

DocuNECT™ Capture

Technical Product Overview

This guide provides an overview of the DocuNECT™ v4.2 Product, a solution for Electronic Document Capture.

May 16, 2010 – Release 2.3



www.portfordsolutions.com

Contents

DocuNECT™ Overview	3
The DocuNECT™ Document CaptureLifecycle.....	4
Capturing Documents from the Desktop.....	6
DocuNECT Virtual Print Driver	6
Frequently Asked Questions	8
DocuNECT™ Desktop Capture	9
Frequently Asked Questions	10
DocuNECT™ Microsoft Windows and Outlook Integration	10
Frequently Asked Questions	10
Capturing Documents from Other Sources.....	11
Email Integration.....	11
Network File Locations	11
FTP, sFTP or SSH.....	11
Web Sites.....	12
Multi-Function Device Integration	12
EMC Documentum ApplicationXtender or D6	12
Indexing Documents from the Desktop.....	13
Automating Document Indexing	16
Distributing Documents and Data	19
Developing with DocuNECT™	19

DocuNECT™ Overview

DocuNECT™ is an electronic document capture, indexing and distribution software platform which provides a central hub for electronic document capture. DocuNECT™ allows content to be captured, processed to retrieve index values, and then distributed into the target *Enterprise Document Management System* (EDM) or business application.

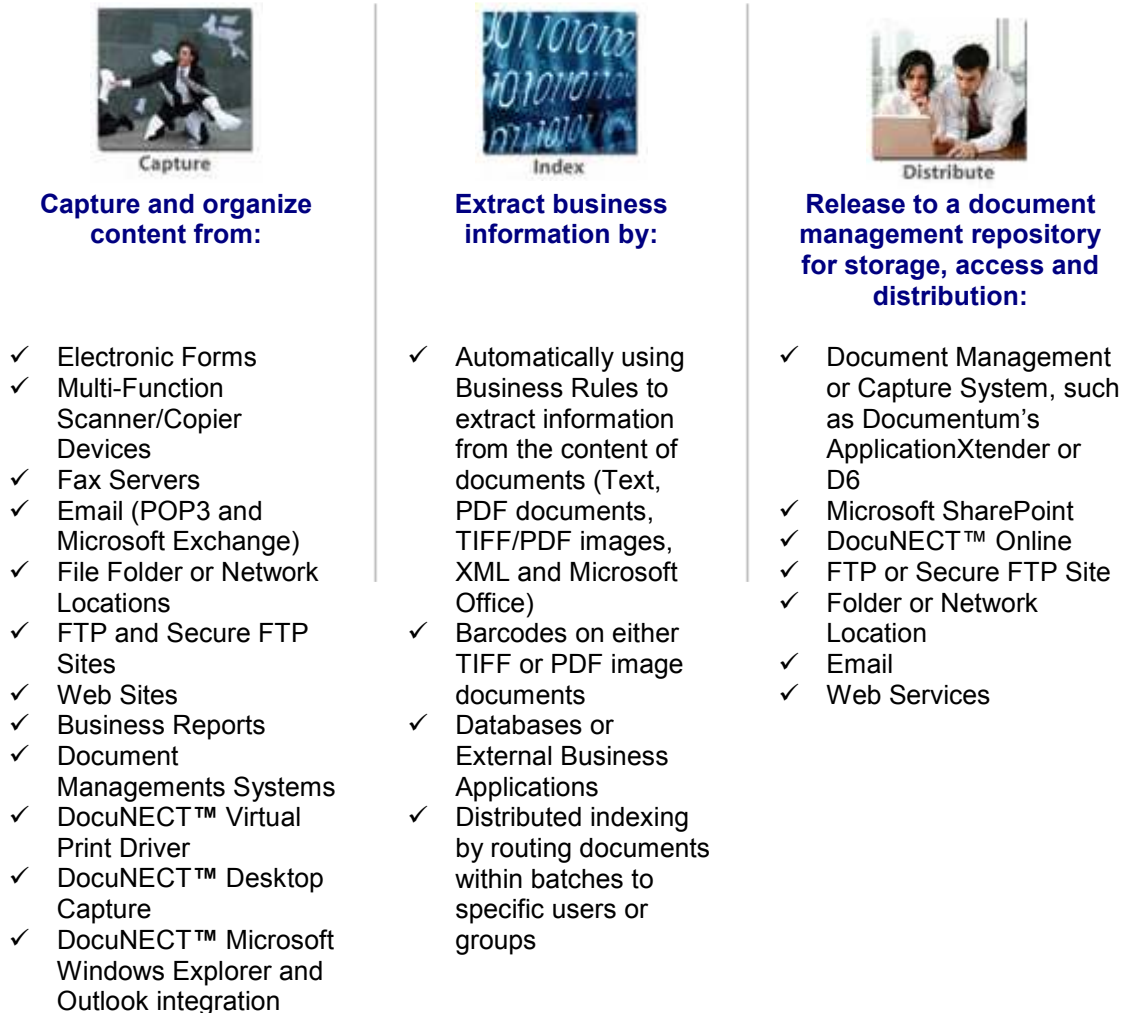
This document provides a complete overview of the product and is broken down into the following sections:

-  **The DocuNECT™ CaptureLifecycle**
-  **Capturing Documents from the Desktop**
-  **Capturing Documents from Other Sources**
-  **Indexing Documents from the Desktop**
-  **Automating Document Indexing**
-  **Distributing Documents and Data**
-  **Developing with DocuNECT™**

The DocuNECT™ Document Capture Lifecycle

At the heart of the DocuNECT™ capture technology is the *CaptureLifecycle* that contains the business rules for capturing, indexing and distributing documents. The *CaptureLifecycle* is a set of rules that is applied to a batch of document during the capture, indexing and distribution process.

The following diagram provides a summary of the three basic stages of the *CaptureLifecycle* that define how and where the documents are capture from, how the index data is extracted and where they are finally distributed to:



Capture Lifecycles are created using an editor through which business rules can be applied to the capture process. The editor allows an administrator to easily create capture lifecycles, and includes a high-level scripting language called DocScript to accommodate complex capture processes and business rules.

DocuNECT looks for data the same way we do. When we look at documents for reference data we search for particular keywords. For example, if we pick up an invoice we search for the word *Invoice No.* and then look at the associated value. The Capture Lifecycle work exactly the same way. Using this method the rules can work with any type of document whether it is an image that has been scanned, a PDF document, a Microsoft Office document, an email or text document.

While this search technology works well with unstructured textual documents, DocuNECT™ can also work with specific document structures. Microsoft Excel spreadsheets can reference data by columns and rows, PDF forms have form field names that can be referenced, and XML files have tags that can be read. The editor also allows document templates to be created that describe a particular document type. Template rules can be applied to automatically separate documents and extract data. A good application for capture lifecycle templates is invoice processing as the vendor information can be added as a template for high-volume invoices to reduce the time it takes to index documents.

