPlugin interface
        public PSConfigPanel getConfigurationUI()
        {
            return new PSItemExtractorConfigPanel();
        }

        // process fields, one at a time
        PSClientItem clientItem = resource.getItem();
        Iterator fieldProperties = getExtractorDef().getFieldProperties();
        while (fieldProperties.hasNext())
        {
            PSFieldProperty property = (PSFieldProperty)
fieldProperties.next();
            PSItemField itemField = getParentField(clientItem,
property.getName());

            // set your value to this field ...
        }

        PSItemContext[] items = new PSItemContext[1];
        items[0] = resource;

        return items;
    }
}
```

After you create an Extractor plugin using IPSItemExtractor, create an <Extractor> element for the plugin in <Rhythmyx root>/rxconfig/ContentConnector/contentconnector.xml in the <Extractors> element. Add the new Extractor to your classpath so that the Enterprise Content Connector can locate it.

If you create an Static Item Extractor plugin, you must register the plugin ContextVariableNames that you specify in contentconnector.xml as context variables in Rhythmyx if they are not already registered. See the topic “Adding Variables for Context-Dependent Resources” in the Rhythmyx online help for information about registering context variables.

Writing Transformers

Transformers are Java extensions that modify content extracted from the source site so that it conforms to Rhythmyx's format. For example, you may need to restructure date formats, add metadata fields, or delete description fields that are not used in Rhythmyx. The reformatting required is dependent on the format of your source content; therefore, Rhythmyx does not provide any Transformers.

Write Transformers, then create `<Transformation>` elements for them in `<Rhythmyx root>/rxconfig/ContentConnector/contentconnector.xml` in the `<Field Transformations>` element or `<Item Transformations>` element.

Writing Loaders

Content Loaders load extracted content into the target location. The Enterprise Content Connector supplies the Rhythmyx Loader and Preview Loader Content Loaders, and you may create custom Content Loaders. All Content Loaders must implement the IPSContentLoader interface. See the IPSContentLoader Javadoc

(`<rxroot>\docs\Rhythmyx\javadocs\com\percussion\loader\IPSContentLoader.html`) for more information.

After you create a Content Loader plugin using IPSContentLoader, create a `<Loader>` element for the plugin in `<Rhythmyx root>/rxconfig/ContentConnector/contentconnector.xml` in the `<Loaders>` element.

CHAPTER 7

The Enterprise Content Connector and Rhythmyx

The Enterprise Content Connector attempts to upload content into Rhythmyx without requiring you to change your Rhythmyx system or to reformat content once it enters Rhythmyx. However, a few enhancements to Rhythmyx make uploading with the Enterprise Content Connector a more efficient process:

- creating a Rhythmyx user whose function is to upload content ensures that the user performing the process always has the necessary access and creates audit trail records that you may trace to uploading;
- adding Workflow Transitions that let you skip review and editing states make it easier to publish uploaded content that has already been edited.

Creating a Rhythmyx User to Load Content

Create a “uploading” user in Rhythmyx with the access required to log in to Rhythmyx, create the uploaded Content Types, and perform the Workflow Actions that you assign in Extractors. In the Descriptor, enter this user as the **Username** for accessing Rhythmyx. Having a separate user for loading content is especially useful for debugging if you are performing incremental uploads because the Workflow Actions assigned in the Descriptor are stamped with the “uploading” user.

In addition, you may want to create users who can perform any *special Workflow Transitions that you add for uploaded content* (see "[Reconfiguring Workflows for Uploaded Content](#)" on page 127).

Reconfiguring Workflows for Uploaded Content

Content that you upload to Rhythmyx may already be edited and ready for publication. In most cases, you should not send uploaded content immediately into a public state, but you may want to limit the edits and approvals required. For example, you may want to add a publish transition that follows an initial review state. Uploaded content could enter Rhythmyx in the Review state, and a user with the correct access could review these items and move them directly into the Publish state.

