# ABBYY FlexiCapture 10
## Project Setup Guide

Dear user!

This guide describes the actions performed during project setup.

If you still need more information, try the following:

- Use other help files. You can open them using the program menu or by clicking "F1" or Start > Programs > ABBYY FlexiCapture 10 Stations> Helps.
- System Administrator's Guide: Start > Programs > ABBYY FlexiCapture 10 Servers > Guides > System Administrator's Guide.
- Project Setup Guide: Start > Programs > ABBYY FlexiCapture 10 Stations > Guides > Project Setup Guide.
- A Guide to Creating Machine-Readable Forms: Start > Programs > ABBYY FlexiCapture 10 Stations > Guides > A Guide to Creating Machine-Readable Forms.
- Help files for applications in FlexiLayout Studio 10 and FormDesigner 10 are opened from the menu of the corresponding application or be pressing F1. You can also click Start > Programs > ABBYY FlexiCapture 10 Stations > Helps.

For the standalone version, ABBYY FlexiCapture 10 Stations is replaced with ABBYY FlexiCapture 10 in the paths described above.

We hope you will enjoy using our product!

# Contents

# 1.  Introduction

## 1.1.  The Purpose of Data Capture

A large variety of documents is used today: in business, production and services. Applications, questionnaires, invoices, drafts and other documents are essential for any company. Modern information technology makes paper documents insufficient, and most data is converted to electronic form for storage, analysis and processing purposes.

The most labour- and time consuming thing about electronic documents was data input. It could only be entered by hand, which was reasonable with a small amount of information. However, this doesn't work well with large document volumes. The speed of manual entry cannot be momentarily increased when the situation so demands, because manual entry groups are hard to manage, and the costs of changes may equal those of starting the process anew.

Thus, manual entry is not the optimal way. Its alternative, a simpler and more effective way, is an automatic data capture system like ABBYY FlexiCapture 10.

## 1.2.  Data Capture Automation

ABBYY FlexiCapture 10 is data capture software for processing structured, semi-structured (FlexiLayout) and non-structured documents.

Automatic data capture consists of the following stages:
- A pack of pages is **scanned** using a document scanner first;
- The scanned pages are then automatically bundled into documents;
- The characters are automatically **recognized**;
- Uncertainly recognized characters are sent to the Operator for **checking** (**verification**);
- Confirmed data is finally **exported** to a file or database, and document images are saved to the specified folder. Images can be saved in a graphical format or as a searchable PDF file.

ABBYY FlexiCapture 10 is an efficacious data capture automation solution allowing easy control over work progress and quality.

## 1.3. Documents You Can Process in ABBYY FlexiCapture 10

ABBYY FlexiCapture 10 is a data capture application supporting different document types.

The following document types can be processed in ABBYY FlexiCapture 10.

- ***Structured documents***. Documents with dedicated data fields that remain constant in quantity, position and formatting throughout the document copies are called structured. These forms are often issued in printed form for filling by hand. In order to identify a structured form and capture the data, a layout must be created indicating field locations to the program. Layouts are created in ABBYY FlexiCapture 10 during project setup.

- ***Semi-structured documents***. Documents with data fields that differ in quantity, position and formatting from copy to copy are called semi-structured or flexible. Invoices are an example of this type, because they are often different in the number of items and formatting, for they are issued by different companies. All invoices include an account number and the amount of payment, but these are located in different parts of the document. ABBYY FlexiCapture 10 uses FlexiLayouts for identification of semi-structured documents and data capture. FlexiLayouts are created in ABBYY FlexiLayout Studio. For information on ABBYY FlexiLayout Studio, see its Help File. Processing of flexible or semi-structured documents is different from that of structured documents only at creation and layout matching stages.

- ***Non-structured documents***. If you need to process non-structured documents with information presented in free form, for example contracts, letters, orders, diagrams, ABBYY FlexiCapture 10 will also be a solution. Non-structured documents can be automatically identified as supplements to structured or flexible documents, or by using FlexiLayouts, and then exported to image and searchable PDF files. Index fields can be captured from non-structured documents automatically (using FlexiLayouts) or manually. A typical non-structured document processing scenario would be converting a paper archive to electronic form, with capture of a couple of index fields required for attribute-based search.

## 1.4. ABBYY FlexiCapture 10 Stations

Depending on the installation type (standalone or distributed), the system will include different Stations. If you are using the Standalone version, projects are set up at the Administrator Station. If you are using the Distributed version, projects are set up at the Project Setup Station and then sent to the Application Server.

# 2. Creating a Project

A ***Project*** includes all the settings necessary for data capture (Document Definitions, Image Import Profiles, Program Settings), and the documents to be processed.

Documents are merged into ***batches***. The batch formatting type depends on the processing type. For example, you can merge documents with the same creation date or from the set of documents scanned at one go. Settings to be used during batch recognition are defined by the batch type. You can create several batch types with the most commonly used processing settings. In this case, the Operator will only have to select the necessary batch type. You can also specify the settings for the whole batch. These settings will be applied to projects with the type set as "Default". First, the project settings are specified, and then, on their basis, batch types are formed (project settings are default settings for batch types).

***Work Batches*** are used for document processing. ***Test Batches*** are used during Document Definition debugging. The difference between these batch types is that Test Batches make use of the local (unpublished) Document Definition, while Work Batches use the published Document Definition. You can access the list of Test Batches right from the Document Definition Editor window.

A ***Document*** consists of an image of one or several pages (i.e. single-page and multi-page documents) and the data captured from them.

Multiple Document Definitions can be included in the project. In this case, documents of different types will be processed within one project. Thus, you don't need to presort the documents for processing. You can combine different document types in one flow. However, if document flows don't intersect, you can create separated projects for them.

First, you need to create a project and at least one Document Definition.

To create a new project, either click **New…** in the **Open Project** dialog box that appears at program start, or click **File > New Project…** on the main menu. Select a folder to save the project to and specify the name.

To add batches to the project, use the **New Batch** command of the shortcut menu that appears when you right-click the batch list. It is possible to load images without creating a batch. In this case, a batch will be created by the program automatically.

To view the documents that were added to the batch, double left-click the batch name. To return to the batch list, click **Project > Work Batches** or press Ctrl+B.

# 3.   Creating Document Definitions

Creating Document Definitions is the main stage of project setup affecting the quality of resulting data after document processing. To create a Document Definition, do the following:

- Create a layout (layout proper – in the Document Definition Editor or by loading a form built in ABBYY FormDesigner, and FlexiLayout – by loading a file built in ABBYY FlexiLayout Studio 10);

- Set properties for each field and specify what data is to be found in each field (it increases recognition quality considerably). Also, specify which fields are to be sent to the Operator for Verification;

- Set Field Value Check rules. The rules help identify documents with values that don't meet the requirements, for example, the field value doesn't correspond to the values from a particular database.

- Set the data export method. Data can be exported to a file, a database, to Microsoft™ SharePoint or according to the script procedure.

When the definition has been created, it must be published. It will then be available for document processing.

The main actions of creating and editing the Document Definition are carried out in the ***Document Definition Editor*** window that opens after a new definition has been created. To open the Document Definition Editor, click **Project > Document Definitions…**, then select the definition name and click **Edit…**.

To create a new definition, click **Projects > Document Definitions...** and, in the dialog box that opens, click **New…**. The Document Definition Creation Wizard will open. In the **Create New Document Definition** window, you can specify the main properties of the definition: the name, comment, language and style. Then set the text type: select ICR (for hand-printed text) or OCR (for machine-printed text) from the dropdown list. Later, it will be possible to change the text type for individual fields, and the type selected at this stage will be used as default.

Next, you need to load or scan the image serving as a prototype for the Document Definition. (If your document consists of multiple pages, load the first page for now. To learn how to add additional pages, refer to the «Creating a Document Definition for Multi-Page Documents»). You can scan the page (preferably not filled in) or load the image from a file. If you are going to process semi-structured documents, use a FlexiLayout. Select **Load FlexiLayout** and enter the path to the FlexiLayout file with the *.afl extension (the file that was created in ABBYY FlexiLayout Studio). If you were creating a flexible Document Description, this will be the last step.

If you are creating a fixed Document Definition, the next step is as follows. Select field types to be automatically found on the image. You can set checkmarks and text entry fields. The best results during automatic search can be achieved if you use text fields with marking and rectangular checkmarks. If entry fields have no marking, and the checkmarks are to be set against a white background, it is recommended that you create them manually.

If your form has anchors (static elements of the given form, for example, squares used for identification and/or matching of a Document Definition), they will be found and automatically marked on the image as well.

# 3.1. Creating a Layout

## 3.1.1. Definition Objects' Geometric Marking

When all the steps of creating a Document Definition are complete, the loaded image will appear in the Document Definition Editor window. The image will have fields of the types you selected at the last step as well as anchors.

Automatic object selection can be carried out later by clicking and clicking somewhere in the area of the element you need to select. The element type and location will be identified automatically.

The Document Definition Editor has a set of easily accessible tools for field and static element marking:

- Field Regions Mode ( ) and
- Static Element Mode ( ).

To create a static element or field manually, you need to select one of these modes by clicking the corresponding button on the toolbar, and then, clicking and holding the left mouse button, draw a rectangle around the desired element. You can also hold Shift and left-click somewhere in the object area. In this case the field or static element region will be defined automatically.

This is the list of tools for creation of different element types:

**Fields:**                                          **Static Elements:**

☐ - text entry field         ☐ - anchor

☐ - checkmark                ☐ - separator

☐ - checkmark group          ☐ - static text

☐ - barcode                  ☐ - barcode

☐ - picture

☐ - table

☐ - group (of fields)

A barcodes can be either a recognizable field or a static element. Care must be taken when selecting the mode that depends on the barcode purpose – if information is to be captured from the barcode, mark its region in Field Regions Mode, and if it is used for definition identification and matching, mark it in Static Element Mode.

