Arco

Doma 6
Application administrator manual
03. Files

A new era in document management.

Doma 6

Application administrator manual

03. Files

Document reference: not yet defined
Described Doma version: 6.5.1

Author: Arco R&D - Ferdy Van den Broeck

Manual Modification Follow-Up

Ref.	Date	Author	Modification description	Affected paragraphs
	30/08/2012	Arco R&D - Ferdy Van den Broeck	Manual creation	
	06/09/2012	Kris Steenackers	Minor adjustments	

A digital version of this document is available on www.DocRoom.net.
Please send an e-mail to training@arco.be or info@arco.be to get a login for the system.

Remarks about this manual can be sent to training@arco.be.

Contents

1 Introduction 8
2 File extensions 9
2.1 General 9
2.2 File extension parameters 12
3 File Servers 14
3.1 General 14
3.2 File server Configuration 15
Next to the default installed file servers, one can create an implementation of a custom fileserver (see programming Guide) 16
3.3 Using a file server 16
4 Rendition Servers 17
4.1 General 17
4.2 Rendition server configuration 18
4.3 Using a rendition server 19
5 OCR 20

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

1 Introduction

Doma is able to store all kind of files in any category. Files will be linked to packages or to a document or a workflow.

Remark:

- **Please see the manual *Doma 6_Application Administrator Manual_04_Properties and Edit checks* for more information about properties.**
- **Please see the manual *Doma 6_Application Administrator Manual_05_Packages* for more information about packages.**
- **Please see the manual *Doma 6 - Application Administrator Manual - 06 - DocRoom Categories* for more information about categories.**

By default, files of the most common file types are already defined in the system. DocRoom allows to

- Add other file types.
- Configure quite a lot of options for the file types.
- Define the storage of the files on the DocRoom server (or other servers) using the 'file server' concept.
- Define the storage of various renditions of the files on the DocRoom server (or other servers) using the 'rendition server' concept.

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

2 File extensions

2.1 General

By default, files of the most common file types are already defined in the system.

DocRoom leaves all configuration possibilities open for any kind of file type. It is the responsibility of the system administrator to make sure the select the correct configuration for a new file type.

For every extension a number of options exist. It is possible to combine options that are not compatible.

The user can specify the DocRoom categories for which an extension should be allowed. By default all extensions are allowed for all categories. It is good practice to allow only a limited number of extensions for a category.

Two direct editing methods are supported by DocRoom: In-Line editing for ASCII and HTML files, WEBDAV for file types that are WEBDAV enabled.

In Line editing: for ASCII files and HTML documents, DocRoom offers an integrated editor allowing users to directly edit their documents in DocRoom.

With WEBDAV the users can edit directly from DocRoom. Once the user stops editing the file, this file will be saved to the actual DocRoom repository by the Application Server.

WEBDAV is by default enabled for office documents. Acrobat Professional Edition also supports the WEBDAV protocol. If you add new extension, DocRoom cannot guarantee that the associated applications support WEBDAV

By default, DocRoom can convert the following formats to PDF:

- Gif
- Jpg
- Wmf
- Png
- Bmp
- Tif
- Rtf
- Doc
- Docx
- Htm
- Csv
- Xlsx
- Xls
- Ppt
- Pptx
- Txt

HTML renditions can be created for following file formats:

- Csv
- Xls
- Xlsx
- Doc
- Docx
- Rtf

If this option is selected, for every file with this extension that is uploaded in DocRoom, a HTML rendition will be kept in DocRoom that is stored on the default Rendition Server

For following image formats, a thumbnail can be created in DocRoom:

- Bmp
- Gif
- Jpg
- Png
- Tif
- Image pdf's

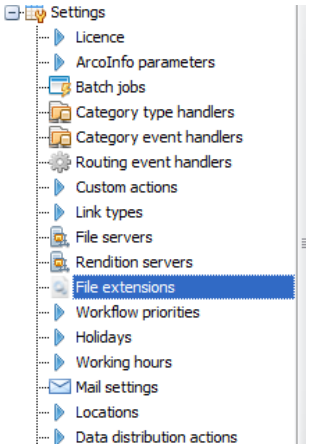
When the movie option is selected, it can be watched online (direct streaming) instead of first having to download the file to your pc before being able to watch the movie.