Syndicated content and other incrementally uploaded content is often time-sensitive. Use aging transitions with this content so it does not remain on a site when it is no longer current. In addition, when you upload a group of syndicated content items, some of them may not be appropriate for your site. You may find it useful to include transitions from Review to Archive states, so that you can avoid publishing these items or further processing them in Rhythmyx.

Index

A

- Adding a Descriptor from another Host • 24
- Adding a New Extractor • 40
- Adding New Plugins • 111
- Adding New PluginsConfiguration File
 - Plugin ElementsPlugins
 - Adding • 111

B

- Batch Processing • 72
- Binary Extractor • 35, 45

C

- Changing Configuration File Default Values • 31, 58, 59, 101, 111
- Changing Configuration File Default ValuesConfiguration File
 - Changing Defaults • 101
- Changing Content Types • 68, 91
- Choosing a Content Loader • 54, 77
- Choosing a Content Selector • 25
- Configuration File
 - Changing Defaults • 101
 - Defined • 100
 - Plugin Elements • 111
- Configuring a File Content Selector • 27, 80
- Configuring a File Content Selector Filter • 28, 29
- Configuring a List Content Selector • 27
- Configuring an Extractor • 42
- Configuring Content Loaders • 17, 21, 22, 54
- Configuring Content Selectors • 17, 21, 22, 25
- Configuring Descriptor Error Handling • 18, 21, 22, 59
- Configuring Field Transformers • 21, 22, 50
- Configuring Item Transformers • 21, 22, 52
- Configuring Logging • 18, 21, 22, 57
- Configuring the Preview Loader • 56
- Configuring The Rhythmyx Enterprise Content Connector • 66, 100
- Configuring the Rhythmyx Loader • 56
- Configuring Transformers • 17, 49

- Content Connector Interface
 - Content Tree • 11
 - Introduction • 7
 - Log View • 15
 - Menu Bar • 8
 - Metadata View • 14
 - Status View • 13
 - Toolbar • 10
- Content Extractors
 - Adding • 40
 - Assigning to Descriptor • 32
 - Binary • 45
 - Configuring • 42
 - Content Type Tab • 34
 - Defined • 31
 - Fields Tab • 35
 - Filters Tab • 36
 - Item Extractors • 33
 - Location Tab • 39
 - Page • 46
 - Static • 47
 - Static Item Extractors • 38
 - Variables • 48
 - Workflow Tab • 37
 - Writing • 119
 - XML • 44
 - XSL • 43
- Content Loaders
 - Choosing • 54
 - Configuring Preview Loader • 56
 - Configuring Rhythmyx Loader • 56
- Content Selectors
 - Choosing • 25
 - Configuring File Selector • 27
 - Configuring List Selector • 27
 - Defined • 25
 - Filtering • 29
- Content Syndication Connector License • 94
- Content Tree • 7, 11, 14, 65, 91
- Content Type Tab • 34
- contentconnector.xml • 100
- Copyrights and Trademarks • 2
- Creating a New Descriptor by Modifying an Existing Descriptor • 23
- Creating a Rhythmyx Enterprise Content Connector Descriptor • 8, 18, 21, 64
- Creating a Rhythmyx User to Load Content • 126

D

- Default Descriptor Preference • 16
- Default Extractors • 43
- Descriptor Setup Dialog • 4, 9, 18, 19, 21, 22, 23, 25, 27, 50, 52, 54, 56, 58, 59
- Descriptors
 - Creating • 21
 - Creating through Modification • 23
 - Descriptor Setup Dialog • 19
 - From other Hosts • 24
 - Introduction • 17
 - Modifying • 22
 - Saving • 61

E

- Editing a Rhythmyx Enterprise Content
 - Connector Descriptor • 8, 18, 22, 64
- Error Handling • 59
- Examples
 - Incremental Upload Example
 - Batch Processing • 97
 - Description • 94
 - Descriptor • 96
 - Initial Upload • 96
 - Logging and Error Settings • 97
 - Procedure • 94
