

Course Guide

IBM Datacap 9.0.1: FastDoc Multi-Page Document

Course code F254 ERC 1.0



January 2017 edition

Notices

This information was developed for products and services offered in the US.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
United States of America*

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

© Copyright International Business Machines Corporation 2017.

This document may not be reproduced in whole or in part without the prior written permission of IBM.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Trademarks	v
Course description	vi
Agenda	viii
Unit 1. Add a Multi Page Document	1-1
What this unit is about	1
What you should be able to do	1
How you will check your progress	1
References	1
Unit Objectives	1-2
Lesson 1.1. Create a Multi-Page Form Application	1-3
Create a Multi-Page Form Application	1-4
Topics	1-5
Why is this lesson important to you?	1-6
Image Samples	1-7
Application template choice	1-8
Configure document, pages and fields view	1-9
Add an HRF_Document and HRF_Page	1-10
HRF_Page Settings	1-12
Recognition considerations	1-13
Full page zonal recognition	1-15
Recognize hand written text fields (ICR)	1-16
Recognize mark zones (OMR)	1-17
Field Settings tab	1-19
Configure Fingerprints	1-21
Add a Donation_Receipt Page	1-22
Demonstrations	1-23
Exercise: Create a Multi-Page Form Application	1-24
Exercise introduction	1-25
Lesson 1.2. Recognize with ICR and OMR	1-27
Recognize with ICR and OMR	1-28
Topics	1-29
Why is this lesson important to you?	1-30
Overview	1-31
Recognize with the ICR Engine	1-32
More Recognize Actions	1-33
Datacap Studio Zones tab	1-34
ICR/C Tab Settings	1-36
Recognize OMR	1-38
Establishing OMR Parent fields	1-39
Select OMR Properties View	1-40
Set OMR Length Property	1-41
The OMR Settings	1-42
Demonstrations	1-44
Exercise: Recognize with ICR and OMR	1-45
Exercise introduction	1-46
Lesson 1.3. Validate with database lookup	1-47
Validate with database lookup	1-48

Topics	1-49
Why is this lesson important to you?	1-50
Steps to Configure Lookup	1-51
Configure a lookup database connection string	1-53
Database Lookup Actions	1-55
Validate Fields Zip Lookup Function	1-56
Exercise: Validate with database lookup	1-57
Exercise introduction	1-58
Unit summary	1-59

Trademarks

The reader should recognize that the following terms, which appear in the content of this training document, are official trademarks of IBM or other companies:

IBM® and the IBM logo are registered trademarks of International Business Machines Corporation.

The following are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide:

DB2®

FileNet®

Initiate®

Tivoli®

WebSphere®

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other product and service names might be trademarks of IBM or other companies.

Course description

IBM Datacap 9.0.1: FastDoc Multi-Page Document

Duration: 3.0 hours

Overview

This course shows you how to create an application in FastDoc Server mode for processing a document with multiple pages. You also configure Recognition rulesets for Intelligent Character Recognition (ICR) for handwriting, Optical Mark Recognition (OMR) for check boxes, and database lookup.

You work with a fully functioning IBM FileNet Content Manager system configured with IBM Content Navigator and IBM Datacap 9.0.1 to practice the skills that are required to implement and configure data capture solutions.

Audience

- Application Builders who are responsible for building Datacap applications
- Anyone who needs to know the capabilities of FastDoc to create an application for processing a document with multiple pages.

Prerequisites

The following courses or equivalent knowledge are required:

- IBM Datacap 9.0.1: Introduction (F251)
- IBM Datacap 9.0.1: FastDoc Local Mode (F252)
- IBM Datacap 9.0.1: FastDoc Server Mode (F253)

Skills taught

Upon completion of this course, participants will be able to:

- Create a Form Template based Application for processing a document with multiple pages
- Use Page Separator pages to configure a multi-page batch
- Recognize hand written fields with Intelligent Character Recognition (ICR)
- Recognize handmade digital marks with Optical Mark Recognition (OMR)
- Recognize machine print fields with Optical Character Recognition (OCR)
- Validate using a lookup Database

Course outline

Add a Document Type

- Create a multi-page form application
 - Recognize with ICR and OMR
 - Validate using a database lookup
-

Agenda

Day 1

(01:00) Lesson 1 - Create a multi-page form application

(01:00) Lesson 2 - Recognize with ICR and OMR

(01:00) Lesson 3 - Validate with database lookup

Unit 1. Add a Multi Page Document

What this unit is about

This unit demonstrates adding a document with multiple pages to the document hierarchy and recognizing hand written input and mark input. It also demonstrates data validation with database lookup.

What you should be able to do

- After completing this unit, you should be able to:
- Create a multi-page form application
- Configure ICR and OMR fields
- Implement Validation with database lookup

How you will check your progress

- Successfully complete the activities in the Student
- Exercises book.

References

IBM Knowledge Center

http://www.ibm.com/support/knowledgecenter/SSZRWW_9.0.1/com.ibm.datacaptoc.doc/datacap_9.0.1.htm

Unit Objectives

- After completing this unit, you should be able to:
- Create a multi-page form application
- Configure ICR and OMR fields
- Implement Validation with database lookup

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-1. Unit Objectives

F2541.0

Notes:

Lesson 1.1. Create a Multi-Page Form Application

Create a Multi-Page Form Application

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-2. Create a Multi-Page Form Application

F2541.0

Notes:

Topics

- ▶ Create a Multi-Page Form Application
 - Recognize with ICR and OMR
 - Validate with database lookup

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-3. Topics

F2541.0

Notes:

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users.
- To configure most Datacap capture applications, you must know how to configure a document with multiple pages. The page images that you use in the class are rigid forms. They are best processed with Form Template based application. Also important is configuring Recognition rulesets for ICR (Intelligent Character Recognition) for handwriting and OMR (Optical Mark Recognition) for check boxes.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-4. Why is this lesson important to you?

F2541.0

Notes:

Image Samples

- Humanitarian Relief Fund

Humanitarian Relief Fund
This is not a bill. It is a form for the Humanitarian Relief Fund.