Remark: If a rendition is not selected at the time a file is uploaded, the file rendition will not be created afterwards, when the option is activated. However if the file is changed in DocRoom, the new set of selected options will be used to generate the appropriate renditions at the time of the change.

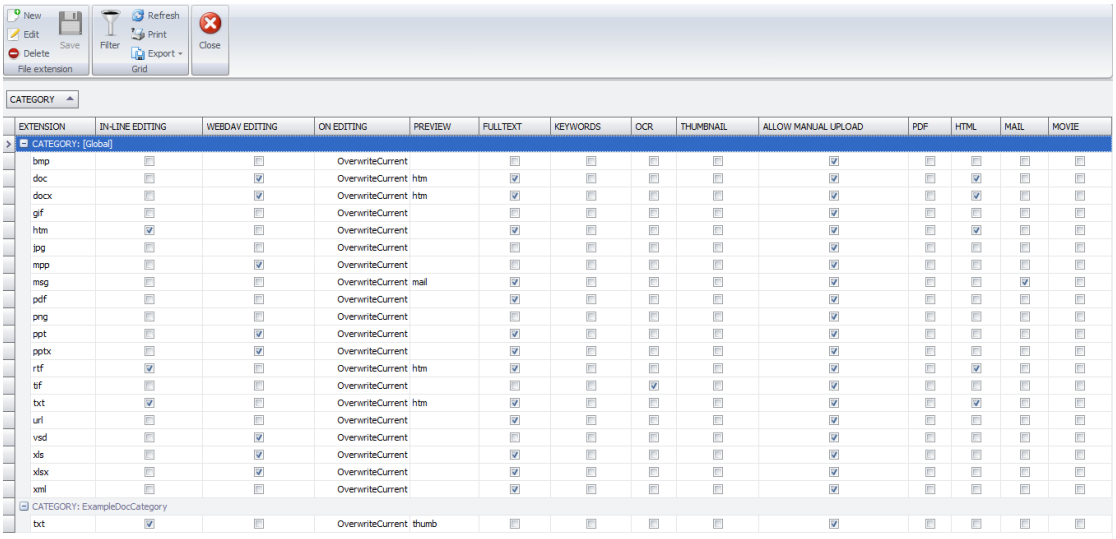
Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

The file extensions supported are defined in the 'Settings' → 'File Extensions' entry of the DomaAdmin tree.

File type that are not configured in this screen, will not be allowed in DocRoom. An error message will be generated when trying to add a file with an extension not defined here.



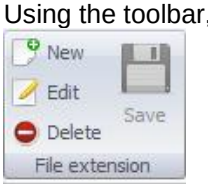
When opening the screen, an overview is given of the currently defined file types:



The first part of the screen CATEGORY [Global] gives the 'Global File Extensions'. These are the file extensions that are available for all DocRoom Categories (Documents, Folders, Dossiers, Workflow).

The lower part of the screen CATEGORY [*Categorynamel*] gives the 'File Extensions' that are available for the specific DocRoom Categories. These can be the same extensions as in the global definitions. In that case the files for this extension will have different configurations when added to this specific category type.

Using the toolbar, the user can



- Add new file extensions
- Edit existing file extensions

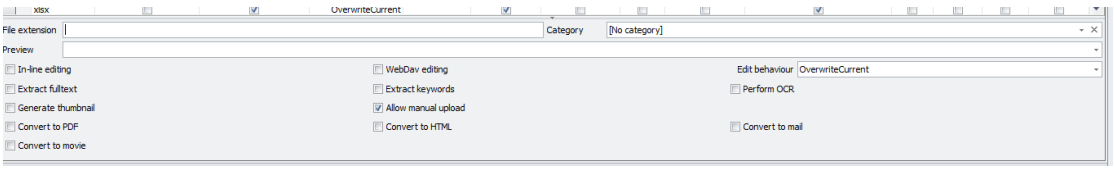
Omega Business Park
Wayenborgstraat 24 – B-2800 Mechelen
www.arco.be
info@arco.be
t +32 (0)15 289 030
f +32 (0)15 289 031

- Delete existing file extensions

2.2 File extension parameters

Following screen allows editing the parameters for a file type.

Remark: not all combinations are meaningful for all file types. Using meaningless combinations can result into an error.