Q – Can I mix document formats in a batch?

A – *Yes, the rules can be trained to detect different document formats and apply different rules accordingly. This is useful if you are capturing emails where the attachments often range in format.*

Q – Can I ignore documents in a batch?

A – *Yes, again the rules can be defined to ignore certain types of document based on their format or content. A good application for this feature would be to ignore emails being sent to a central email inbox that do meet specific criteria.*

Q – Can the rules send the different document to different users or group for indexing?

A – *Yes, there may be a batch of documents that require different expertise and rules can be created to send documents to those users who have the expertise based on the type.*

Q – What if the rules fail?

A – *If the rules fail then an email notification is sent out and the batch goes into an error status for the error to be resolved.*

Q – You mentioned scripting, do I need to be a developer?

A – *No, the script is very high-level and is there to provide a high-degree of flexibility to manage different capture scenarios.*

Capturing Documents from the Desktop

DocuNECT™ provides the following methods for capturing documents from the desktop:

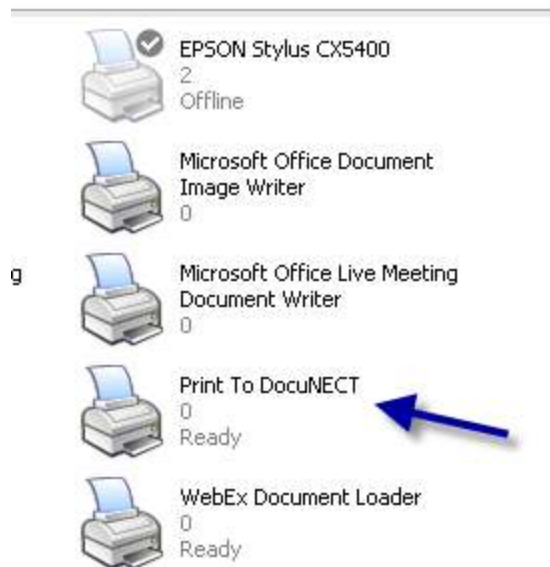
- DocuNECT™ Virtual Print Driver
- DocuNECT™ Desktop Capture
- DocuNECT™ Microsoft Windows and Outlook Integration

DocuNECT Virtual Print Driver

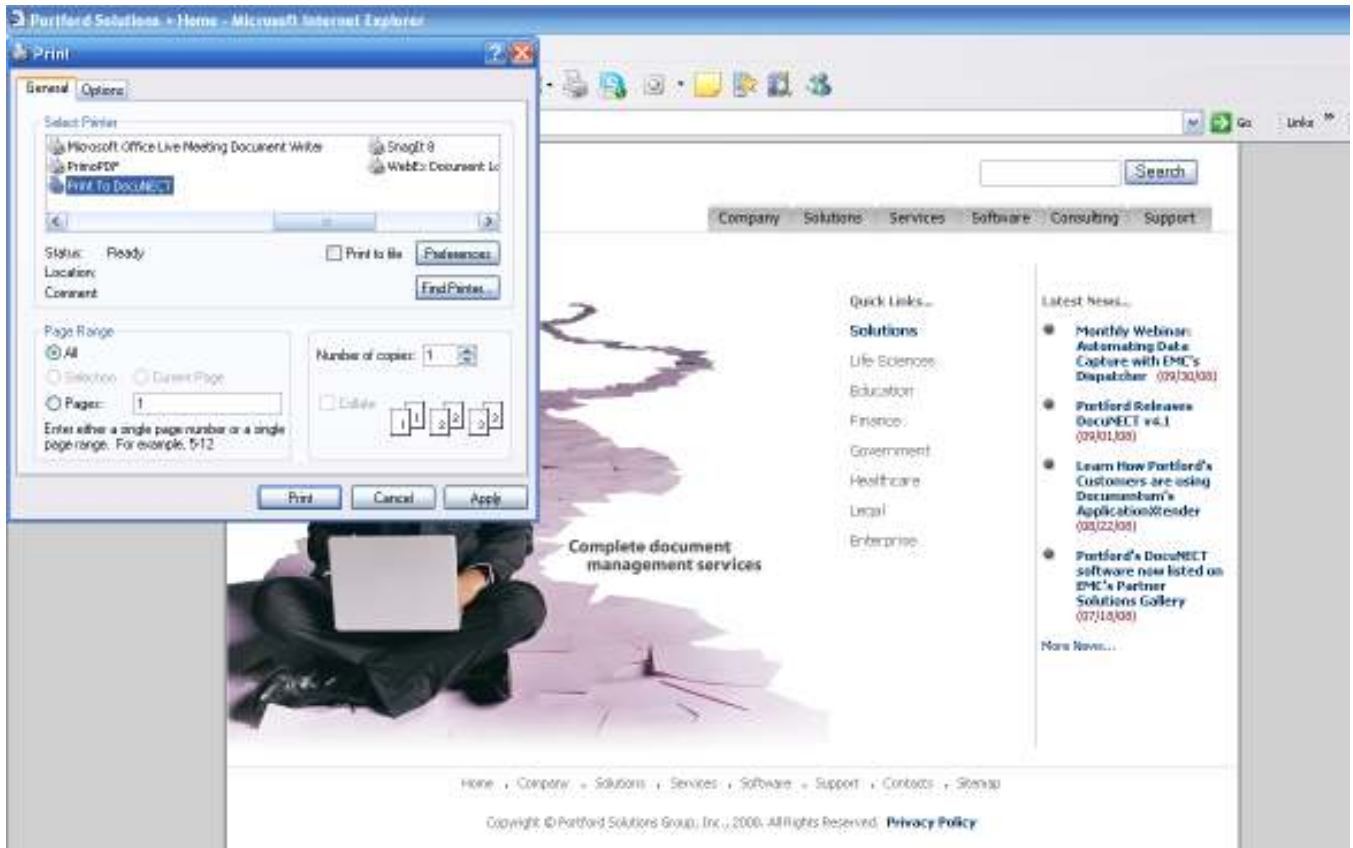
The DocuNECT™ Virtual Print Driver (VPD) is a great way to capture content from business applications. The VPD works exactly the same way as a regular print driver, except the document is directed to the DocuNECT™ system.

Depending on the business rules defined in the *CaptureLifecycle* definition, the user either indexes the document from the desktop or the document is sent to the DocuNECT™ server for automatic indexing.