 - Workflow • 95
 - Site Migration Example
 - Descriptor • 76
 - Reviewing Upload • 92
 - Roles and Workflows • 75
 - Scanning and Uploading • 89
 - Upload Information and Source Server Data • 74
- Examples of Rhythmyx Enterprise Content
 - Connector Operations • 73
- Excluding Items • 68
- Extractor Editors • 33
- Extractor Variables • 48

F

- Fields Tab • 35
- Filtering • 29

- Filters Tab • 36, 38

H

- How The Rhythmyx Enterprise Content Connector Works • 4

I

- Implementing Site Migration with Enterprise Content Connector
 - One Time Migration Example • 74
- Incremental Upload Example • 94
 - Batch Processing • 96, 97
 - Creating the Descriptor and Performing the Initial Upload • 96
 - Creating the Syndicated Workflow • 94, 95
 - Introduction • 94
 - Procedure • 94
- Introduction to the Rhythmyx Enterprise Content Connector • 3
- Item Extractor Editor • 21, 22, 33

L

- Location Tab • 39
- Log View • 7, 15
- Logging • 57

M

- Maintaining Content Extractors • 11, 17, 31
- Maintaining Rhythmyx Enterprise Content Connector Descriptors • 4, 17
- Managing the Extractors Assigned to a Descriptor • 32
- Metadata View • 7, 14
- Migration Example
 - Creating a Descriptor • 76, 96
 - Introduction • 74
 - Reviewing an Upload • 92
 - Scanning and Uploading • 89
 - Setting up Rhythmyx Roles and Workflows • 74, 75
- Modifying the Content Tree • 65, 67

P

- Page Extractor • 35, 46
- Plugins

- Writing Content Selector Plugins • 118
- Writing Extractor Plugins • 119
- Writing Loader Plugins • 123
- Writing Transformer Plugins • 122

R

- Reconfiguring Workflows for Uploaded Content
 - 126, 127
- Rescanning • 71
- Reuploading • 71
- Rhythmyx Enterprise Content Connector
 - Functions • 4
 - Introduction • 3
- Rhythmyx User for Uploading • 126
- RhythmyxContentConnector.exe • 72
- Running Content Connector
 - Automatic Running • 72
 - Changing Content Types • 68
 - Excluding Items from a Run • 68
 - Interactive Running • 64
 - Methods of Running • 64
 - Uploading a Single Item • 68
 - Uploading Items in a Content Tree Folder • 69
 - Uploading New and Changed Content Only • 70
- Running The Enterprise Content Connector
 - Interactively • 64, 65, 68, 69, 72
- Running the Rhythmyx Content Connector • 63
- Running The Rhythmyx Enterprise Content Connector Automatically • 72

S

- Saving a Status File After Scanning • 70
- Saving Descriptors • 21, 22, 61
- Scanning and Uploading Again to Correct Errors
 - 67, 71
- Scanning for New and Changed Content • 64, 70
- Setting Up Rhythmyx to Use Content Connector
 - 125
- Setting User Preferences • 9, 16
- Site Migration Connector License • 74
- Starting the Rhythmyx Enterprise Content Connector • 5
- Static Image Extractor • 47
- Static Item Extractor Editor • 21, 22, 38
- Status File
 - Defined • 70
 - Saving • 70
- Status View • 7, 13

T

- The Enterprise Content Connector and Rhythmyx • 125
- The Menu Bar • 7, 8, 10
- The Rhythmyx Enterprise Content Connector Interface • 7
- The Toolbar • 7, 10
- Transformers
 - Configuring Field Transformers • 50
 - Configuring Item Transformers • 52
 - Defined • 49

U

- Uploading Selected Items and Folder Contents • 68
- Uploading Specific Items • 67, 68
- Uploading the Contents of a Content Tree Folder
 - 67, 69

W

- Workflow Tab • 37
- Workflows for Uploaded Content • 127
- Writing Content Selectors • 118
- Writing Extractors • 119
- Writing Loaders • 123
- Writing Transformers • 122

X

- XML Extractor • 35, 44
- XSL Extractor • 35, 43