PLEASE CHOOSE YOUR PREFERRED PREFERENCE FROM ONE OF THE FOLLOWING OPTIONS:

CONTRIBUTION: ☒ One Time ☐ Quarterly ☐ Monthly ☐ Annual

PLEASE PRINT: 10050000 10050000

DATE: 03/02/05

PLEASE PRINT: 4398204956814326

EXP. DATE: 03/08

PLEASE PRINT: Barbara Gill

PLEASE PRINT THE FOLLOWING INFORMATION:

LAST NAME: GILL

FIRST NAME: BARBARA

ADDRESS 1: 101 MAIN STREET

ADDRESS 2:

CITY: TARRYTOWN

STATE: NY ZIP: 10898

SOCIAL SECURITY NUMBER: 082539100

PLEASE SELECT ALL THAT APPLY:

☒ Your Honor ☐ Fund Raising

☒ Publications ☐ Awareness

YOU MAY HAVE YOUR CONTRIBUTION OR DONATION ON:

RED CROSS

Signature: Barbara Gill

Add a Multi Page Document

- Donation Receipt

Donation Receipt from Humanitarian Relief Fund	
Questions? Contact the Humanitarian Relief Fund at 777-888-9999.	
Donated by	Barbara Gill
Address	101 Main Street
City	Tarrytown
State/Province	NY
ZIP/Postal Code	10898
Phone	678-555-6789
Type of donation	CASH
Description	Credit Card donation
Value	\$1005.00
Thank you for your generosity. We appreciate your support!	

© Copyright IBM Corporation 2017

Figure 1-5. Image Samples

F2541.0

Notes:

The image on the slide illustrates sample Humanitarian Relief Fund form and a donation receipt.

Humanitarian relief Fund Form

- Structured layout
- Hand written text fields
- Red Boxes represent potential OMR fields

Donation Receipt

- Structured layout
- Machine printed txt

Application template choice

- The Humanitarian Relief Fund form document and Donation Receipt have a structured format.
 - They always have the same fields.
 - The fields do not move, they are at a consistent location.
 - Ideally suited for the Form Template application.
- Create the basic application.
 - From FastDoc Admin (local mode).
 - Run the Application Wizard.
 - Create an RRS application.
 - Application name: HRF_Form.
 - Select Form Template.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-6. Application template choice

F2541.0

Notes:

Configure document, pages and fields view

- FastDoc (Admin) Datacap Server mode:



- Configure the Batch Structure.
 - Add documents, pages, and fields.
- Configure Ruleset Properties.
 - Configure ruleset properties and Test profiles and rulesets.
- Configure Fingerprints.
 - Add fingerprint classes, add a fingerprint page, and map page fields to image zones.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-7. Configure document, pages and fields view

F2541.0

Notes:

The screen capture shows the FastDoc “Configure documents, pages, and fields” window.

Add an HRF_Document and HRF_Page

- Add a document (HRF_Document) to the document hierarchy structure.
- Add a page (HRF_Page) to the HRF_Document.

- Add fields to the HRF_Page:
 - Zip
 - City
 - State
 - Frequency

	HRF_Page
<input type="checkbox"/>	Zip
<input type="checkbox"/>	City
<input type="checkbox"/>	State
<input type="checkbox"/>	Frequency

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-8. Add an HRF_Document and HRF_Page

F2541.0

Notes:

The screen capture shows the options for copying page configuration from an existing page, when you are creating a new page.

In this example, HRF_Document, HRF_Page (first page), and four fields on the HRF_Page are used:

HRF_Document

|

HRF_Page

/ / \ \

Zip City State Frequency

Remember:

- When defining new Documents and Pages to always click Enable for the Use rulesets from?
 - If you are defining a new Document, select Document from the option list.
 - If you are defining a new Page, select Page from the option list.
- If you select the Document or page option, then the default rules from the template for the document or page are copied for your new object.
- If you do not enable the use of the default rulesets, then you are responsible for creating your own rulesets for managing the objects.
- Note; Click the? Next to Use rulesets from:? for helptext that describe this option.

HRF_Page Settings

- With HRF_Page selected and the Settings Tab selected:
 - Define Minimum, Maximum, and Order for the HRF_Page
 - Minimum 1
 - Maximum 1
 - Order 1

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-9. HRF_Page Settings

F2541.0

Notes:

It also defines the Min, Max, and Order for the HRF_Page. All three variables have a value of 1

- There must be one, and only one, HRF_Page in an HRF_Document and it is always the first page.

Recognition considerations

- For the HRF_Page:
 - Configure recognition as full page zonal recognition.
- For the HRF_Page, Fields:
 - Text fields are hand written.
 - Recognition needs to be configured for ICR.
 - The Frequency, Payment Type, Credit Card Type, and Contribution Usage fields are optical marks fields.
 - Recognition needs to be configured for OMR.
- For Donation_Receipt pages, the fields are machine printed.
 - Recognition needs to be OCR.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-10. Recognition considerations

F2541.0

Notes:

ICR = Intelligent Character Recognition (for hand writing)

OMR = Optical Mark Recognition

OCR = Optical Character Recognition (machine print text)

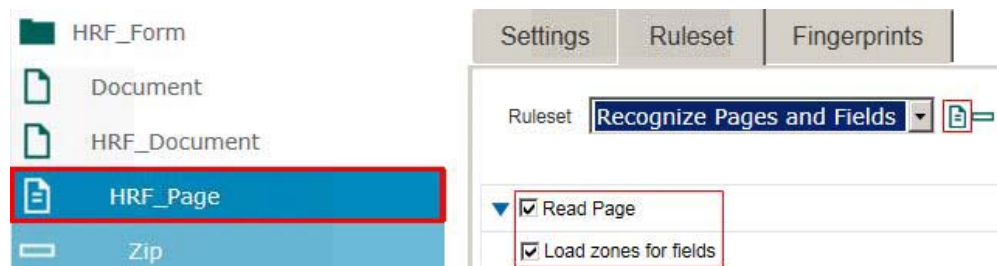
Data recognition is the stage during which you locate the fields that you want to capture and then convert the fields into character-based data.

- The data that is obtained from recognition is stored in the page data files that you set up in the document assembly stage.
- You can use several techniques that to identify pages.
 - Fingerprint matching: (The most widely used)

- Other non-fingerprint methods to analyze the page and identify and locate the fields are:
 - Text matching
 - Pattern matching
- If you used fingerprint match for page identification, you most likely used the fingerprint images to define the recognition zones.
- These zones are the fields that you want to read on each page.

Full page zonal recognition

- Page level recognition for HRF_Page and Donation_Receipt.
- In FastDoc, set these options at the Batch Structure Page level.