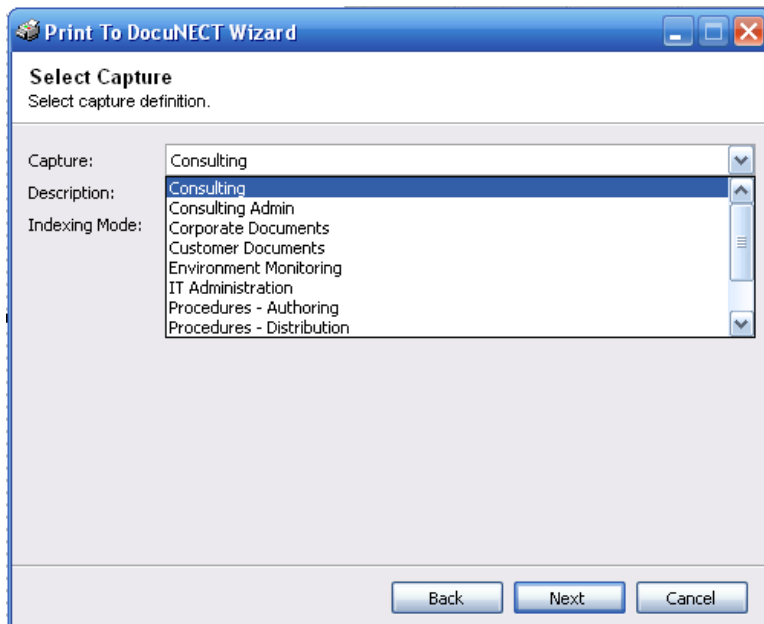
As shown in the screenshot below the VPD is listed as a Microsoft Windows printer. Any application has the ability to print can use the VPD.



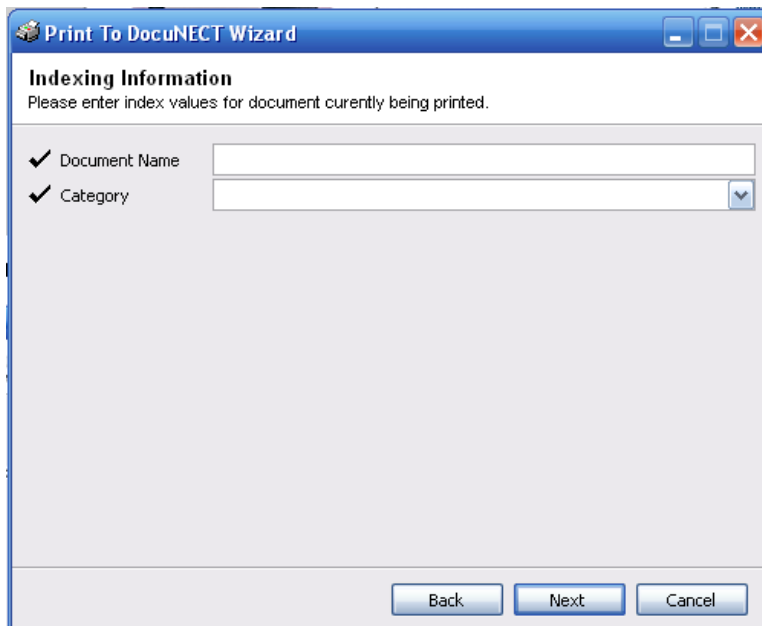
The following screenshot shows a Microsoft Internet Explorer web page being printed into DocuNECT™.



Once print, the VPD wizard is displayed prompting for a definition to print to:



Depending on how the definition is setup the user will be prompted for the associated index values.



The screenshot shows a Windows-style dialog box titled "Print To DocuNECT Wizard". The main heading is "Indexing Information" with a sub-instruction: "Please enter index values for document currenly being printed." There are two checked items: "Document Name" with an empty text input field, and "Category" with a dropdown menu. At the bottom, there are three buttons: "Back", "Next", and "Cancel".

Frequently Asked Questions

Q - Once printed, what format is the document when appears in the Document Management System?

A - *Capture definitions can be configured to create either Adobe PDF or TIFF files. If index values match an existing record in the system, VPD can replace or append that record with the additional printed page(s).*

Q - Do I have to enter the index values manually?

A - *Again this depends on the how the capture definition is setup, but each definition can be setup to index from within the VPD wizard or automatically on the server using business rules. If business rules have been defined to extract information automatically then documents can be printed directly into this process.*

Q – What if documents are being printed external to my firewall?

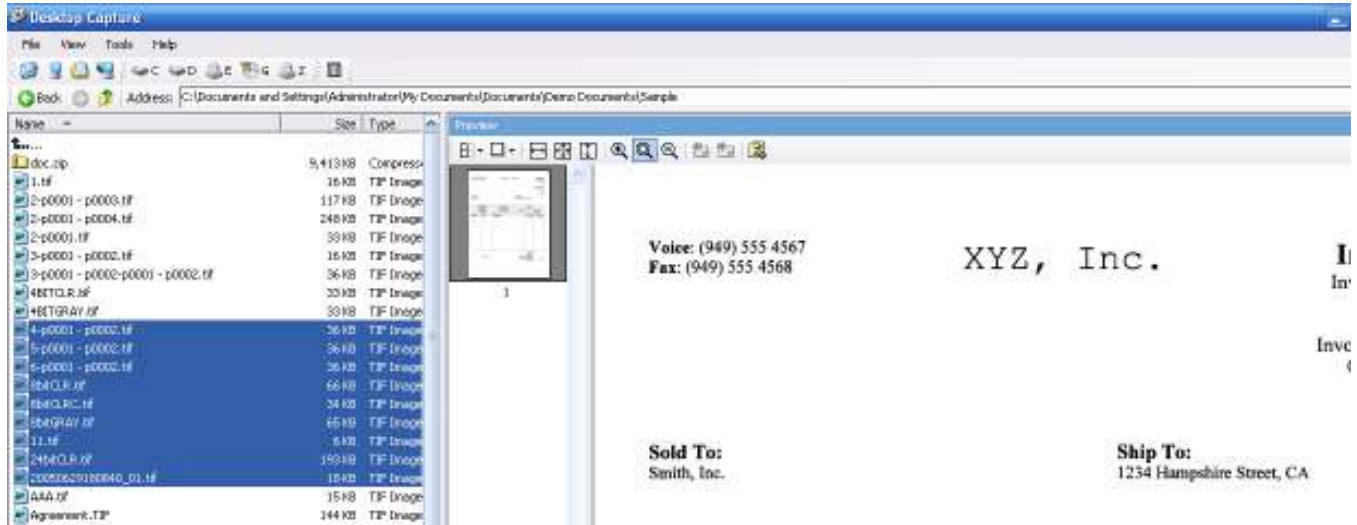
A – *All DocuNECT™ Desktop software use web services over HTTP or HTTPS, which makes it easier to facilitate document uploads outside the firewall.*

Q – Can I use the VPD in a Citrix (www.citrix.com) or Provision Networks (<http://www.provisionnetworks.com/>) environments?