Field name	Description
File extension	*The extension of the file being configured (TXT,JPG …)*
Category	*[No category]: global definition for all DocRoom categories* *Select the name of a category if specific configuration of the file type for this DocRoom category*
Preview	The rendition of the document when the document is viewed in the web interface (preview tab/clicking on the document file name) Native: the file is opened with the native program to view the file, for example Microsoft Word in case of .docx files PDF: PDF rendition is shown (the PDF option must be checked) HTML: HTML rendition is shown(the HTML option must be checked) Thumbnail: a thumbnail is shown(the thumbnail option must be checked) Download: the file will have to be downloaded Text: TXT rendition is shown, (the 'Extract fulltext' option must be checked) Remark: not all renditions are possible for all file types.
In Line editing	☑ *File will be edited inline.*
Extract Fulltext	☑ *DocRoom will try to extract the full text (and the language) for this file type, so content search will be available. The necessary filters for the full text extraction must be installed on the DocRoom server* ☐ No full text extraction for this file type
Generate thumbnail	☑ A thumbnail rendition of the file is created at indexation. Only for image file types ☐ No thumbnail rendition of the file is created
Convert to PDF	☑ A PDF rendition of the file is created at indexation ☐ No PDF rendition of the file is created

Convert to Movie	☑ A movie rendition of the file is created at indexation. ☐ No movie rendition of the file is created
WEBDAV Editing	☑ Inline editing of the file will be done using WEBDAV technology. WEBDAV must be enabled and configured in IIS ☐ Inline editing using native program **Remark:** The WEBDAV protocol is supported for Microsoft Office formats and Acrobat PDF (professional edition required). For a complete list see http://www.WEBDAV.org. **Remark:** If you want to allow users to edit documents over WEBDAV (only for WEBDAV enabled file types) a WEBDAV share has to be created. The installation of DocRoom creates the share by default.
Extract Keywords	☑ Keywords will be extracted at indexation (if possible) based in full text data. This option is only available if 'Extract fulltext' is selected ☐ No keywords will be extracted
Allow manual upload	☑ Files can be manually added by the user ☐ Files can only be added by DRUM (DocRoom Upload Module) or programmatically in event actions/handlers
Convert to HTML	☑ A HTML rendition of the file is created at indexation ☐ No HTML rendition of the file is created
Perform OCR	☑ OCR (Optical Character Recognition) is performed on the file. The OCR module of DocRoom must be installed and configured. For .TIF, PDF file types ☐ No OCR is done
Convert to Mail	☑ Rendition to internal DocRoom mail is generated. Only for .MSG files ☐ No rendition to internal DocRoom mail is generated
Edit behaviour	Overwritecurrent: current file will be overwritten CreateNewMinor: a new minor version will be created when editing CreateNewmajor: : a new major version will be created when editing

3 File Servers

3.1 General

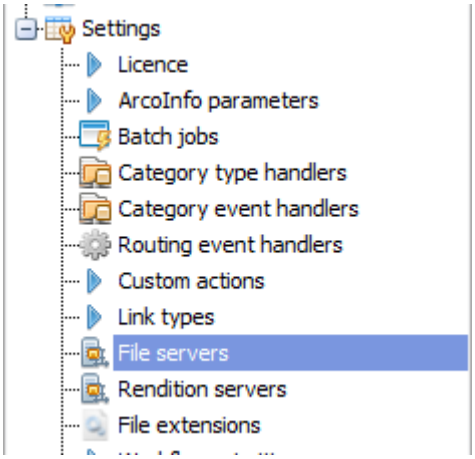
File servers define the implementation of the storage of files within a DocRoom implementation.

DocRoom can store user files (DocRoom attachments) on a file server or in a Binary Large OBject (BLOB) inside the database.

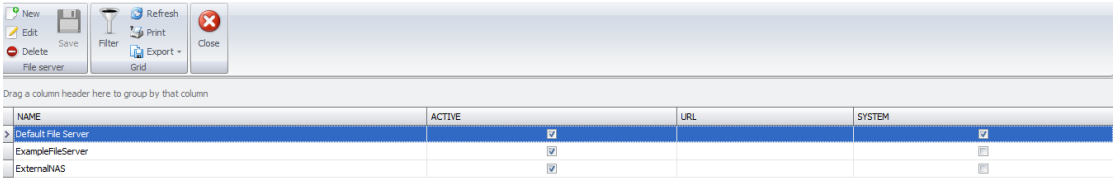
By default DocRoom defines a Default File Server in the [ROOT]\Data\FileServer.