Created fields appear in the list on the **Fields** tab of the **Document Structure** window. Names matching the comments or the field names will be used by default. You can change the field name by selecting it in the document structure and pressing F2. If you want the field to have the name matching the comments, select the field, right-click it and, in the shortcut menu, click **Get Name from Image…**. Then, draw a rectangle around the necessary text.

Field type can be identified by the icon in the list of fields and the frame color. Static elements are not displayed in the list.

Objects can be copied (including to other document sections), deleted, and moved, and their size can be changed. Names of field copies are numbered.

You can also select several objects at a time. To do it, hold Ctrl while selecting. The ☐ tool can also be used.

### 3.1.1.1. Field groups

Fields can be combined into groups to make document structure more illustrative or to created repeated field groups. For example, the city, street and house number fields can be combined into the "Residence Address". Then, you can copy the "Residence Address" field group to create the "Work Address" fields.

To combine documents, use the ☐ tool.

If there are repeated field groups in the documents, you can create several instances of the group you have created. All field properties as well as rules set within the group will work for each instance of the group. See 3.1.1.5 "Fields with multiple instances".

You can also copy a field group, but, in this case, copying will result in creating a new independent group.

### 3.1.1.2. Fields with no marking

There can be fields having no corresponding region on the image. Fields with no marking are marked with a red asterisk in the list. Such fields can be used to store calculation results for values in recognized fields.

Fields with no marking have all the properties characteristic of their type: they can be sent to the Operator for Verification, the format check can be executed, and the field values can be exported.

To create a field with no marking, do one of the following:

1. In the Document Definition Editor, click **Edit > Create Field** and create a field of the required type. The field will appear in the list and will be marked with an asterisk. In this case, a document structure field will be created, but not its region on the image.

2. Delete marking of a regular field. Select a field on the image or list and, on the shortcut menu, click **Delete Region**. The marking will be deleted, and the field will be marked with a red asterisk.

To create a region for fields with no marking, click  on the toolbar. If there are fields marked with an asterisk in the list, the program will prompt you to select a name for one of such fields. This may happen when you set the list of fields first and define their locations later.

### 3.1.1.3. Table marking

ABBYY FlexiCapture 10 gives you the possibility of working with tables. The **Table** fields are used for it.

A range of tools is available for table marking in regular Document Definitions. These tools are used to draw tables, position lines and columns.

Draw a table region using the  tool. The table header must not be included in this region. Next, mark the table cells using the  tool to add separators. To create vertical separators, point the cursor to the table region, then drag the dotted separator to the desired place and left-click to fix it in its place. Horizontal separators are created using the same method, only holding Alt. You can also initiate an automatic separator search. To do it, select the table you have just created and use the following commands of the shortcut menu: **Autodetect Vertical Separators** and **Autodetect Horizontal Separators**. To delete a separator, click . When you have added the required number of separators, set the columns by clicking  and clicking in the column region. Each column contains cells of one type: text, checkmarks, graphic elements or barcodes. The program will prompt you to select the column type when you set it.

When the geometric marking is finished, you will need to set the recognition and verification properties as well as the data type for each column. Column properties are set exactly like the properties of regular document fields.

### 3.1.1.4. Fields with multiple regions

If there are fields with regions consisting of several parts on your form, for example, tables stretching across multiple pages, you can create fields with multiple regions for processing of such data.

Values from all regions of a field are merged and exported as one field. The line break serves as a separator here.

To create a field with multiple regions, create one of the field regions, select it, then right-click it, and, on the shortcut menu, click **Continue Region…** Then, select a place for the region to continue at. Repeat the procedure the required number of times.

### 3.1.1.5.   Fields with multiple instances

Your document can include repeated objects – fields or field groups that are present in several instances and that describe similar objects, for example, the same information about employees or their children, or invoices. Fields with multiple instances are created for such objects.

Any field, except for a table, can have multiple instances, with their regions located far from each other, possibly on different pages. Field instances have the same properties, and fields with multiple instances are exported to separate database files or tables.

Fields with multiple instances can be useful when creating repeated field groups: you create a field group, and then the necessary number of instances is created.

To create a field with instances, create one instance first, them select it, right-click the field region and, on the shortcut menu, click **New Instance…**. Create the necessary number of region instances and move their images to the required location on the page.



**Figure 1. Using fields with multiple instances**

### 3.1.1.6.   Excluding a region from recognition

You may need to exclude a region from recognition in certain cases, for example, a region with comments that hinders data capture of a field (Figure 2). To exclude a region from recognition, click ⬚ and draw a region to exclude with your mouse.
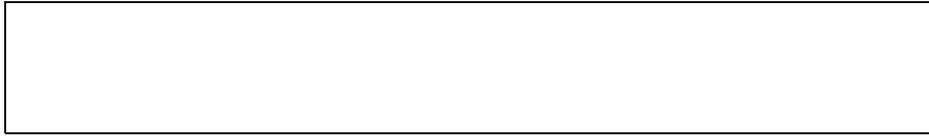
**Figure 2. Excluding an unrecognizable region**

## 3.1.1.7.   Deleting fields

To delete a field, select it and press **Delete** or click **Delete** on the shortcut menu. To delete marking only, leaving the field in the document structure, press **Shift+Delete** or click **Delete Region** on the shortcut menu of the field.

## *3.1.2.   Static Elements*

Static elements are objects that do not provide information for capture. They are used for Document Definition matching and identification. Anchors are a type of static elements.

To begin working with static elements, switch to the special mode by clicking , which will display static element marking.

All static element types can be used to match a Document Definition, but the best results can be achieved only if the documents have standard anchors – black squares, crosses or angles. They must be set manually or automatically as a static element of the *Anchor* type. Their shape must be specified in the Element Properties window, the **General** tab.

Static elements of any type can be used as identifiers as well. Analyzing the position and value of the identifiers, the program finds the document the current page belongs to. If a barcode is used as the identifier, you can specify its value manually – this helps identify the page quickly and precisely.

To use a static element for Document Definition matching and/or identification, select the corresponding option in the **General** tab of the Properties dialog box. To open the Properties dialog box for any element, click **Properties...** on the shortcut menu.

Static elements can be used when matching a Document Definition (select **Use to match Document Definition**) and/or for Document definition identification (select **Use to identify Document Definition**).

*Anchor*: set the anchor type (square, cross, angle or rectangle). For standard anchors, select **Use to match Document Definition**. Select **Use to identify Document Definition** only if the anchor locations are unique for the current document section.

***Static text***: if the static text will be used for identification, you can enter the text value. You need to enter the text value only if the page can't be identified by its location alone (for example, if the pages are different only in heading texts, while the location and size of these headings are the same).

***Separator***: choose to use the separator for document definition identification or matching.

***Barcode***: if the barcode will be used as an identifier, you can enter the barcode value. On the **Recognition** tab, specify the barcode type, its orientation and image processing options.

## 3.1.2.1.   Barcode Specifics

If a barcode is used as an identifier, it is an anchor barcode and it belongs to static elements. Create it when you are working with static elements. The Properties dialog box of such barcode has two tabs: **General** and **Recognition**.

If a barcode is used for data capture, it is a field. Create it when you are working with fields. The Properties dialog box of such barcode has all the tabs of the Field Properties dialog box: **General**, **Data Type**, **Recognition**, **Verification**, and **Rules**. The value of such barcode will be recognized and, if the settings provide for it, sent to verification and export.

## 3.1.3. *Field Properties*

It is important to set field properties correctly, because they affect the quality of field recognition and define whether the field values are to be exported and sent to the Operator for verification. Some properties are important for data recognition. For example, the text field marking property is to be set carefully, so that the marking that doesn't disappear during scanning is separated from the characters. It will lead to the recognition resulting in text only, without unnecessary elements.

Correct field properties lead to better recognition results making the Operators' work easier. When you set field properties, document verification can be reduced to minimum.

When you have created a field of a certain type, it will have default properties. Some properties are taken from the Document Definition (for example, language), others are defined automatically (for example, marking type). They can be changed in the dialog box that opens when you click **Properties** on the shortcut menu of any object. Every field type has its own properties.

### 3.1.3.1. General Field Properties

The Properties dialog box of every field has the **General** (Figure 3) tab. The **Name** and description are specified on this tab. You can change the name automatically given to the field at the creation stage which corresponds to the comments that are most close to the field. **Caption** is the field name as displayed as data. The field type is displayed on this tab as an icon to the right of the name.