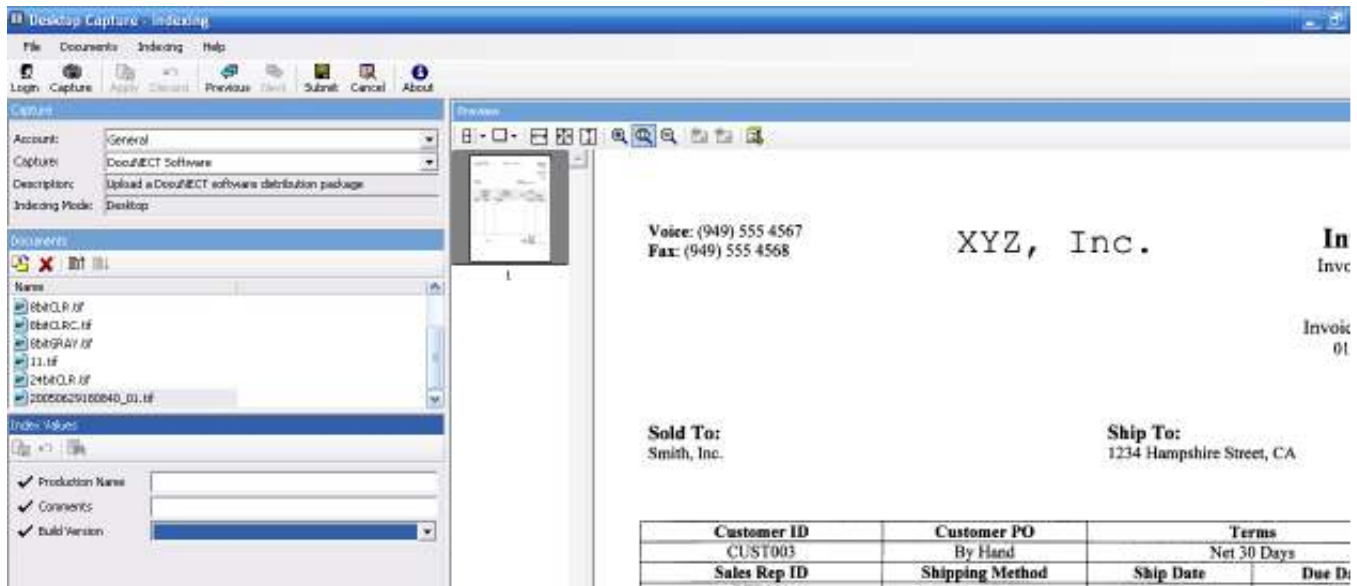
A – *Yes, the VPD is compatible with both environments.*

DocuNECT™ Desktop Capture

The DocuNECT™ Desktop Capture application allows for electronic documents to be uploaded for batch processing. The application has a built in file browser and image viewer to make it easier to search the desktop for documents to upload:



Multiple documents can be selected and uploaded in a batch and indexed before being uploaded to DocuNECT™.



Frequently Asked Questions

Q – Can I upload any document using the Desktop Capture?

A – Yes, the browser views image formats (TIFF, BMP, PNG, JPEG, and GIF), Adobe PDF and Microsoft Office, however, any document can be uploaded and stored in its native format.

Q – What if I have an image file that I want to break up into separate documents for indexing?

A – The application has the ability to break up one TIFF documents into separate documents.

DocuNECT™ Microsoft Windows and Outlook Integration

Documents can be capture directly from Microsoft Windows Explorer or Microsoft Outlook. This functionality is part of the Desktop Capture application.

Frequently Asked Questions

Q – Can you drag and drop more than one document at a time?

A – Yes, you can select multiple files from the desktop and then drag the files to the drop zone.

Q – Can you drag and drop emails and attachments?

A – Yes, you can drag and drop emails and associated attachments.

Capturing Documents from Other Sources

In addition to capturing documents from the desktop, DocuNECT™ can capture documents from “server-side” sources such as:

- Email (either POP3, IMAP or Microsoft Exchange)
- Network File Locations
- FTP or Secure FTP, or SSH
- Web Sites
- Multi-Function Devices
- EMC Documentum ApplicationXtender or D6

Email Integration

This functionality allows for a central email inbox being setup as an input to a business process. For example, an email address such as invoices@company.com can be setup to allow vendors to automatically email documents directly into the document management system, or into an indexing queue depending on the business rules that have been setup.

The rules allow for the email body and/or attachments to be captured along with the email metadata such as the *From, To, Cc* etc.

Q – What if I have Lotus Notes or Groupwise email?

A – *Although DocuNECT™ has native integration with Microsoft Exchange it can also work by using the POP3 or IMAP protocols which make this functionality accessible to other email systems such as Lotus Notes or Groupwise.*

Q – Can I send back an email confirmation?

A – *Yes, an email confirmation can be sent back to the originator confirming receipt.*

Q – Can I process the attachments and the body of the email?

A – *Yes, this is configurable to capture the body of the email, the attachments, or both.*

Network File Locations

This feature monitors a folder structure and will import the documents it finds. Business rules such as folder and file naming can be applied to automatically assign index data.

This process is often used for document migrations as it can apply business rules to an entire network share and import the documents into a document management system.

Q – Can I determine what documents are processed?

A – *Yes, the rules can determine which file types to include or ignore.*

FTP, sFTP or SSH

Documents can be captured from an FTP site, or a Secure FTP site. SSH can also be used via a product called Bitvise (www.bitvise.com). By using the SSH and sFTP protocols makes it an ideal process for secure transfer of documents between organizations.

Q – Can I apply the same rules as the network share?

A – *Yes, the system sees an FTP site and essentially another drive so the same business rules can be applied.*

Web Sites

Documents hosted on a web site can be capture via a URL which is useful for tracking marketing information or importing documents only available via a web link.

Q – How does it know when there is a new document to import?

A – *HTML pages include a time updated flag that indicates the last time the document was imported. Alternatively the file size can be compared if the host HTML page does is not utilizing this feature.*

Multi-Function Device Integration

Organizations are utilizing their existing copiers as scanning devices. These devices allow documents to be scanned and, in some cases, basic indexing to be applied. Typically, the interface on a multi-function device allows user to select a document type and enter an index value.