- For full page recognition, you get the field data directly from the full page recognition results.
- Otherwise, you need to run the recognition engine on each field zone to capture the data.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-11. Full page zonal recognition

F2541.0

Notes:

The image on the slide illustrates the Ruleset tab view of the of the FastDoc Configure, documents, pages, and fields window. The options shown are the options for full page zonal recognition.

Recognize hand written text fields (ICR)

- In FastDoc, set these options at the Batch Structure field level.



- More ICR configuration is done in Datacap Studio (covered in next lesson).

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-12. Recognize hand written text fields (ICR)

F2541.0

Notes:

The screen capture shows the Ruleset tab view of the of the FastDoc Configure, documents, pages, and fields window.

ICR = Intelligent Character Recognition (for hand writing)

For the HRF_Page > Fields

Read field check box

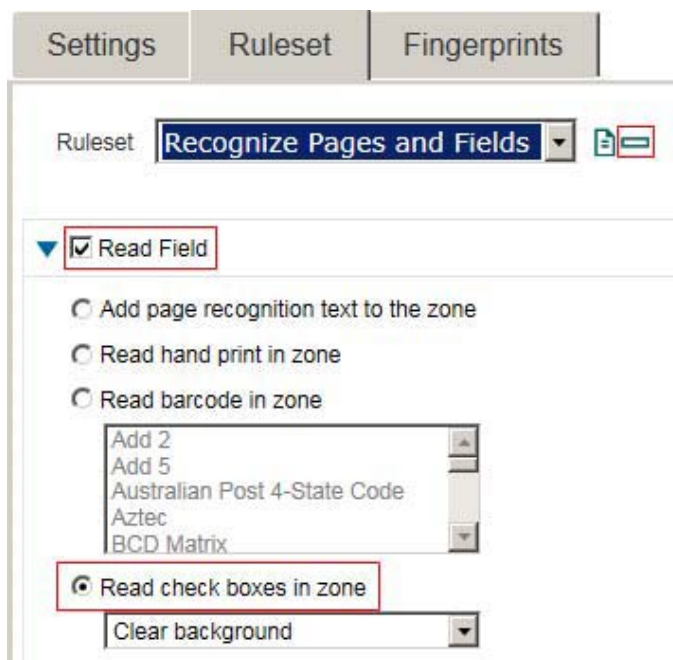
- It is a general option required for any recognition to be done on the field.

Read hand print in zone option

- If text fields are hand written, this option activates ICR field level actions in the “Recognize Pages and Fields” ruleset.
- More configuration is needed in the Datacap Studio Zones tab, in the ICR/C tab of the properties pane. (The Datacap Studio configuration is covered in the next lesson).

Recognize mark zones (OMR)

- For the HRF_Page > mark zone fields
- In FastDoc, set these options at the Batch Structure field level.



Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-13. Recognize mark zones (OMR)

F2541.0

Notes:

The screen capture shows the Ruleset tab view of the of the FastDoc Configure, documents, pages, and fields window. The options shown are the page recognition option.

OMR = Optical Mark Recognition

For the HRF_Page > Fields

These options are set for the zone that surrounds all of the mark zones. (In your example, it is the Frequency field.)

Read Field check box

- Is a general option required if any recognition it to be done on the field.

Read check boxes in zone option.

- If fields are check boxes, this option activates OMR field level actions in the “Recognize Pages and Fields” ruleset.
- All zones that are defined within this zone are interpreted as OMR zones.
- More configuration might be needed in the Datacap Studio Zones tab, in the OMR tab of the properties pane. (The Datacap Studio configuration is covered in the next lesson).
- More configuration is required on the Settings tab in FastDoc.

Field Settings tab

- Set Optical mark Enable.
- Configure Optical mark option list for the Verify step.

Settings Ruleset Fingerprints

Field type: Frequency

Display name:

Optical mark: ? ☒ Enable

Dictionary: ☐ Allow multiple selections

Value	Display Text	
1	One Time	↓ ↑
12	Monthly	↓ ↑
4	Quarterly	↓ ↑
2	Annually (this year & next)	↓ ↑

Add

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-14. Field Settings tab

F2541.0

Notes:

The screen capture shows the Settings tab view of the of the FastDoc Configure, documents, pages, and fields window.

For the HRF_Page > Fields

These options are set for the zone that surrounds all of the mark zones. (In your example, it is the Frequency field.)

Optical mark Enable check box

- Select to be able to run OMR recognition
- You must add a dictionary item for each check box.

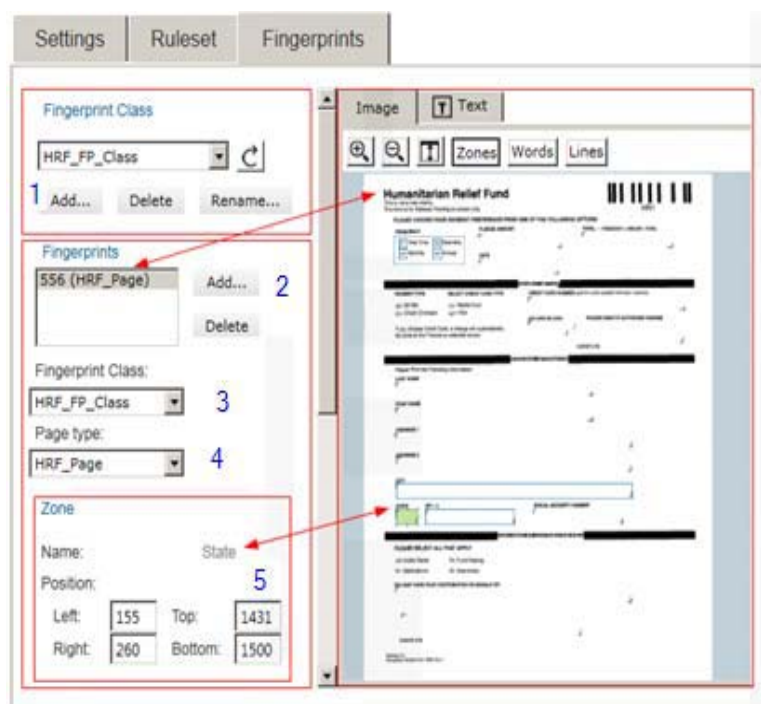
Add Dictionary items

- Building a dictionary table does two things
 - It adds an OMR field under the (Frequency) Check box zone field.
 - Creates an entry in an option list that can be used on the Verify panel if you need to override the OMR value for this field.