Additionally, the following options can be selected on the **General** tab:
- **Export Field Value** – clear this option if you don't need to export the field value. This may be necessary, for example, if the field value is used to get the value of the calculated field (see <u>Rules Checking</u>), and you need to export the final result only.
- **Read only** – select this option to prohibit any changes by an Operator. You may want do select the option for fields with values to be calculated automatically (according to rules), for example, for fields where you plan to save a sum or combination of values of other fields (see <u>Rules Checking</u>).
- **Show in data form** – clear this option if you don't want the field to appear on the data form when viewing documents. If an error occurs in such field, you will be unable to correct it in the Document Editor window, because the field will not be displayed. Thus, it is recommended to specify the settings that will not cause any errors in such fields.
- **Cannot be blank** – select this option if the field is required to be filled in. If the field is empty, a format error will occur.
- **Index field** – select this option if you are going to index the field. In this case, the field value will appear for each document in the list, and the Operator will be able to sort and search for documents by the value of this field. The index field value can also be used for file naming during export.
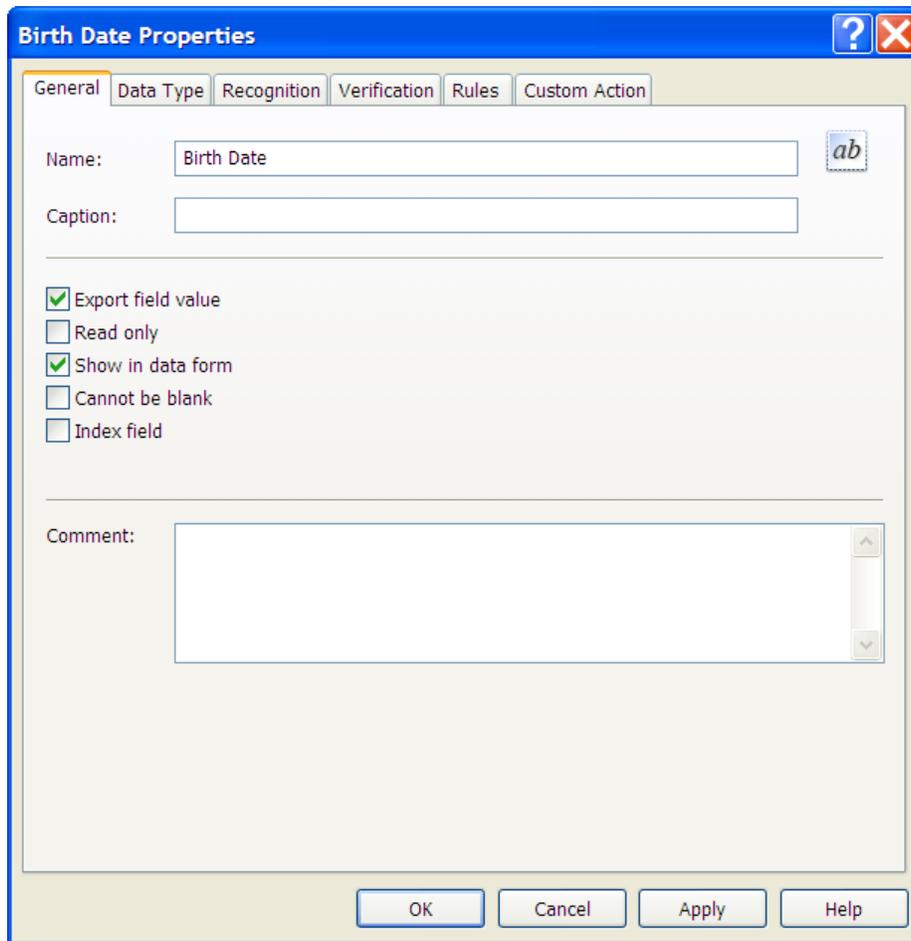
**Figure 3. The Field Properties dialog box, General tab**

## 3.1.3.2.  Data Type

*Data Type* defines the set of possible field values and the allowed field format. If the value entered in the field doesn't correspond to the specified data type, the operator will receive a verification error message. The text data type usually has a simple area of valid values: the date, time, address, taxpayer's account number (INN), and sum. As for the checkmarks, the data type means values the field assumes when the checkbox is selected or cleared.

### 3.1.3.2.1.  *Data types of the text entry field*

It is important to set the *data type* of the text entry field correctly. This tells the program what data type is to be found in the field – figures, certain letters, specific characters, date or other. The program allows the user to flexibly set data types. The user gets a ready set of data types which includes the types used most often. Besides, the user will be able to create his/her own data types to suit the task at hand.

When you set data types, you can configure the system to check the format of input value as well as set value restrictions, such as maximum and minimum numbers, time period for the date, and maximum number of characters for text fields.

The data type is set on the **Data Type** tab of the Field Properties dialog box (Figure 4).

Select a category from the **Content** list. In the **Details** field, the description of one of the data types (chosen by default, or specified manually earlier) that belong to the category will be displayed.

If the **Process value as text** option is selected, the values of fields with any content will be processed and exported as text. The field format check will not be carried out either.

To change the recognition **Language** or specify a more restricted data type, click the **Edit…** button located to the right of the description.

For standard (or general) types (**General** is selected in **Content settings** list), you can do the following:

- For Text – select several recognition languages (the "…" button). You can use the built-in dictionary and/or custom dictionary. Information in the dictionaries will be taken into consideration during field recognition then.

- For Number – select **Integer** if the value is expected to be an integer number;

- For Date – select the date component order, choose (not) to show months in words, choose (not) to show time and day of week;

- For Address, Name, Code – use a custom dictionary.

Special types (select **Special** in the **Content settings** list) contain predefined data types you can choose the most suitable from. Mind the description on the bottom of the dialog box when making a selection.

You can create your own data type if none in the list suits your needs.

1. To create a new data type, select one of the values in the **Contents** list on the **Data Type** tab. You can select any value in list that would correspond to the purpose of your type. The resulting data type will be stored in the selected category, although the category itself doesn't influence the new type.

2. Click the **Edit…** button located to the right of the **Details** field. In the dialog box that opens, select **Special** in the **Content settings** list. Click **New…**.

3. Follow the New Data Type Wizard's instructions.

14
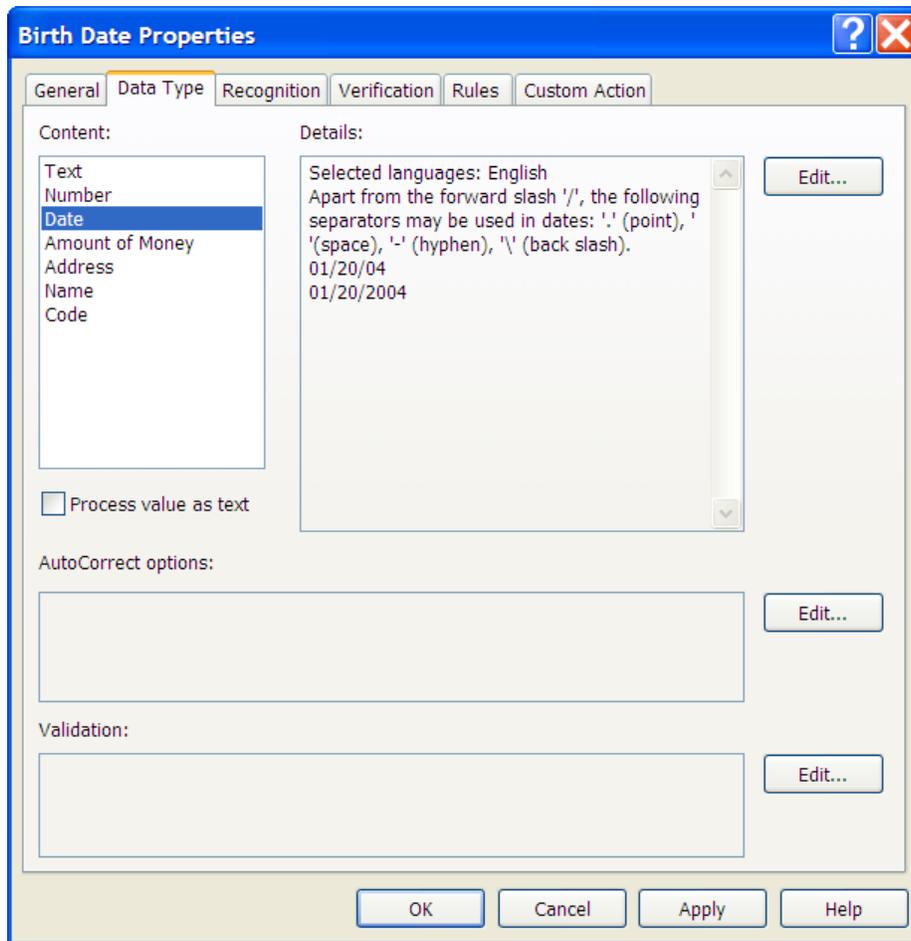
**Figure 4. Field Properties dialog box, Data Type tab (text entry field)**

Automatic processing of the recognized value can be carried out for any data type. Unnecessary spaces will be deleted, capitalization and spelling corrected. To start automatic processing, click the **Edit**… button located to the right of the **Replace characters** field. In the dialog box that opens, specify the necessary text processing parameters (Figure 5).
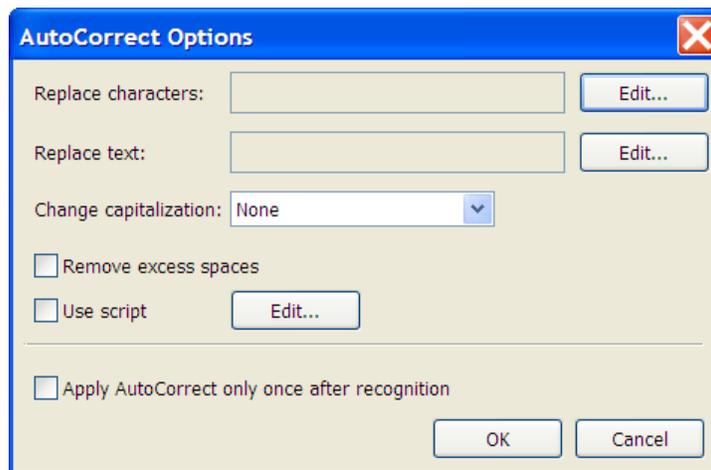


**Figure 5. The Autoreplace Settings dialog box**

You can also configure the program to check the recognized field value for compliance of the value with the set conditions (in this case, for date the program will check if the value is within the specified time

15

period, for text fields, if the value is valid and if it has the required format, etc.). To specify value restrictions for a field, click the **Edit…** button located to the right of the **Value Check** field. (Figure 6).
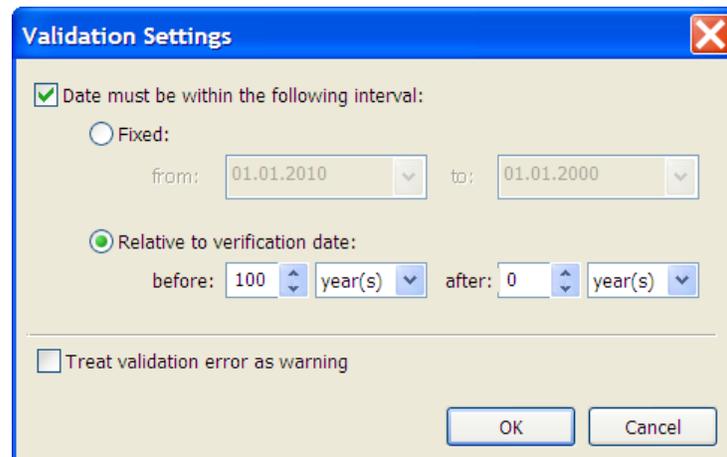


**Figure 6. Value Check Settings dialog box**

### *3.1.3.2.2. Data types for checkmarks*

For checkmarks, you can specify the values assigned to fields at checkmark selection/clearance. It is done on the **Data Type** tab (Figure 7).