DocuNECT™ can import images scanned on a multi-function device and process them. DocuNECT™ has the following image processing features:

- Process either TIFF or PDF image documents
- Document separation by using a patch code, barcode or contextual value
- Image cleanup including despeckle and deskew
- Blank page removal
- Automatic rotation
- Barcode reading, including: Patch Code, Code 11, Code 39, Code 128, CodaBar, Inter2of5, EAN13, EAN8, UPCE, PDF417, Data Matrix, QR Code
- Image compression, using page analysis to determine the most appropriate compression algorithm for both PDF and TIFF images
- Supported PDF image compression: JBIG, JPEG, and PNG
- Supported TIFF image compression: Group 3, Group 4, JPEG, JPEG Type 7, CCITT RLE, LZW, Pack Bits, and Deflate
- OCR engine for text extraction. This works for both PDF and TIFF images that can produce Microsoft Word, HTML, Text, Rich Text Format (RTF) and PDF image plus text with text underlay

If the multi-function device does not have the ability to identify the user then DocuNECT™ has a barcode generation utility that generates the barcode based on the user's current Active Directory login. This can be used to separate scanned documents and lookup the users email so a notification can be sent with a URL link to the batch to be indexed.

EMC Documentum ApplicationXtender or D6

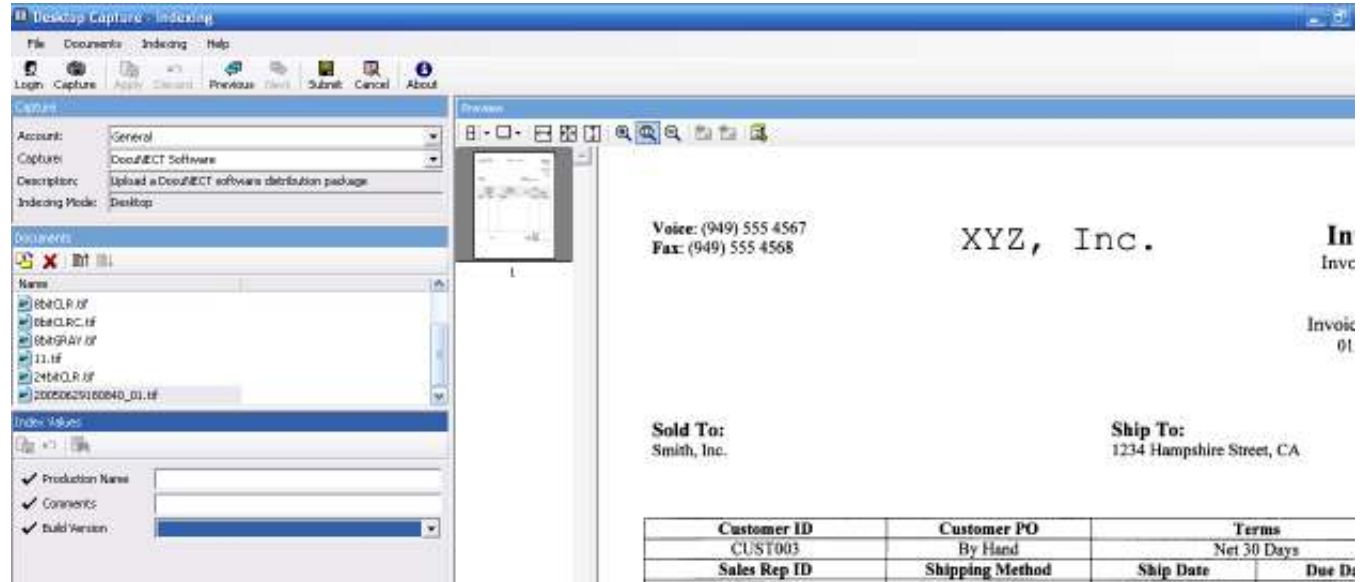
EMC's document management systems can be defined as a capture source to distribute documents to an external source, such as a network share or a FTP site.

Q – Can I determine which documents to export?

A – *Yes, a list of search expressions can be determined that identify a distinct group of documents to be exported.*

Indexing Documents from the Desktop

In a capture lifecycle, documents can either be indexed at the desktop (point of capture), in batch through a web application, or automatically by applying business rules. The screenshot below shows indexing from the Desktop Capture application:



The web based batch indexing has a more powerful interface as it is designed for high-volume indexing and data verification.



Inbox

This page displays the cabinets that you are allowed to access documents from.

ID	Name	Description	My Batches	All Batches
00001	Accounts Payable	This cabinet manages the capture lifecycle for Accounts Payable documents	5	5
00005	Contract Documents	Extracts data from legal contract	0	0
00003	Corporate Documents	This cabinet manages documents from a File-Share migration	0	0
00004	Human Resource	Manages Human Resource documents imported from electronic forms	0	0
00006	Loan Data	Manages Loan Data and Documents	0	0
00002	Loan Processing	This cabinet manages the capture lifecycle for Loan Processing documents	5	6

The following screenshot shows the DocuNECT batch management feature which summarizes the different batches going through the *CaptureLifecycle* process.



[Inbox](#)

View Batches: Accounts Payable

This page displays the batches containing documents assigned to you.

Search

Batch Status: (all) ▾
 Assigned To: (me or not assigned) ▾
 Locked By: (me or not locked) ▾

Results

This section displays the batches matching the above filter.

<input type="checkbox"/>	<input type="checkbox"/>	Batch ID	Status	My Docs	All Docs	Total Size	Created At	/		
<input type="checkbox"/>	<input type="checkbox"/>	00037	Indexed	5 / 5	5 / 5	61.00 KB	3/2/2010 8:44:46 AM		Index	Verify
<input type="checkbox"/>	<input type="checkbox"/>	00036	Indexed	2 / 2	2 / 2	5.12 MB	3/2/2010 8:43:46 AM		Index	Verify
<input type="checkbox"/>	<input type="checkbox"/>	00031	Indexed	1 / 1	1 / 1	92.84 KB	3/1/2010 1:52:52 PM		Index	Verify
<input type="checkbox"/>	<input type="checkbox"/>	00026	Created	1 / 1	1 / 1	46.50 KB	1/26/2010 9:36:54 AM		Index	Verify
<input type="checkbox"/>	<input type="checkbox"/>	00008	Indexed	3 / 3	3 / 3	196.18 KB	1/25/2010 4:51:45 PM		Index	Verify
<input type="checkbox"/>	<input type="checkbox"/>	00004	Indexed	3 / 3	3 / 3	107.79 KB	1/25/2010 3:19:58 PM		Index	Verify

The batch summary displays a tabular view of the document and data in the batch.

View Batch: 00008

The page displays the batch properties and depending documents.

Batch Properties

This section displays the batch properties.