There is also an option to allow multiple selections from the list. “Allow multiple selections”.

Configure Fingerprints

- Add a Fingerprint class
- Add a Fingerprint image
- Select a Fingerprint class
- Select page type
- Mark Zones



Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-15. Configure Fingerprints

F2541.0

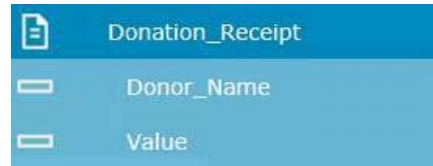
Notes:

The screen capture shows the Fingerprints tab view of the of the FastDoc Configure, documents, pages, and fields window.

- Fingerprints are defined to enable the identification of images by the position of data items on a page.
- When an image is identified by matching it to a fingerprint, it is quick to extract the data values from the fields that the zonal information defines that is stored in the fingerprint.
- Locating data on an image with the zonal information stored in the fingerprint is more efficient than other methods of locating data like doing key word searches.
- You can assign the fingerprints that you create to fingerprint classes to differentiate the forms that you receive from various sources.
- Fingerprint classes can be used to group fingerprints based on the source of the form.

Add a Donation_Receipt Page

- Add a second page to the HRF_Document and name it Donation_Receipt.
- Add fields:
 - Donor_Name
 - Value
- Define Min, Max, and Order for
 - the Donation_Receipt:
 - Minimum 1
 - Maximum 0
 - Order 0

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-16. Add a Donation_Receipt Page

F2541.0

Notes:

The screen capture shows the rule for identifying a second or subsequent Donation_Receipt.

If you set Minimum to 1, then must always be at least one donation receipt included in a document.

If you set Minimum to 0, then having no donation receipt is acceptable.

Note: If you set Maximum to 1 and then do not include a donation receipt in your batch, then the batch is routed directly to FixUp to correct the problem before processing the batch.

Demonstrations



- Configure a multi-page document type
- Configure rulesets for a multi-page document type
- Add a fingerprint to a page

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-17. Demonstrations

F2541.0

Notes:

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

Exercise: Create a Multi-Page Form Application

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-18. Exercise: Create a Multi-Page Form Application

F2541.0

Notes:

Exercise introduction

- Create a Form Template based application
- Add a Second page to the document



Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-19. Exercise introduction

F2541.0

Notes:

Lesson 1.2. Recognize with ICR and OMR

Recognize with ICR and OMR

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-20. Recognize with ICR and OMR

F2541.0

Notes:

Topics

- Create a Multi-Page Form Application
- ▶ Recognize with ICR and OMR
- Validate with database lookup

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-21. Topics

F2541.0

Notes:

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users.
- To build a multi-page Datacap application, you must improve your ability to recognize pages and fields with more advanced techniques. In this lesson, you use Datacap Studio to expand on the configuration for hand written text fields recognition with ICR, and multiple option check boxes recognition with OMR.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-22. Why is this lesson important to you?

F2541.0

Notes:

Overview

- OCR and ICR engines:
 - ocr_a: ABBYY FineReader OCR engine
 - OCR_s: Nuance (formerly ScanSoft) OmniPage OCR engine
 - ocr_sr: Newer implementation of the Nuance OmniPage OCR engine
 - icr_c: Open Text RecoStar ICR engine
- The Forms Template uses the following actions for recognition:
 - On a page that only does hand written text fields and optical mark fields:
 - Page recognition
Zones: ReadZones() or FPXML: ReadZonesFPX()
 - Field recognition

For hand writing	icr_c: RecognizeFieldICR_C()
For optical mark	ocr_a: RecognizeFieldOCR_A()
 - On a page with machine print:
 - Page recognition
Machine print ocr_sr: RecognizepageOCR_S()
 - Field recognition
Machine print Recog_Shared: SnapCCOtoDCO()

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-23. Overview

F2541.0

Notes:

The RecognizePageOCR_S() is the action that is used for full-page recognition creates the .cco file that is later used for fingerprinting and locating data on a page.

- Datacap includes: three optical character recognition (OCR) engines,
 - OCR_a: ABBYY FineReader OCR engine
 - OCR_s: Nuance (formerly ScanSoft) OmniPage OCR engine
 - OCR_sr: Newer implementation of the Nuance OmniPage OCR engine
- Datacap includes: one Intelligent Character Recognition (ICR) engine that you can use to do full-page recognition:
 - ICR_c: Open Text RecoStar ICR engine
- The OCR engines work well with machine-printed text,
- The ICR engine works well with hand-printed and machine-printed text.
- The ICR_sr engine is not used to do full-page recognition in this application.

Recognize with the ICR Engine

- Uses the OpenText RecoStar engine to recognize constrained (unconnected) hand- or machine-printed characters.

Recognize Actions	Description
EnableLoggingICR_C	Enables or disables event log for the ICR/C engine.
RecognizeFieldICR_C	Does character recognition on the current field.
RecognizeFieldVoteICR_C	Does recognition on the current field zone and compares the result to the existing field value, character by character, raising the confidence level when the characters match and lowering it when they do not.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-24. Recognize with the ICR Engine

F2541.0

Notes:

The image on the slide illustrates a table of Recognize actions with descriptions.

ICR recognition

The ICR, or Intelligent Character Recognition, engine is an advanced optical character recognition (OCR) engine. It uses the OpenText® RecoStar™ engine to recognize constrained (unconnected) hand- or machine-printed characters.

The length of a field is important for ICR. If the field length is set to 9, the ICR engine divides a zone into nine pieces and looks for a character in each piece.