The following values can be assigned to checkmark fields:

- *Yes* – if the checkbox is selected, *No* – if the checkbox is clear;
- *1* – if the checkbox is selected, *0* – if the checkbox is clear;
- *Checkmark name* – if the checkbox is selected, *Empty line* – if the checkbox is clear;
- User-created values can also be entered by selecting the corresponding option.

*Note.* If the checkmarks are grouped, they share the same properties which are created for the whole group. Checkmark values are created in the group properties as well.
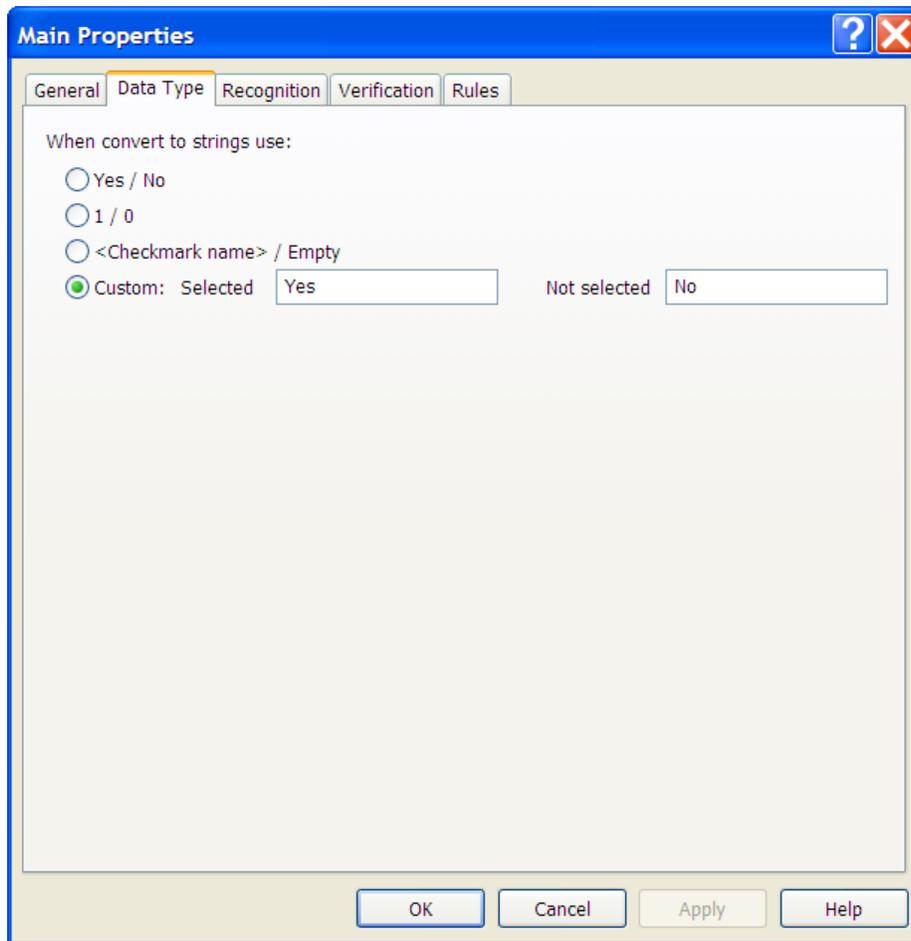
**Figure 7. Field Properties dialog box, Data Type tab**
**(checkmark not in the group)**

### *3.1.3.2.3. Data types for checkmark groups*

In the Checkmark Group Properties dialog box, on the **Data Type** tab, the list of names of checkmarks in the group will be displayed (Figure 8).

Clear the **Allow empty selection** option if you want at least one checkmark to be in the group.

If  it is possible to select multiple checkmarks in the group, select **Allow multiple selection**.

You can define the exported values in case no checkmark is selected, or if more than one checkmark is selected. Select either *<Empty selection>* or *<Multiple selection>* and click **Edit…**. In the dialog box that opens, enter a value in the **Exported value** field. If the export value is not set, an empty line will be exported in case no checkmark is selected. In case multiple checkmarks are selected, the values will be exported separated by commas.

If the **Treat validation error as warning** option is selected, a warning will be issued instead of an error message.
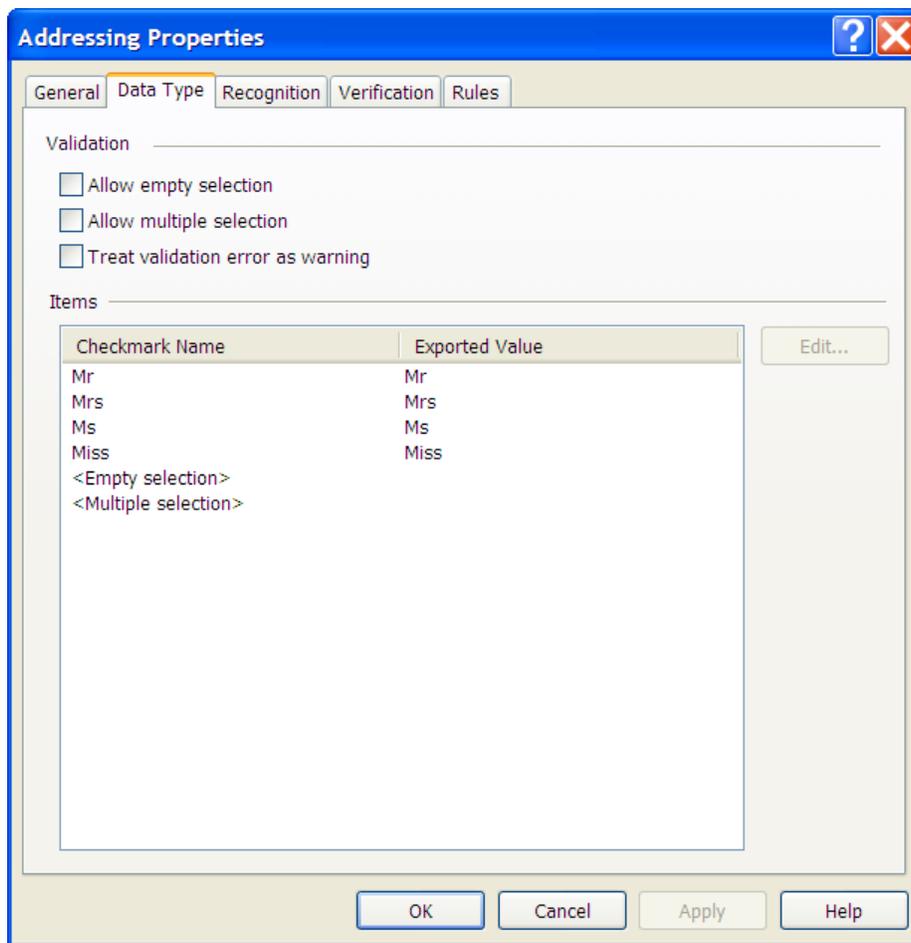
**Figure 8. Field Properties dialog box, Data Type tab (checkmark groups)**

## 3.1.3.3.  Field recognition properties

ABBYY FlexiCapture 10 allows specifying recognition settings for each field. Field properties defined correctly on the **Recognition** tab of the Field Properties dialog box will increase recognition quality and minimize the possibility of error. The properties differ for various field types. The recognition properties by field are as follows.

### *3.1.3.3.1.  Text entry field recognition properties*

Select the **Don't recognize (Key from Image field – will be entered manually)** option if you don't want the field to be recognized, and if the value will be entered manually by an Operator. This may be necessary, for example, if the field value is unrecognizable due to small intercharacter spacing. In this case, you don't need to specify other recognition properties, because this kind of field will not be recognized. The Operator will be prompted to enter the field value during the field value check.

Now select the text type: **ICR (for handprinted text)** or **OCR (for printed text)**. If you selected OCR, choose the printing type from the dropdown list (typographic printer, matrix printer, typewriter, etc.). To set multiple text types or use a template, select **Advanced** and click **Modify…**

Select **Marking type** from the dropdown list of examples. If the marking disappears during scanning, it is recommended to choose the type suitable for fixed-pitch text (*Gray rectangles*). If the marking doesn't disappear and is split into character cells, specify the character cell number (it can be defined

automatically). The *Simple* type is used for fields with no marking, usually for texts printed on typographic printers.