This file server defines the location of the documents stored in DocRoom and can be configured as a File Data Store (using UNC to access the documents) or as a BLOB Data Store (in this case the Database connectivity is used to store and retrieve files).

The file servers are defined in the 'Settings' → 'File Servers entry of the DomaAdmin tree.



Opening the screen will give an overview of the existing file servers:



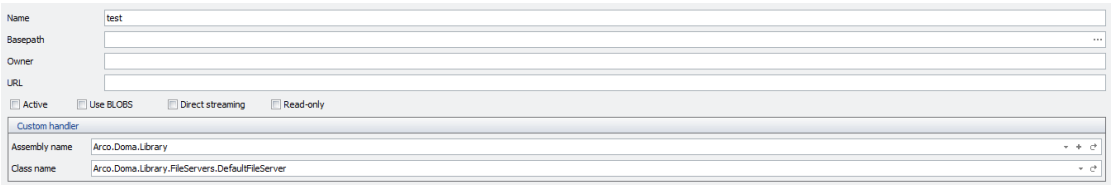
Using the toolbar, the user can

- Add new file servers
- Edit existing file servers
- Delete existing file servers

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

3.2 File server Configuration

The following can be configured:



General

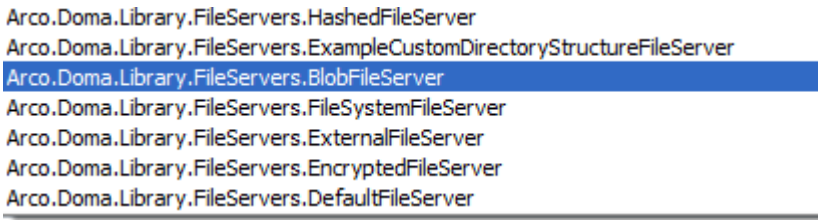
Field name	Description
Name	*The name of the file server*
Basepath	*Enter the base path on the server where the files are physically stored.* *Example: D:\data\DR6_SQL2012\Data\ExampleFileServer* *Example: [DATADIR]fileserver*
Owner	Obsolete
URL	*NA for file servers*
Active	Obsolete
Use BLOBS	☑ The file is not stored on a physical disc, but stored in the database. ☐ The file is stored on a physical disc, in the directory defined in Basepath Only applicable for the DefaultFileServer Always enabled for BlobFileServer
Direct Streaming	☑ File content is streamed directly to the user (without using the application server service). Security of the application pool in IIS is used ☐ File content is streamed using the application server service. Security of the account running the application server service is used
Read Only	☑ Files can only be read from this file server, no files can be added or modified ☐ Files can be added, deleted, modified

Custom handler

Field name	Description
Assembly Name	*Enter the name of the .net assembly dll where the code of the file server is defined.* *Default: Arco.Doma.Library*
Class Name	*Enter the name of the .net assembly class where the code of the file server is defined.* *Default: Arco.Doma.Library.FileServers.DefaultFileServer*

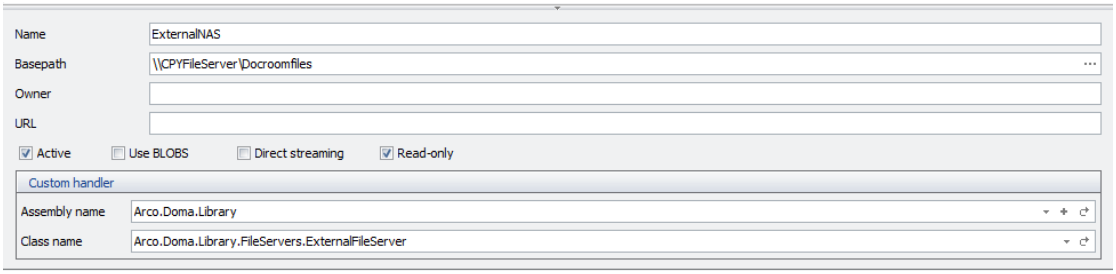
The following fileserver definitions are standard available.