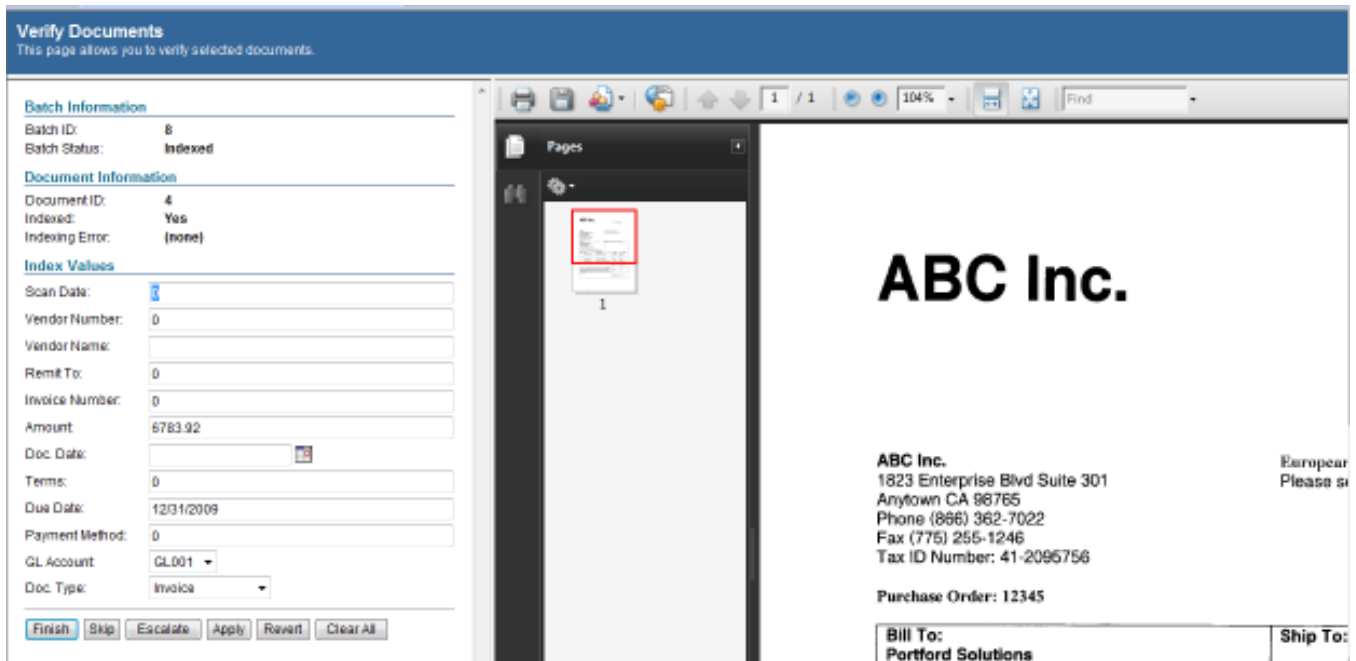
Batch ID: 00008
 Status: Indexed
 Cabinet: [Accounts Payable](#)
 Author: [cmssadmin](#)
 Storage: [General Storage](#)
 Date Created: 1/25/2010 4:51:45 PM
 Docs Count: 3
 Total Size: 196.18 KB
 Locked By: -

Documents

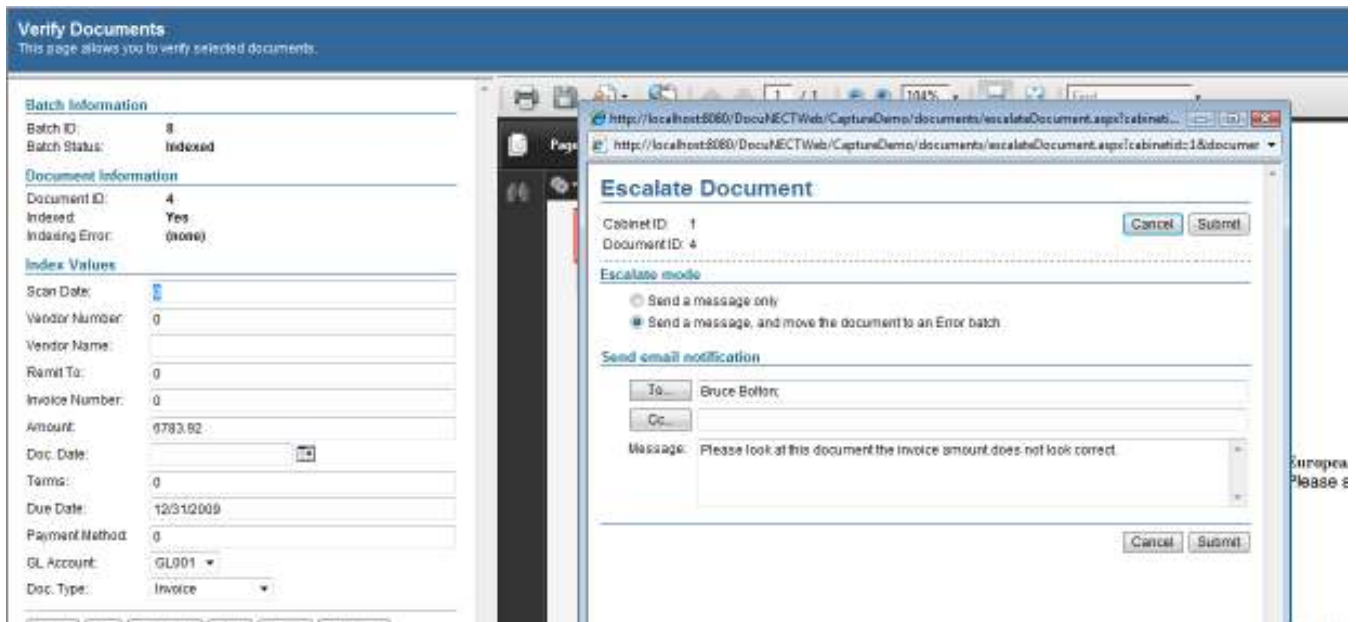
This section displays the documents assigned to you or not assigned.

<input type="checkbox"/>	<input type="checkbox"/>	Document ID	Indexed	Scan Date	Vendor Number	Vendor Name	Remit To	Invoice Number	Amount	Doc. Date	Terms	Due Date	Payment Method	GL Account	Doc. Type
<input type="checkbox"/>	<input type="checkbox"/>	00004	Yes	0	0		0	0	6783.92	0	0	12/31/2009	0	GL001	Invoice
<input type="checkbox"/>	<input type="checkbox"/>	00005	Yes	0	0		0	0	1740	0	0	1/13/2010	0	GL001	Invoice
<input type="checkbox"/>	<input type="checkbox"/>	00006	Yes	0	0		0	0	3000	0	0	1/21/2010	0	GL001	Invoice

The indexing module allows quick data entry and verification of the document.