You can select the **Letter case** for letters of a particular case to be found. If the field can be filled with both small and capital letters, choose **Auto**.

Select either the horizontal or vertical text **Orientation**.

For a one-line field, select **One line**. For a field that will always contain one word (i.e. without spaces), select **One word**.

Specify the image preprocessing properties. For example, the image can be despeckled with adjustable size of "garbage" to be deleted. You can also invert the image or remove the texture.

*Note*. If the **Despeckle image** option is selected, the size of "rubbish" is defined automatically by default. To set your own size, select **Clear the garbage of specified size only** and specify the size.
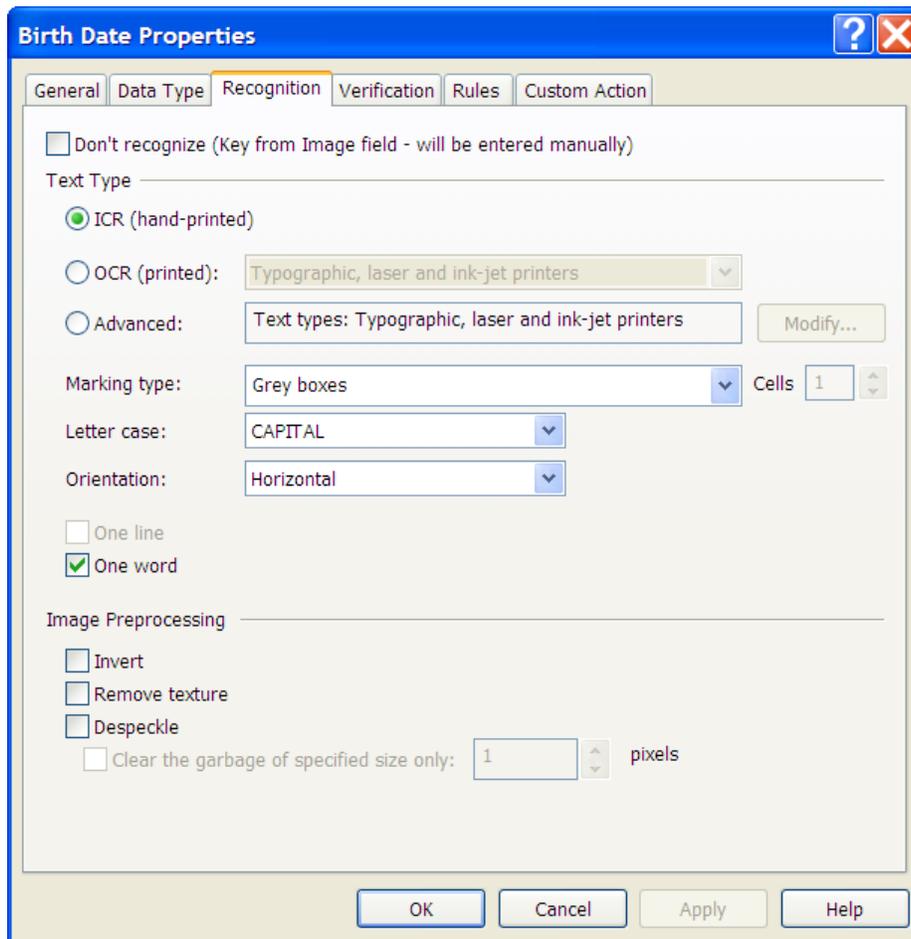


**Figure 9. Field Properties dialog box, Recognition tab (text entry field)**

### 3.1.3.3.2. *Checkmark and checkmark group recognition properties*

For the checkmarks to be recognized correctly, specify the **Checkmark type** by using the dropdown list. If the checkmark is in a square, select **Square**; if the checkmark is set against a white background without any frame (or if the frame disappears during scanning), select **Without frame**. Select **Auto** for the program to define the checkmark type automatically. In this case, the checkmark must be empty, because the program will define whether there's a checkmark by comparing the image of this area in the processed document with that in the form the Document Definition is based on.

19

You can allow corrections for certain checkmarks – if a checkmark was put incorrectly, the person who did it can shade the whole checkmark. Completely shaded checkmarks will be deemed void. However, if you selected the **Auto** type, you won't be able to allow corrections.

Image preprocessing can be configured for checkmarks just like for text fields.

If checkmarks are grouped, they will share the same properties. Recognition properties are defined likewise, but for the whole checkmark group.
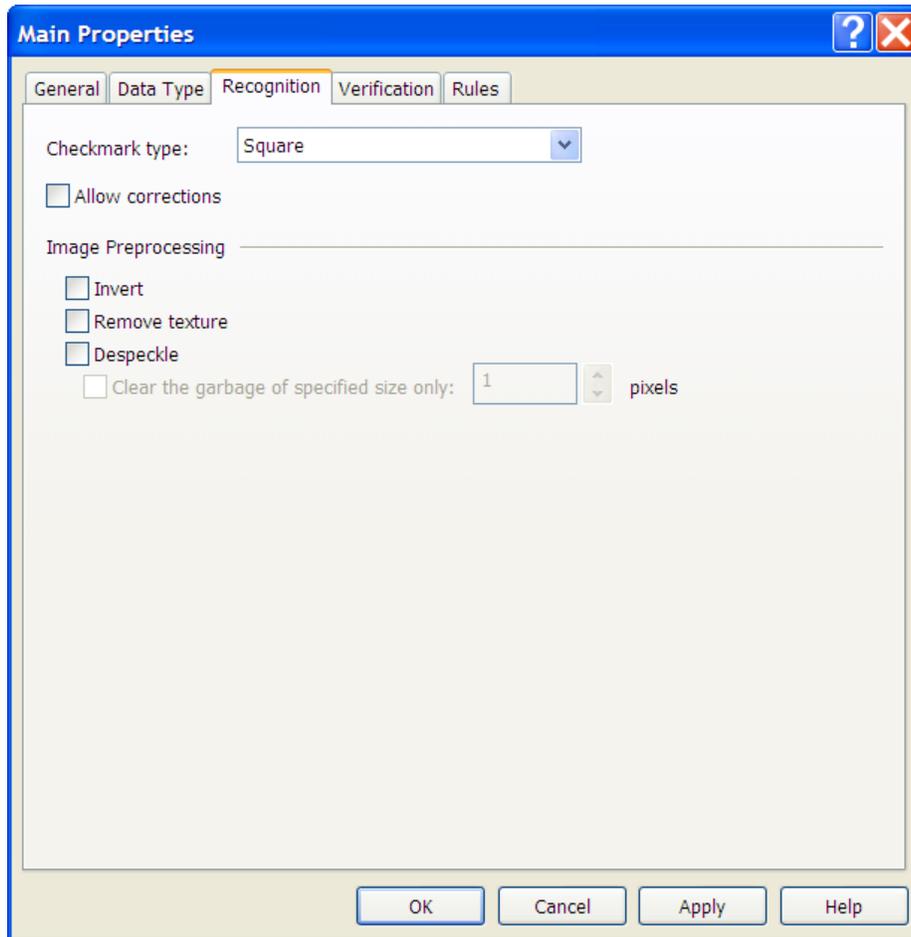


**Figure 10. Field Properties dialog box, Recognition tab (checkmark field)**

### 3.1.3.3.3. *Barcode recognition properties*

The properties of a barcode that is a field are the same as those of a Static Element barcode. In case of a field barcode, the barcode type, orientation and image despeckle settings are specified. The only difference is that the field value can be entered by the Operator. To allow this, select **Don't recognize (Key from Image field – will be entered manually.**

### 3.1.3.3.4. *Image recognition properties*

On the **Recognition** tab of the Image Field Properties window, select **Exclude from recognition** if the image area is to be excluded from recognition. This option is only available for the Image field type and is used for compatibility with FormReader 6.5 templates).

## 3.1.3.4. Verification settings

*Verification* is checking recognized data by an operator. During Document Definition creation, you can configure the verification settings on the **Verification** tab of the Field Properties dialog box (Figure 11). Uncertainly recognized characters will be highlighted by the program and sent to an Operator for checking. However, it is also possible to configure verification in such a way, that the field will be sent to an Operator even if there are no uncertain characters, or vice versa, won't be sent for checking even if there are uncertainly recognized characters. Choose *all* (required verification) if you want to be absolutely sure there is no error in the field value.

Character inclusion in group and field verification can be enabled by selecting the corresponding option(s). Group verification applies to checkmarks and digits most. Other characters can also be verified in group verification, if the corresponding option is selected for the field. Checkmark groups can be also sent to field verification. You can select whether to send field to field verification always or only if it has highlighted characters or format errors.