More Recognize Actions

Recognize Actions	Description
RecognizePageFields2CCO_ICR_C	Does recognition for all zoned fields on the page.
RecognizePageFieldsICR_C	Does recognition on all fields that are configured for ICR/C.
RecognizePageICR_C	Does full page recognition with the ICR/C engine.
RecognizePageToPDFICR_C	Does full page recognition and stores the results in a PDF file.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

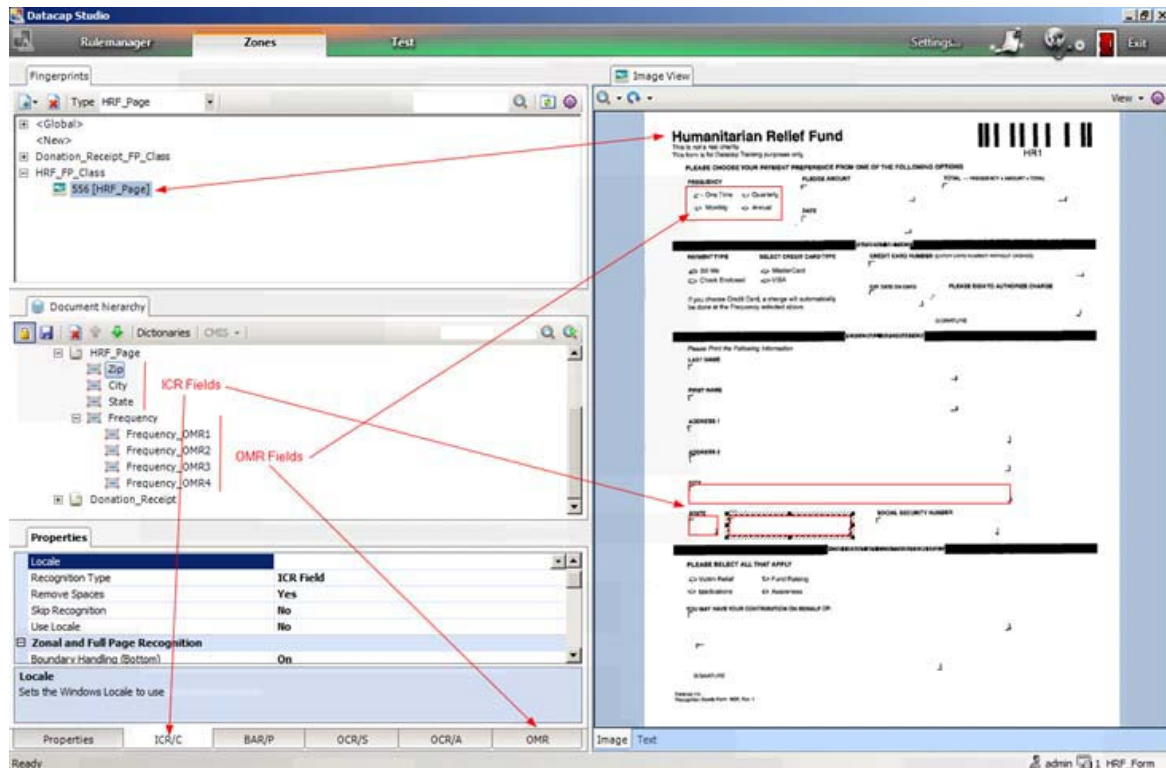
Figure 1-25. More Recognize Actions

F2541.0

Notes:

The image on the slide illustrates a table of Recognize actions with descriptions.

Datacap Studio Zones tab



Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-26. Datacap Studio Zones tab

F2541.0

Notes:

The screen capture on the slide shows a view of the Datacap Studio Zones tab.

- The previous lesson stated that more configuration for the ICR fields is required in Datacap Studio.

ICR configuration

- Verify that zones are correctly defined for each ICR field.
 - Click each ICR field that is defined in the Document hierarchy pane and verify that the correct field in the Image View gets highlighted.
- Note:** you must lock the document hierarchy view before the zones become visible in the Image View.
- Click each ICR field in turn in the Document Hierarchy pane.
 - Click the ICR/C tab in the properties pan in the lower left corner.
 - Configure the ICR properties as shown on the “ICR/C tab setting” slide.

OMR Configuration

- Verify that the OMR parent area is zoned. (Click the Frequency field in the Document Hierarchy pane and verify that the zone below the Frequency heading is marked around the four optical mark fields.
 - Notice that the four OMR fields Frequency_OMRx are defined in the Document Hierarchy field but there are no zones that are defined for them.
 - These fields were created when you defined the Frequency Dictionary table in FastDoc.
- Zone the four OMR fields.
 - Zoom in on the Frequency zone so that it is large enough to properly zone the OMR fields.
 - Make the Zones rectangles of the same size and space line them up vertically and horizontally.
 - Remember that you must define the zones in the same order that you used when you created the Dictionary table.
 - Review the OMR properties as shown on the “Select OMR Properties View” slide.

ICR/C Tab Settings

- Recognition Type
- Boundary handling
- Character set
- Country
- Delete intermediate project file
- Font
- Length
- Logical context
- Number of lines
- Pattern
- Pitch
- Reader
- Syntax
- Triagram Mode

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-27. ICR/C Tab Settings

F2541.0

Notes:

ICR field setup

ICR settings for a field must be done on the ICR/C tab. Ensure that some white space is left around the boundaries while marking a field in fingerprints.

Key settings on the ICR tab:

- **Local:** Warning. Clear this field.
- **Recognition Type:** Select ICR for ICR fields.
- **Boundary handling:** Turn on boundary handling, especially if there is not much space around each field. The application reads and extracts data from the outer parts of marked zones if these fields are on.
- **Character set:** Select the appropriate option if you want to extract specific data from the field. For example, select 0-9 for ZIP code.
- **Country:** Select the country or language of the data to be extracted.

- **Delete intermediate project file:** Set this value to False if you want the application to generate a log file for debugging. The file is saved in your Temp folder with a .rsp extension.
- **Font:** Select handprint for ICR.
- **Length:** Enter the number of parts that you want the zone to be divided into before extracting the data. For example, if you are identifying the ZIP code field, enter 9 here. The application divides the ZIP code into nine sections and looks for data in each field.
- **Logical context:** Set this value to **On** to set the application to check for the most likely value of a questionable character. For example, if you set this value to **On** for a ZIP code, the application identifies the character “L” or “I” as “1”.
- **Number of lines:** Set this value to **Unknown** if the number of lines in the zone is unknown.
- **Pattern:** Use this option to match the specified RegEx (regular expression) pattern.
- **Pitch:** Select **Fixed** if the values in the zone are of the same size. If the values vary, select **Variable**. If you are not sure, select **Unknown**.
- **Reader:** Use this option to assign either the RecoStar or AEG reader to read the data. Select the Voter option here. This option is a combination of RecoStar and AEG.
- **Syntax:** Indicates the most likely data type in the zone. For example: date, numeric, text.
- **Triagram Mode:** Set this value to On. The application checks if the values next to the recognized value are any indication of the most likely value of the character.