The indexers have the ability to escalate the document to another user, or move the document into an error batch.



The recipients of the escalation will receive an email with a link to the batch and the actual document. On clicking on the link they will be able to review the error and resolve the escalation.

Qty	Item	Description	Each	Total
1	ABCASSTO-001	ABC Server Software	\$2,480.00	\$4,880.00
5	ABCNYTO-001	ABC Client Software	\$160.00	\$780.00
1	ABCMAINT-001	Annual Software Maintenance and Support	\$1,004.00	\$1,004.00

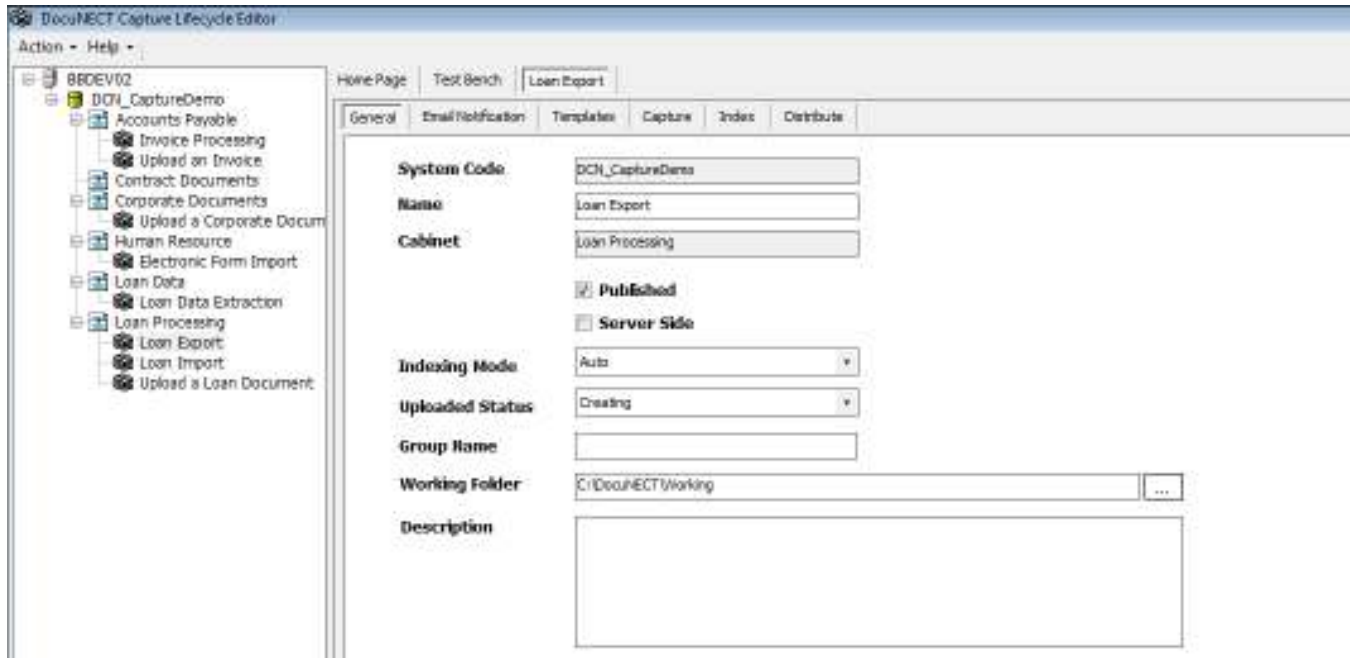
Automating Document Indexing

There may be situations where documents can be automatically indexed to reduce the labor involved in processing documents. DocuNECT™ supports the following automated indexing features:

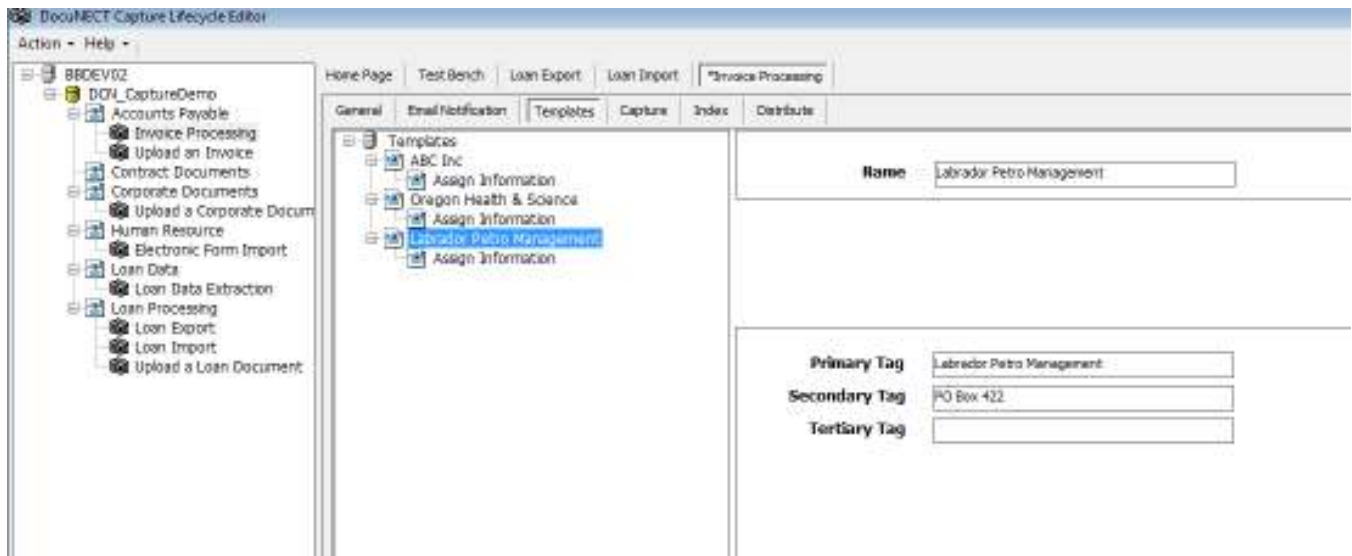
- Extract the content of a PDF or TIFF image by using the DocuNECT™ OCR server.
- Extract the content of PDF text document, Microsoft Word, HTML, XML and Text documents using business rules. The rules can use search tags or regular expressions to extract specific data from documents and assign them index values
- Extract content from Microsoft Excel document by referencing rows and columns
- Extract content from PDF forms by referencing form fields
- Read the following barcode formats Patch Code, Code 11, Code 39, Code 128, CodaBar, Inter2of5, EAN13, EAN8, UPCE, PDF417, Data Matrix, QR Code
- Database lookup and processing functionality to retrieve data from external databases
- Converts PostScript, TIFF and Microsoft Office to PDF
- Converts PDF to TIFF
- Indexes from an external reference file. If you receive documents with index data in a text or XML file for example then rules can be defined to import this format
- Documents can be split by rules. For example, a text or PDF business report can be split based on the occurrence or change in value such as a department or project code.