You can also use double verification mode where field is verified by two different operators. If merged verification results match the document is sent to the next stage. Otherwise additional check is carried out.
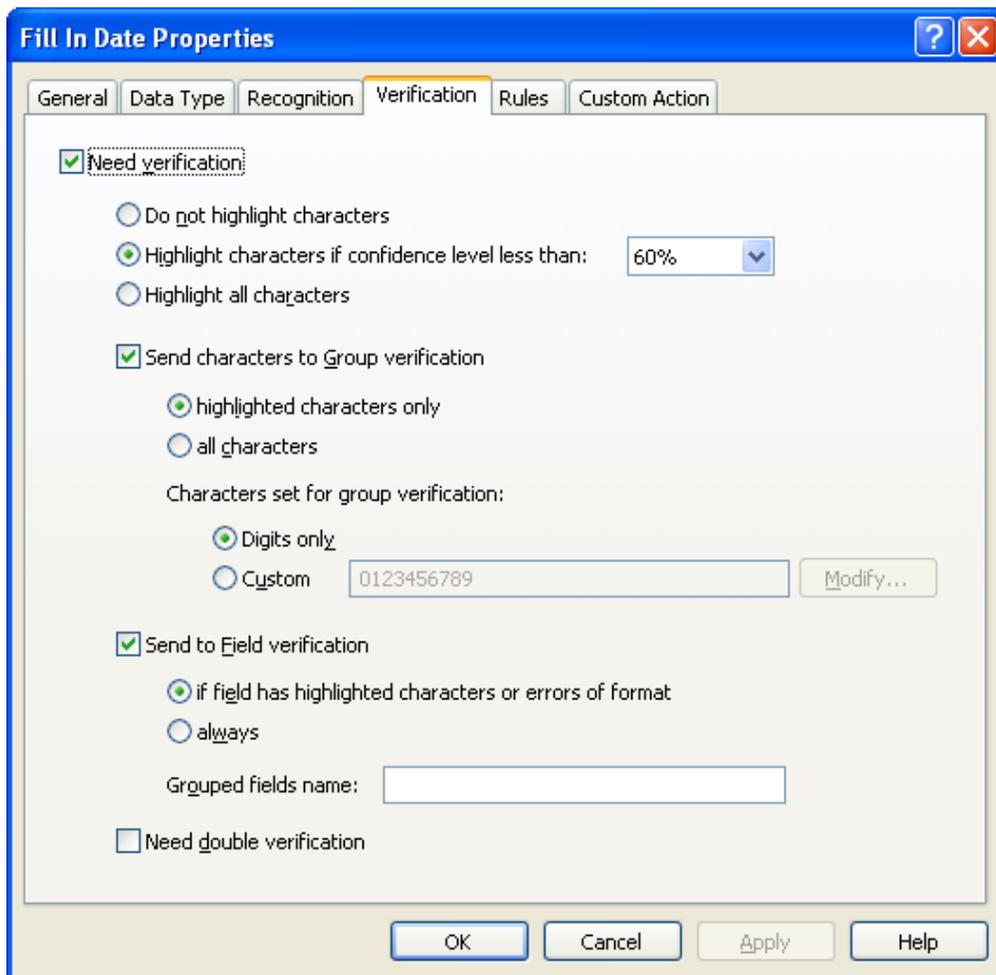


**Figure 11. Field Properties dialog box, Verification tab**

21

## 3.1.3.5.    Image Export Parameters

In the **Image** Field Properties, you can specify export parameters, such as the exported file type, quality, color type and resolution. To do it, go to the **Export** tab of the Field Properties dialog box (on the field's shortcut menu, click **Properties…**). You can configure the following:

- **File type** (TIFF, JPEG, BMP, JPEG2000, PCX packbits, PNG).
- **Quality**. For TIFF, JPEG and JPEG2000 files, you can choose the exported file quality (best, high, normal, low).
- **Color type** – color, grayscale, halftone (b&w).

To specify the resolution of the exported image, select **Change resolution to** and enter your resolution or select a predefined resolution from the list.

## 3.1.3.6.    Rules Check

*Rules* are needed to automatically check recognized data. The rules, along with data types, allow imposing data restrictions, i.e. defining the conditions the field values must meet. If not, pages with such fields will be flagged and marked with the corresponding message. Rules are mainly used for Document Integrity Checks. Rules can also be used to process recognized data – combine several field values or replace recognized values with values from the database.

Rules Check is a low priority task coming after automatic value correction, checking data for field format compliance, and data validation (checking if the date falls within the specified period, the numbers, within the specified interval, etc.).

ABBYY FlexiCapture 10 allows specifying the following Rules:

- Check Sum – checks the sum of several field values. The sum is then compared to the specified number or value of another field. For example, if your document describes an employee's salary an bonus, with an overall income field, you can set a rule that will check if the salary and bonus sum matches the overall expenditure. A warning message will be issued.

- Compare Fields – compares several field values (including those from different sections). This rule can be used if your document has several fields with values that must be equal. If the values are not equal, the Rules Check error will be issued.

- Database Check – entered values are checked by being compared to the values from a certain.

- Merge Fields – merging of several field values. Field values can be separated by dots, spaces or other separators. For example, it may be more convenient to compile a date from already recognized fields (Day, Month, Year), separating them with dots, than to recognize it as a whole. The merging results can be stored in any field of the Document Definition. It is convenient to use fields with no marking for this purpose (see Fields with no marking).

- Sum in Figures – Sum in Words (Russian) – compares a numerical sum with the same sum written in words (Russian language only).

- Script – the user describes constraints with the help of a script. The script language is described in the application help file in detail. As the order of rules to be checked is not specified, a field cannot

be available from two Script rules at a time. It will be available from any number of rules only in read-only mode.

Rules are specified on the **Rules** tab of the Field Properties dialog box (Figure 12). The rules can affect the values of one or multiple fields.

Rule severity can be specified (choose either error or warning). The rule will be flagged red if an error occurs, and yellow, if a warning is issued.
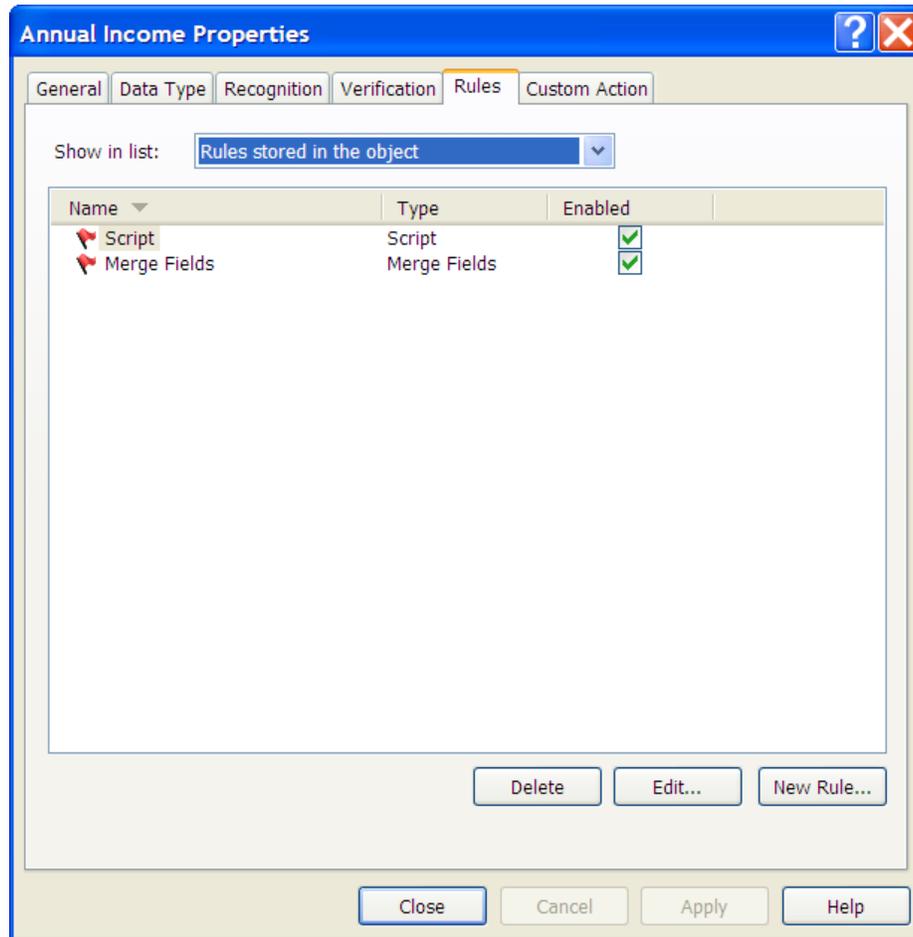


**Figure 12. Field Properties dialog box, Rules tab**

## 3.1.3.7. Custom action/script editing

It is possible to create a custom script. As opposed to the script action triggered automatically, the user-defined custom action is triggered by the user during field verification or when viewing the document in the Document Editor. The script is triggered by pressing the **…** button located to the right of the field value. The script language is described in the help file of the application.

Custom action is configured on the **Custom Action** tab (Figure 13). To add or delete fields available in the custom script, use the **Add…** and **Delete** buttons respectively. To open the Script Editor, click **Edit…**. To enable the custom action/script, select **Enable Custom Action**.
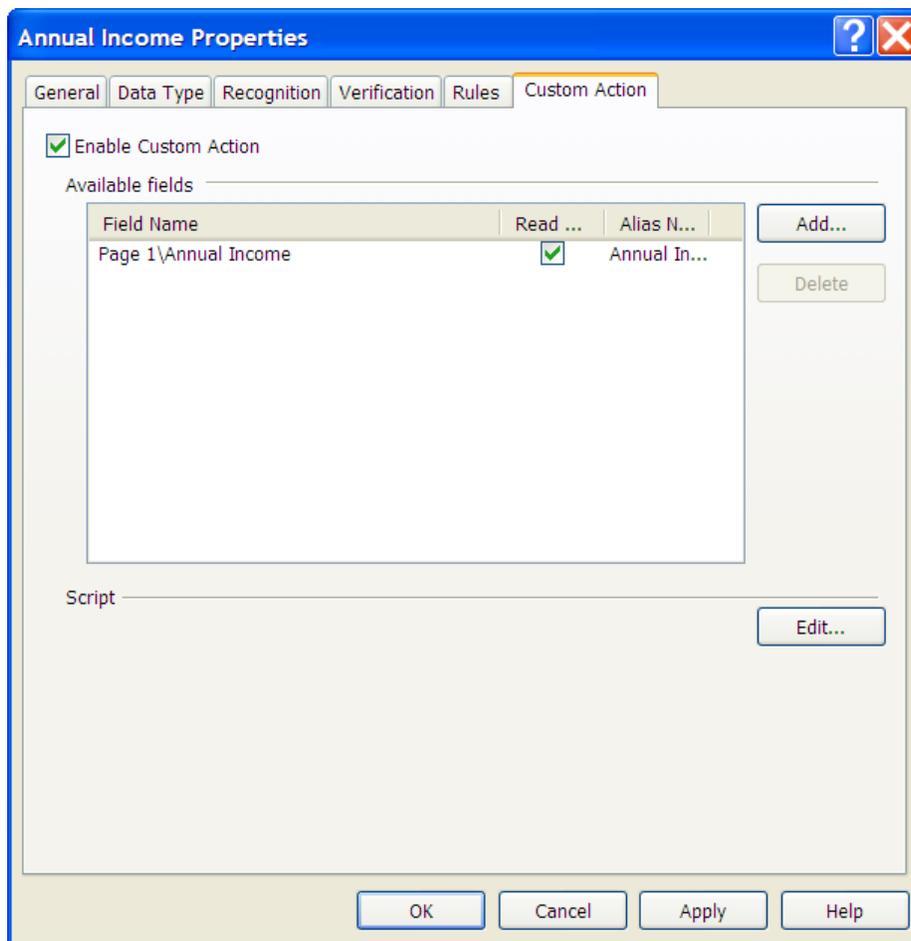
**Figure 13. Field Properties dialog box, Custom Action tab**