Omega Business Park
Wayenborgstraat 24 – B-2800 Mechelen
www.arco.be
info@arco.be
t +32 (0)15 289 030
f +32 (0)15 289 031



- HashedFileServer: the file server will try to store a file only once, even when it is uploaded twice in the system (using the hash of the file)
- BlobFileServer = default file server, blobs enabled
- FileSystemFileServer = default file server, blobs disabled
- ExternalFileServer: used when files are stored on a different server.
- EncryptedFileServer: files are encrypted before they are stored (RIJNDAL algorithm)
- DefaultFileServer

Example of an External file server definition:

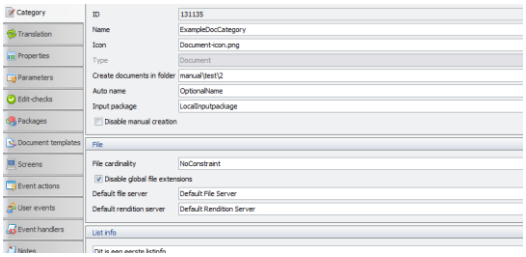


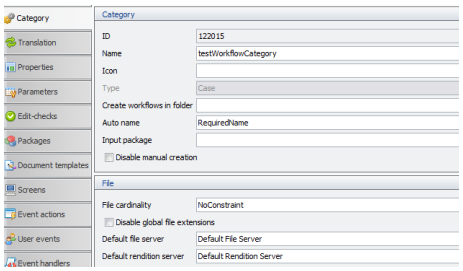
Next to the default installed file servers, one can create an implementation of a custom fileserver (see programming Guide)

3.3 Using a file server

The file servers can be selected in the definition of a

- Document category
- Workflow category





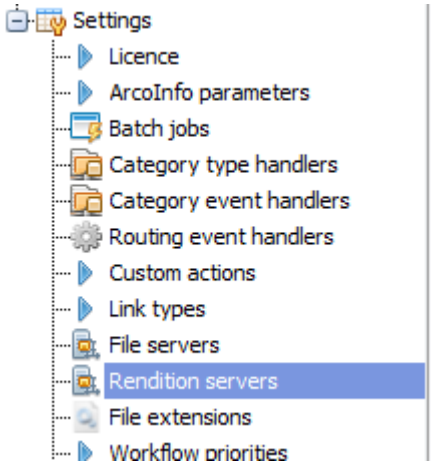
4 Rendition Servers

4.1 General

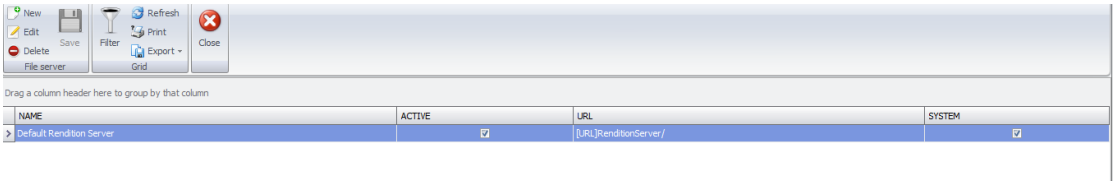
Rendition servers define the implementation of the storage of the various renditions of the files within a DocRoom implementation.

By default DocRoom installs the Default Rendition Server: This rendition server defines the location of the generated renditions. This server is always defined as a File Data Store.

The rendition servers are defined in the 'Settings' → 'Rendition Servers entry of the DomaAdmin tree.



Opening the screen will give an overview of the existing rendition servers:

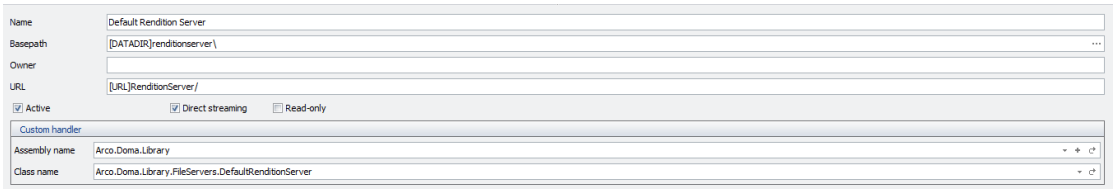


Using the toolbar, the user can

- Add new rendition servers
- Edit existing rendition servers
- Delete existing rendition servers