Recognize OMR

- Optical mark recognition (OMR)
 - Used to electronically extract intended data from marked fields on printed forms.
 - Scans a printed form and reads predefined positions and records where marks are made on the form.
- OMR methods:
 - Pixel threshold evaluation method:
 - Is more reliable for dropout check boxes.
 - Can also be used to read fill-in bubbles on a response form.
 - Is difficult to set up.
 - OCR/A check box recognition method:
 - Is easy to set up.
 - Works well with non-dropout check boxes.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-28. Recognize OMR

F2541.0

Notes:

OMR fields

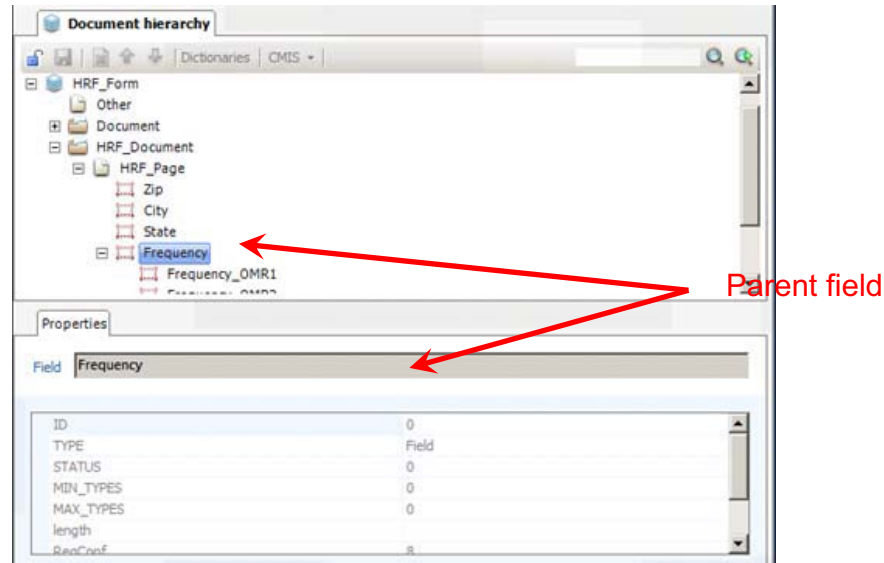
Optical Mark Recognition (OMR) is a technology that is used to electronically extract intended data from marked fields on printed forms. These marks are check boxes and complete fields like bubbles. The OMR technology scans a printed form and reads predefined positions and records where marks are made on the form.

OMR field setup

Datacap employs optical mark recognition (OMR) to determine whether a check box option is selected or not. There are two basic techniques:

- OCR/A check box recognition method
- Pixel threshold evaluation method

Establishing OMR Parent fields



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-29. Establishing OMR Parent fields

F2541.0

Notes:

The image on the slide illustrates the Document hierarchy and Properties panes and highlighting the OCR parent name filed.

Establishing Parent fields

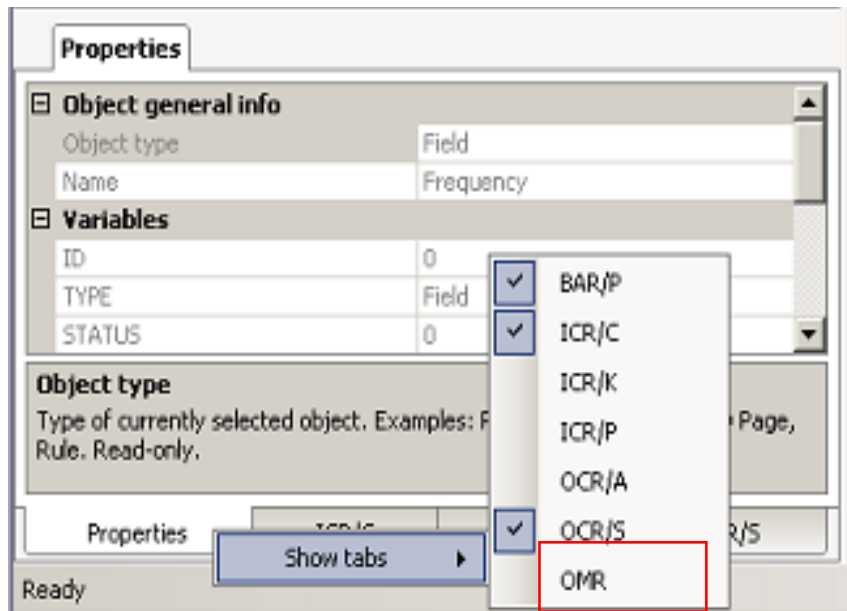
Processing pages with check box options:

- The system automatically defines the check box options as subfields of the parent Frequency field within the document hierarchy.
- You must outline the subfields and the parent field when drawing the recognition zones.

When you define the recognition zones, you need to define the positions of the parent fields and the subfields.

Select OMR Properties View

- For pixel threshold evaluation method



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-30. Select OMR Properties View

F2541.0

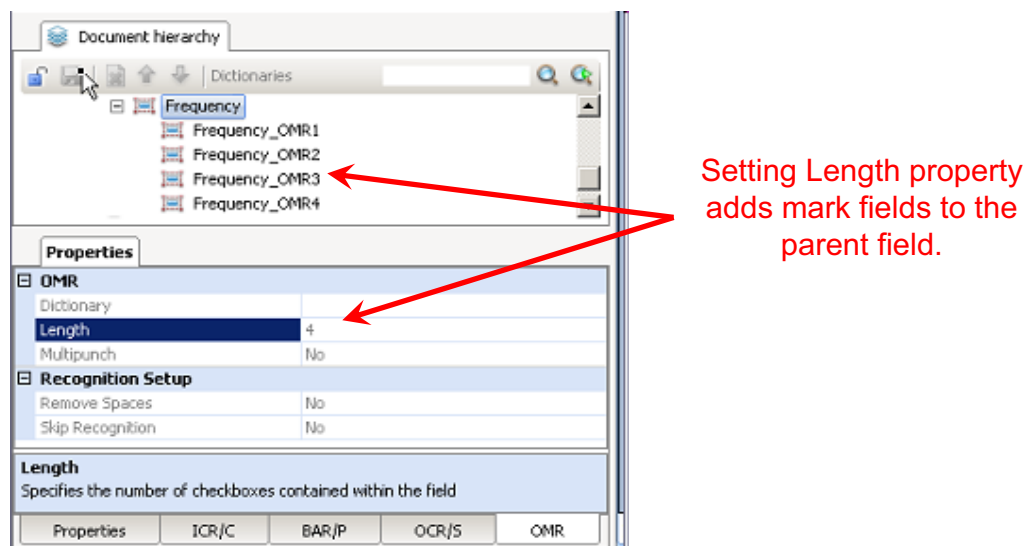
Notes:

The image on the slide shows the selection of the OMR option from the properties pane.