These supported features combine traditional report management and image processing into electronic document capture.

The *CaptureLifecycle Editor* helps users to create and manage the *CaptureLifecycles*, including the data extraction rules.

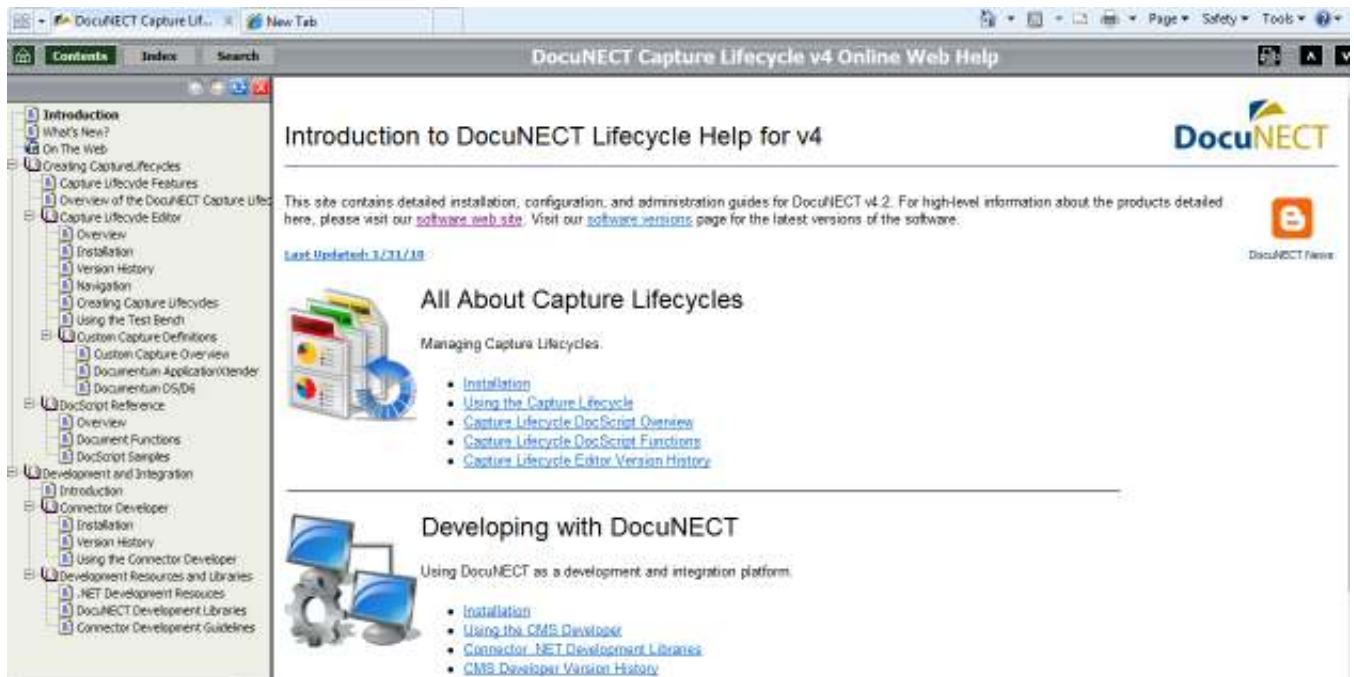


For more complex scenarios, the DocuNECT™ DocScript scripting language can be used to specify the rules. This is an easy to use scripting language that does not require development experience. Templates can also be applied to identify different types of document which may have a different set of business rules.



For example, a loan package may require different data to be extracted for the different types of loan documents and invoices may have different rules for each vendor.

There is a complete online reference for creating and managing *CaptureLifecycles*:



Introduction to DocuNECT Lifecycle Help for v4

This site contains detailed installation, configuration, and administration guides for DocuNECT v4.2. For high-level information about the products detailed here, please visit our [software web site](#). Visit our [software versions](#) page for the latest versions of the software.

Last Updated: 3/31/18

All About Capture Lifecycles

Managing Capture Lifecycles.

- [Installation](#)
- [Using the Capture Lifecycle](#)
- [Capture Lifecycle DocScript Overview](#)
- [Capture Lifecycle DocScript Functions](#)
- [Capture Lifecycle Editor Version History](#)

Developing with DocuNECT

Using DocuNECT as a development and integration platform.

- [Installation](#)
- [Using the CMB Developer](#)
- [Connector .NET Development Libraries](#)
- [CMB Developer Version History](#)

Distributing Documents and Data

Once the documents have been processed and the data extracted they can be distributed to an external source. The following external sources are support:

- ❑ EMC Documentum ApplicationXtender. Maps the documents and indexes to an ApplicationXtender application. DocuNECT™ can add pages to an existing document by checking whether the combination of index values already exists; create a new document, or new version
- ❑ EMC Documentum D6. Maps the documents and indexes to D6 document types and attributes. Custom permission sets (ACLs) can be applied, custom DQL can be executed to update registered tables, and folders can be created dynamically from index values
- ❑ Microsoft SharePoint. Maps the documents and indexes to a Microsoft SharePoint document libraries and columns. Custom permission sets (ACLs) can be applied and folders can be created dynamically from index values
- ❑ File Location or Network Share. The documents can be published to a file share with the index data being exported as a text file, an external database
- ❑ FTP or Secure FTP site. The documents can be published to a file share with the index data being exported as a text file, an external database
- ❑ Email. Documents and data can be emailed out to a specific target audience
- ❑ Web Service Upload. If the documents and/or data is to be uploaded to a third party business application then a web service can be called with an XML structure

Developing with DocuNECT™

DocuNECT™ provides a feature rich platform for developing document capture and management applications. The DocuNECT™ development platform is based Microsoft .NET v2.0 technology. There are three libraries/methods available in the development platform:

1. **DocuNECT Foundation Classes (DNFC)**. Provides access to all the functionality within the DocuNECT Web application system
2. **DocuNECT CaptureLifecycle Runtime Library (DCRL)**. Provides access to the document capture, indexing and distribution functions. This library also underpins the DocScript language used for developing Capture Lifecycle Definitions
3. **DocuNECT Capture Web Services (DCWS)**. A web services library it also available for uploading and indexing documents to make this functionality available over the web. Click on the link below to view the library functions
3. **ContentConnector Development**. This is used to provide a custom plug-in to the capture lifecycle process called ContentConnectors. This process is ideal for developing custom processing within a