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

4.2 Rendition server configuration



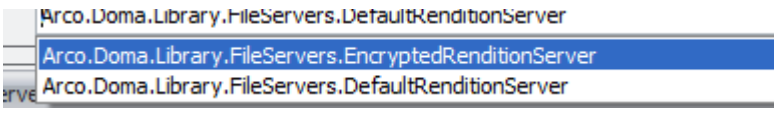
General

Field name	Description
Name	*The name of the rendition server*
Basepath	*Enter the base path on the server where the files are physically stored.* *Example:* *D:\data\DR6_SQL2012\Data\ExampleFRenditionServer* *Example: [DATADIR]renditionserver*
Owner	obsolete
URL	*Default: [URL]RenditionServer/* *The Virtual folder in IIS where the renditions are accessible*
Active	obsolete
Direct Streaming	☑ File content is streamed directly to the user (without using the application server service). Security of the application pool in IIS is used ☐ File content is streamed using the application server service. Security of the account running the application server service is used
Read Only	Not applicable for rendition server

Custom handler

Field name	Description
Assembly Name	*Enter the name of the .net assembly dll where the code of the rendition server is defined.* *Default: Arco.Doma.Library*
Class Name	*Enter the name of the .net assembly class where the code of the rendition server is defined.* *Default: Arco.Doma.Library.FileServers.DefaultRenditionServer*

The following fileserver definitions are standard available.

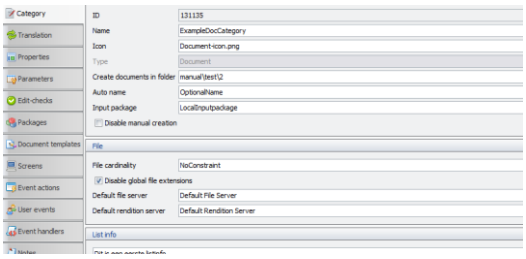


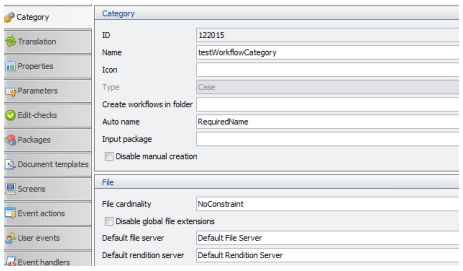
- EncryptedRenditionServer: file renditions are encrypted and then stored (RIJNDAL algorithm)
- DefaultRenditionServer

4.3 Using a rendition server

The rendition servers can be selected in the definition of a

- Document category
- Workflow category





Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031

5 OCR

DocRoom gives the possibility for OCR scanned images, in order to retrieve the text content from .tif, .pdf files.

In order to do this, ABBYY Finereader is installed with the system.

Every time a user (or process) adds a file with a file extension that is flagged for OCR the file will be flagged for OCR (OCR_STATUS = 1, IDX_STATUS = 1).

Every x minutes the batch job „*SendFilesToOCR*" exports the file to the ARCOINFO .ocr.Inputdir (by default [ROOT]\data\OCR\InputFromDoma) folder and adds a record to the OCR queue (dm_object_file_ocr).

The ocr status of this files in the OCR queue = 1.

The „*OCR service*" (By default DomaOCRService) polls on the OCR queue for files with ocr_status = 1. Every file with this status is moved to the IN subdirectory of the OCR processing (by default [ROOT]\data\OCR\processing\in).

After a successful OCR the OCR status is changed to 0. The newly created files (PDF and TXT file) are placed in the OUT subdirectory of the OCR processing (by default [ROOT]\data\OCR\processing\out).

If an error occurred in the OCR process, the OCRSTATUS of the file becomes 9 and the error is logged in the OCR queue.

Every x minutes the batch job „*CollectFilesFromOCR*" polls the OCR queue for files that were successfully ocred. The text rendition is uploaded in the fulltext zone of the file and the ocr"ed pdf result is placed back in DocRoom and **replaces** the original version (unless the arcoinfo parameter ocr.keeptiff=1, then the original tiff file is kept).

After the upload to DocRoom, the temporary files are deleted from the OCR processing directory and the ocr status is updated to 2 in the ocr queue.

Remark: If a file extension is selected for OCR, every file that is uploaded will be sent to the OCR queue. PDF files that contain text will not be sent to the OCR process if DocRoom can extract the full text successfully.

Remark: When a file is still in the OCR queue, it cannot be deleted in DocRoom. At that time, a user can only download the file and change the title of the file.

Omega Business Park www.arco.be t +32 (0)15 289 030
Wayenborgstraat 24 – B-2800 Mechelen info@arco.be f +32 (0)15 289 031