You must configure the OMR settings for specific zones with the OMR tab in the Properties pane on the Datacap Studio Zones view. The OMR tab is not shown by default, so you must enable it, as described in the following section.

Right-click any existing tab and click **Show tabs**. Then, select the **OMR** option.

Set OMR Length Property



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-31. Set OMR Length Property

F2541.0

Notes:

The image on the slide illustrates Document hierarchy tab showing the Frequency OMR fields and the corresponding Length field in the Properties pane the represents the number of OMR zones.

With the Pixel threshold evaluation method

Click the OMR tab to show the settings that the OMR recognition engine uses when doing recognition on the selected field.

The Length property defines how many mark fields are defined for the parent field.

Note: In this class, you define the Frequency field in FastDoc. When you create the OMR dictionary list for manually selecting the OMR field options at Verify time, the subfield corresponding to the list options are automatically defined. All that is left to be done is to associate the mark fields with the zone on the image where the data is read from.

The OMR Settings

Check mark type	Select "Square background" to read non-dropout Check boxes. This setting is stored in the document hierarchy with the OMRTYPE variable, where 0 is "Square background": <V n="OMRTYPE">0</V>
Length	This setting reflects the number of OMR subfields and is set automatically.
Multipunch	This setting is the same as the MultiPunch variable where 1 is "Yes": <V n="MultiPunch">1</V>

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-32. The OMR Settings

F2541.0

Notes:

The image on the slide illustrates the OMR setting options.

With the Pixel Threshold Evaluation Method

The pixel threshold evaluation method uses the RecogOMRThreshold action in the "Recog_Shared" library and is the method that is used in an exercise.

Specifying the Threshold and Background levels

The RecogOMRThreshold action takes two parameters:

- **Threshold:** Specifies the percentage of black pixels, which cause the option to be considered selected.
- **Background:** Used to determine the confidence level and specifies the percentage that can be attributed to the check box outline plus any scanner noise:

Any zone with below the background value is considered not selected with high confidence. Any zone with between the background value and the threshold value is considered not selected with low confidence.

Any zone with greater than $(2 * \text{Threshold} - \text{Background})$ is considered selected with high confidence. Any zone with between Threshold and $(2 * \text{Threshold} - \text{Background})$ is considered selected with low confidence.

Demonstrations

- Configure ICR and OMR fields



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-33. Demonstrations

F2541.0

Notes:

If you are taking this course as a self-paced virtual class, return to the main course menu to play the pre-recorded demonstrations.

Exercise: Recognize with ICR and OMR

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-34. Exercise: Recognize with ICR and OMR

F2541.0

Notes:

Exercise introduction

- Configure ICR and OCR fields



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-35. Exercise introduction

F2541.0

Notes:

Lesson 1.3. Validate with database lookup

Validate with database lookup

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-36. Validate with database lookup

F2541.0

Notes:

Topics

- Create a Multi-Page Form Application
- Recognize with ICR and OMR
- ▶ Validate with database lookup

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-37. Topics

F2541.0

Notes:

Why is this lesson important to you?

- As a Datacap business analyst, you build and deploy applications with the Datacap Capture system and communicate solution details to the solution architect, administrator, and users.
- To build intelligent Datacap applications, you must be expert at doing database lookups to validate data that is interpreted on the captured pages. In this lesson, you validate a postal code with database lookup to illustrate this important skill.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-38. Why is this lesson important to you?

F2541.0

Notes:

Steps to Configure Lookup

- Configure a connection string.
- Validate the new database connection.
- Configure Validate Fields Ruleset for field validation.
- Configure Lookup parameters in the Validate Fields Ruleset.
- Save your edits.

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-39. Steps to Configure Lookup

F2541.0

Notes:

Overview

The purpose of validation is to determine whether captured data conforms to specified business rules. For example, you need to determine whether the vendor information matches the information that is stored in a database of approved vendors.

Use database lookup to validate a ZIP Code

In this section, you use the Zip field on an existing page to demonstrate how to test whether a field contains a value that is in a list of allowed values. For each Zip field that the ICR process detected, you test the field value.

Configure a connection string

- Use Datacap Application Manager to connect to the lookup database.
- Copy the lookup database into the application root folder.

Sample: C:\Datacap\Expense

- Click the ellipsis at the end of the database string to open the 'Data Link Properties' window to configure the link to the database.
- Select the 'Database type and properties name': **Microsoft Access (Jet)**
Browse to the database: **C:\Datacap\Expense\Zipac.mdb**

Validate the new database connection

- Click test connection and check that you get Success status.

Configure Validate Fields Ruleset for field validation

- Use FastDoc to open the application.
- Select the index field for the lookup database.
- Configure validation parameters for the field.

Configure Lookup parameters in the Validate Fields Ruleset

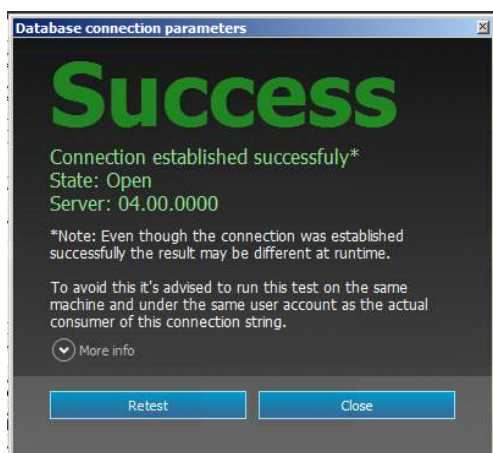
- Click the Lookup check box and configure the lookup parameters.
- Select the database.
 - Type the SQL statement to do the lookup.
Sample for the Zip field:
Select Zip From Zipac Where Zip='+@p\Zip+';
 - Click the 'Populate field' check box.
 - Is any data in the Zip field, is the minimum field length satisfied.