### 3.1.4. *Creating a Document Definition for Multi-Page Documents*

ABBYY FlexiCapture 10 allows creating multi-page Document Definition. A definition can consist of a random number of *sections*, each of them containing one or multiple *pages*. Section order, quantity and the document assembly rules are specified for multi-page Document Definitions.

Document Definition Creation Wizard's work will result in a simple Document Definition (without sections) appearing. You can add pages manually. When a page is added, its image will appear in the image area in the Document Definition Editor window. Fields created on this page will be added to the general list of fields. In this case, the document will consist of one section with several pages.

You may want to add several pages to one section if the document contains blocks stretching over multiple pages, repeating blocks or tables that occupy more than one page. The same method is used when loading a multi-page FlexiLayout (see Creating a FlexiLayout). All fields of the FlexiLayout will have to belong to the same Document Definition.

You can also add sections to a Document Definition. In the simplest case, one section contains one page. This may be useful if field sets on the pages are independent. This method is rather transparent: you can easily see the section the fields belong to, and set your own document assembly structure. For example, your document has 3 pages, with the 1$^{st}$ and 3$^{rd}$ page occurring once, and the 2$^{nd}$, 2-5 times. You will have

to create a separate section for each page, and then set the document structure, i.e. section order and number of repetitions.

A more complex case is a definition containing several sections that include more than one page each. For example, it can be a Document Definition consisting of non-flexible section and a multi-page flexible section, or a Document Definition describing documents consisting of a double-faced page that can repeat a certain number of times. Assume you have to create a Document Definition with a title page (page A) and a number of double-faced pages (page order B,C,B,C…). In this case, create the A section containing one A page and the BC sections containing two pages: B and C. The document structure should be as follows. The A section has one repetition followed by the BC section that repeats 3-7 times. Pages can also be repeated within a section. Multiple repetitions should be used wisely so as not to cause misassembly of the documents.

For Document Definitions consisting of several sections, *key fields* check can be enabled, in addition to order and the number of repetitions. In each document section, you will have to define the key fields with matching values in different sections. For example, you can take the name of a person who fills in the document, and make it a key field. In this case, the name must be written on every page. You can also use an identification number of some sort.

During batch processing, the program will attempt to assemble successive pages into documents while checking assembly rules. Key field values will also be verified. A warning message will appear if key field values don't match on pages of one and the same document. It may sometimes happen when document pages get mixed up during scanning. In this case, you need to change page order, for the document assembly rules to be met.

To add a page to a section, on the Document Definition Editor menu, click **Document Definition > Add Page…** or click **Add Page…** on the shortcut menu of the image. Load the image of an empty page and select object types to be marked on the page automatically.

To add a new section, on the Document Definition Editor menu, click **Document Definition > Add Document Section…**. The **Create New Section Layout** window will open, in which you will specify all the necessary parameters. Follow the on-screen instructions and:
- Specify the section name;
- Load an image of an empty page (scan or load from file);
- Load a FlexiLayout, if necessary;
- Select the types of objects to be marked automatically.

The new section and its fields will appear in the Document Structure window.

To view the resulting document structure when creating a Document Definition, click the **Pages** tab in the in the **Document structure** window on the right side of the Document Definition Editor window. The document structure will be presented with page thumbnails. On this tab, you can change the number of sections in the document. To do it, enter a new value (number or range) in the square to the right of the section name. You can also add annex pages, change page order in sections, and move pages from one section to another using the drag-and-drop method.
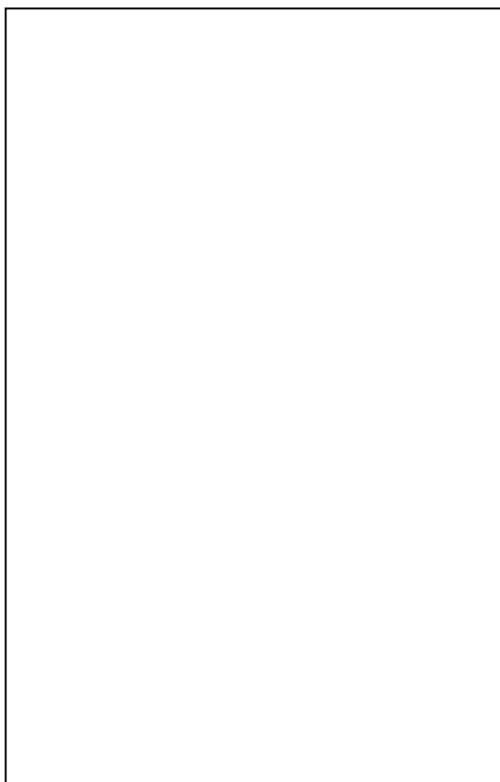
**Figure 14. Document structure, Thumbnails view mode**

You can also specify the order and number of section repetitions by clicking **Document Definition > Document Definition Properties**… in the Document Definition Editor window.

On the **Assembly** tab (Figure 15), specify the minimum and maximum number of section repetitions in the document (the default number is 1).

If you want to check the key field values, select **Check equality of key fields** and then select a key field for each section. Selection is made by double left-clicking.

In some cases, it may be necessary to disable the sections' order check, for example, if it is not critical for document assembly. To do it, select **Disable sections order check**. The presence of all sections in the document will be checked anyway, and the sections' order, will not.
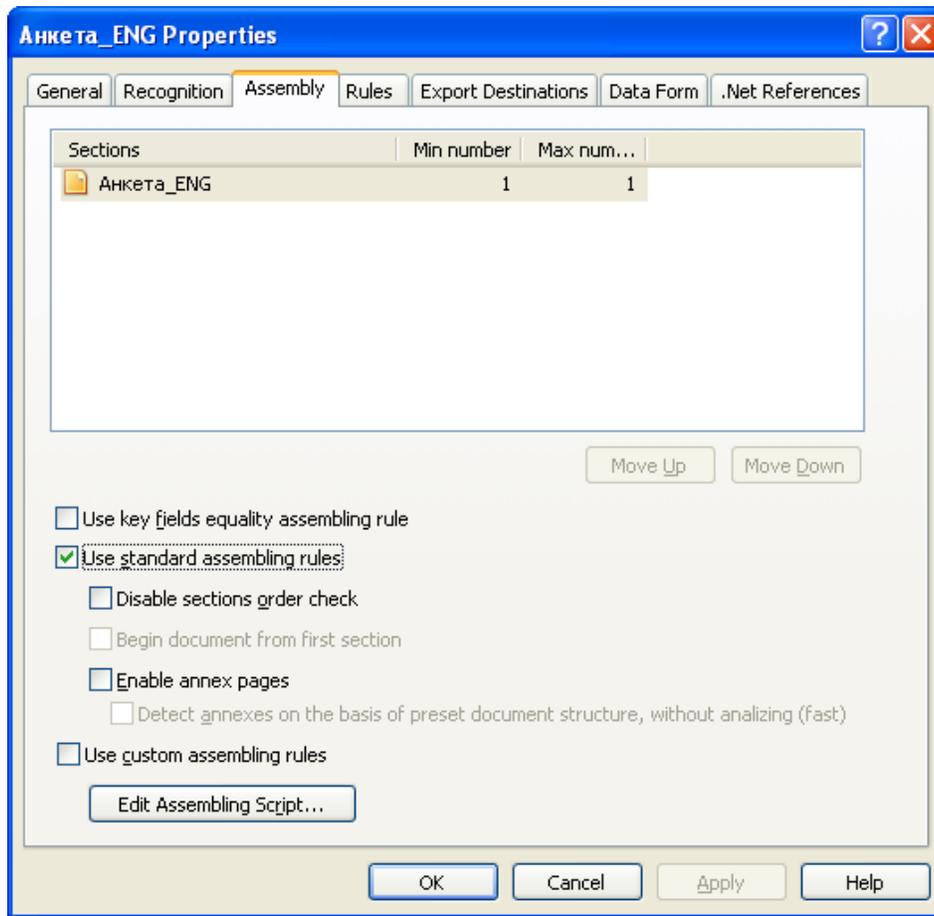
**Figure 15. Document Definition properties, Assembly tab**

## 3.1.5. Creating a Document Definition with Annex Pages

ABBYY FlexiCapture 10 allows creating Document Definitions for documents with annexes.