Save your edits

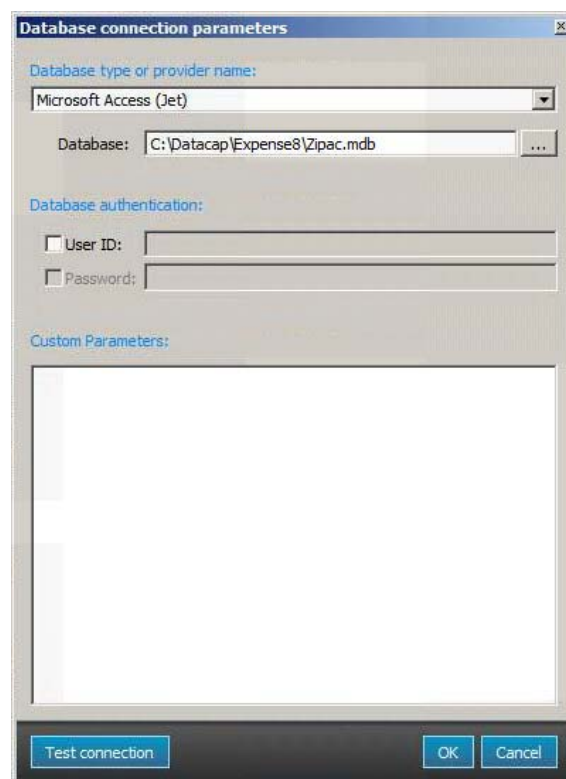
- Click Save and close the FastDoc window.

Configure a lookup database connection string

- Configure a connection string.
- In the Application Manager.
- Click Lookup database ...
- Select the Database type.
- Browse to the database.
- Test the connection.



Add a Multi Page Document



© Copyright IBM Corporation 2017

Figure 1-40. Configure a lookup database connection string

F2541.0

Notes:

The images show:

- The Database connection parameter window where you create the connection to the lookup database.
- I shows the Success message window after you have done the test connection operation.

Configure a connection string

Create a file on the desktop named **zipcode.udl**.

Double-click zipcode.udl to open the Data Link Properties window to configure the link to the database.

Provider option: **Microsoft Jet 4.0 OLE DB Provider**

Database location: **C:\Datacap\Expense\images\HR\Zipac.mdb**

Look for the connection string in desktop\zipcode.udl.

Update the Lookup database string

Start Datacap Application Manager:

Start > All Programs > IBM Datacap Services > Datacap Application Manager

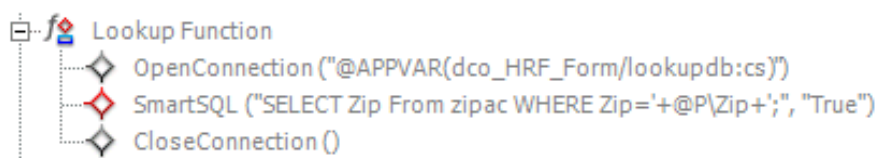
Update the Lookup database string on the Application tab.

Database Lookup Actions

- The Lookup library includes actions for connecting to external data sources and running SQL statements.

Lookup Actions	Description
OpenConnection	Uses a data source name or connection string to open a connection to a database.
SmartSQL	Run a SQL statement with support for smart parameters.
CloseConnection	Closes an open database connection.

- Validate Field > Lookup function:



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-41. Database Lookup Actions

F2541.0

Notes:

The Lookup Function shown in the screen capture shows one of the default functions activated when the Lookup option is selected in the Validate Fields ruleset.

- When you are using one of the Application templates (Learning or Form) for your application, the template automatically creates the necessary actions in the Standard Validate field ruleset. It also provides the plumbing for connecting to the Batch Structure objects and the Task profiles.
- These Lookup library actions that are shown in the slide table, are used in the template Validate Field Ruleset when the lookup option is selected in the Validate Fields ruleset in the FasDoc interface. They connect to external data source, run SQL statements, disconnect from the database, and populate the field.

Validate Fields Zip Lookup Function

- FastDoc Lookup Configuration:

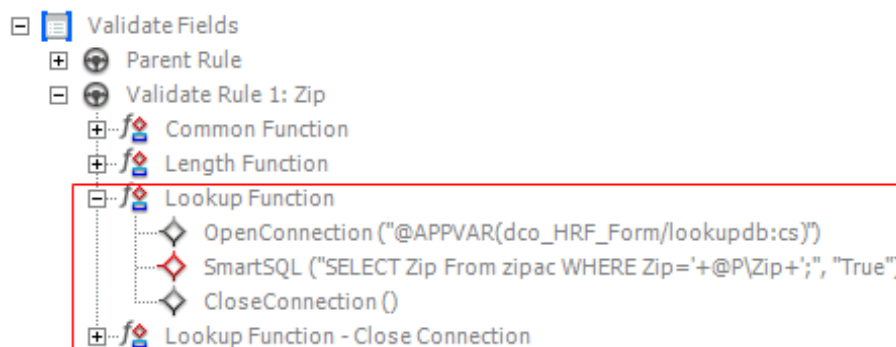
Validates the current field value by allowing it to be used in a database query, optionally populating fields with lookup results.

Database:

SQL statement:

Populate field: ☒

- Standard Template Validate Fields Ruleset:



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-42. Validate Fields Zip Lookup Function

F2541.0

Notes:

The screen capture images on the slide show:

- The UI from FastDoc for the Validate Fields Ruleset and the Lookup validate option.
 - Notice that the Lookup check box is set.
 - The configured parameters define:
 - The Database to read from: HRF_FORM: lookupdb
 - The SQL statement to read the zip value from the database with the zip field value as an index into the table.
 - The option to populate the Select field
- The Ruleset that is created based on the FastDoc options.
 - Notice that the Open Connection action and SmartSQL action parameters match the values that are entered in FastDoc.

Exercise: Validate with database lookup

[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-43. Exercise: Validate with database lookup

F2541.0

Notes:

Exercise introduction

- Configure a zip code database lookup
- Configure a zip code truncate ruleset



[Add a Multi Page Document](#)

© Copyright IBM Corporation 2017

Figure 1-44. Exercise introduction

F2541.0

Notes:

Unit summary

- After completing this unit, you should be able to:
- Create a multi-page form application
- Configure ICR and OMR fields
- Implement Validation with database lookup

Add a Multi Page Document

© Copyright IBM Corporation 2017

Figure 1-45. Unit summary

F2541.0

Notes:



IBM Training