Annex pages can accompany any document. Separate fields are not required to be searched on these pages, so you don't need to match Document Definitions for them. However, they are taken into account during document assembly. For example, a credit application form (structured document) can be accompanied with a reference letter written in free form. This reference can be processed as an annex page.

To create a Document Definition with an annex page? Do the following:
- In the **Document Structure** window on the right side of the Document Definition Editor window, click the **Pages** tab and select **Enable annex pages in documents**. Enter the number (or range) of annex pages in the square that appears to the right of the section name (Figure 14), or
- In the Document Definition Editor window, click **Document Definition > Document Definition Properties**… and open the **Assembly** tab. Select **Enable annex pages**. Then, specify the minimum and maximum number of annex pages (Figure 15).

To save the image of an annex page, enable image saving. To do it, in the Export Settings window, click **Edit** and, in the dialog box that opens, click **Save document images**. Click **Image Format Options** to save the image to a graphical format or a searchable PDF.

## 3.1.6.  Export Settings

To configure saving of data obtained during document processing, you need to configure export for each Document Definition. There are four export types: to a file of the specified format, to an ODBC-compatible database, to a MS SharePoint library, and custom export (using a script). Export is configured in the **Export Settings** dialog box (**Document Definition > Export Settings**). To add a new export destination, click **Add…**. To edit an existing export destination, click **Edit…**

Aside from settings in the Document Definition, export can be configured at batch and project levels. Export configured at batch level is used for default batches. These settings will be used by default, when an Operator starts the export by clicking **File>Export To**.

New export destinations are added using the Export Destination Wizard. First, you will have to specify the general settings (the type and acceptance requirements), and then, the settings of the selected type.

You can specify several export destinations for one document definition. An export destination may be or not be critical. If an error occurs when exporting to a critical destination, the process terminates. If an error occurs when exporting to a non-critical destination, the process continues, and the document will be exported. It is thus recommended to have at least one critical export destination, or export errors will remain unnoticed. You can make an export destination critical in the **Export Settings** window.

Page images can also be exported – together with data or as a separate export procedure. You can specify the sensitive information to be redacted for each export destination. To export images together with data, select **Export to image files** in the Type dropdown box on the first page of the Export Destination Wizard dialog box.

> **Note**. If you are using a distributed version, you must specify file paths in UNC format (\\server_name\share_name[\path]) in the uploaded project for the export to complete correctly. If relative paths are specified, the export results will only be available using the application interface (API). If absolute paths are specified in the <disk>:\[path] format, the export will terminate with an error.
>
> If you have specified relative export paths, you can specify the root path in UNC format when uploading the project to the server. In this case, the relative paths will be formed basing on it.

You can configure the destination in such a way, that only documents without errors, documents with errors, or all documents will be exported. Use the **Documents without errors** and **Documents with errors** in the **Export Settings**.

Export can be carried out automatically or at an Operator's command. If you are using a distributed version, you need to create an export confirmation queue in the Project Settings to allow the Operator to start the export. If you are using a standalone version, select **Enable auto export** in the **Export Settings** window.

### 3.1.6.1.  Exporting to a file

If you want to configure exporting to a file, select **Export to data files** in the **Type** field. Specify the general export settings and click **Next**.

Enter the folder path for storing exported documents in the **Path** field or click **Browse…** to use the folder tree view.

You can export the documents of one batch to one file, or each document to a separate file (select **Create separate file for each document**). You can also choose to create a separate folder for each batch (select **Create separate folder for each batch**).

Select **Overwrite existing files** if you want newer files to overwrite the existing ones with the same name during export.

Next, specify the file naming options for the exported files. To do it, click **File Naming Options…** and select the necessary options.

Click **Next** to proceed.

Select the file type (CSV, DBF, TXT, XLS, or XML) and adjust additional properties, for example, you can change text encoding.

Click **Next** to proceed.

If images are exported together with data, configure the image export parameters (as described in the Exporting Images section).

Enter a name for the export destination and click **Finish**.

## 3.1.6.2.    Exporting to a database

To enable and configure export to a database, select **Export to ODBC-compatible Database** in the **Type** field. Select the necessary options and click **Next**.

Enter a **Connection string** or click **Setup Connection...** and configure the necessary parameters in the **Data Link Properties** dialog box. Then select a schema from the dropdown list.

To test connection to the database, click **Test Connection**.

Then, you will need to define the tables and table columns of the database the Document Description field values will be exported to. To do it, click **Setup Fields Mapping...**

In the left part of the **Field Mapping** dialog box, you will see the document's sections and fields. Match the tables and database fields, located in the right part, to them. If the database already includes tables for document export, select a database table for each section and a table column for each field. The table column format must correspond to the value of the exported field.

If the database doesn't have tables for document export, you can enable automatic database table creation at first export. To do it, click **Create Tables Automatically**. The program will create tables with the structure suitable for exporting. Table columns will be matched to document fields.

Field groups, fields with multiple repetitions, document tables, and sections are exported to separate database tables. The Primary and Foreign keys are used to link the main table to subordinate tables. In the main table, a primary key is added to each entry. In the subordinate table, each entry has a foreign key that records the value of the primary key of the corresponding entry in the main table. The primary and foreign keys are added automatically where required. You only need to specify the field to which the key has to be exported.

The **Show linked columns** option is selected by default. If you clear it, the columns with configured export will not be displayed in the right part of the tree.

Click **Next** to proceed to image export settings.

You can save images to a database or as files. If you save images as files, configure the image export parameters as described in the Exporting images section.

> **Note.** If you chose to save images to a database, a BLOB-type table will be created for them. If you change image export later and choose to save them to a folder, an error will occur during export, because the text path to the image file will be exported to the BLOB-type field.

Enter a destination name and click **Finish**.

### 3.1.6.3.  Exporting to SharePoint

ABBYY FlexiCapture 10 allows exporting document to the Microsoft™ SharePoint library. Each document will have matching columns with values from document fields. These values can be used for document search and indexing.

**Notes**.

1.   To configure export to SharePoint, you must have the Administrator's rights. However, to carry out the export itself, the rights of a Contributor will be enough.
2.   SharePoint columns that are intended for data storage, must have either the *Single line of text* or the *Multiple lines of text* type.

To configure export to SharePoint, select **Export to SharePoint** from the **Type** field. Select the necessary options and click **Next**.

Enter the **Server URL** where SharePoint libraries are stored. Click **Connection Settings…** to configure authentication parameters (Windows Authentication is used by default) and proxy parameters, if necessary.

Click **Connect** to reconnect to the server, if the connection was lost.

**Select a SharePoint library** from the dropdown list**.**

**Select** the document content type in the **Content type** field (SharePoint 2007 or later). In this case, you will be able to configure value export to fields corresponding to the current value type. Click **Setup Fields Mapping…** and configure the necessary settings in the **Field Mapping** dialog box.

Click **Next**.

Configure the image export parameters and, again, click **Next**.

Enter a destination name and click **Finish**.

### 3.1.6.4.  Exporting images

To configure image export, select **Export to image files** in the **Type** field. Select the necessary options and click **Next**.

Enter the path to a folder for storing exported images in the **Path** field or click **Browse…** to use the folder tree view.

You can export documents from one batch to one file (select **Create separate folder for each batch**). Select **Overwrite existing files** if necessary.

Configure the file naming options. To do it, click **File Naming Options…** and select the necessary criteria in the dialog box that opens.

If you need to redact confidential information from exported images, select **Redact sensitive data from images**. Then click **Select Fields to Redaction…** and select the fields to be redacted.

Click **Next**.

Select the format for storing images in the **Type** field. Then configure the quality and color type of the images saved.

If you need to change the initial image resolution in order to reduce file size, select **Change resolution to** and enter a resolution.

If you select the PDF format and select the **Create searchable PDF** option, the document image will be searched in full-text mode, and the recognized text will be saved in the selected format. In this case, you will be able to specify the recognition language: you can either choose to keep the language that was used in the Document Definition, or select one or more languages from the lust (click **Select…**).

Click **Next**.

Enter a destination name and click **Finish**.

## 3.1.6.5. Custom export (script)

Using the scripting method, you can create a custom export procedure which will be more flexible than if you were using standard methods.

To create a custom export procedure, select **Custom export (script)** in the **Type** field. Select the necessary options and click **Next**.

Select a scripting language (**JScript®**, **VBScript**, **JScript.Compact**, **C#.Net**, **Visual Basic.Net** or **JScript.Net**) and enter the scripting text in the Script Editor that appears when you click **Edit Script…** (for detailed description of the scripting language with examples, see the help file).

## *3.1.7.* *Configuring Recognized Data Presentation*

After the data is recognized, the Operator will see it in the document window. The data will be displayed as a list by default, with captions matching the names of Document Definition fields. However, it is also possible to change data display and arrange it in a convenient manner or change element headings. Data presentation can be edited in the bottom right corner of the Document Definition Editor window.

Fields can be moved, renamed and otherwise altered. To add text, click **Insert Label Box** on the shortcut menu.

Text font and size for field names and recognized data values can be specified on the **Data Form** tab of the Document Definition properties dialog box (**Document Definition > Document Definition Properties…**).

## *3.1.8.* *Testing Document Definitions*

Before starting mass document input, you need to test the created Document Definitions.

Document Definitions can be tested directly from the Document Definition Editor using the image the definition was created on. To start testing, click **Testing>Run Test**.

Testing involves documents added to test batches, because they work with the unpublished version of the Document Definition. These batches can be accessed from the main window of the program (**View**>**Test Batches**). You can also access test batches from the Document Definition Editor by clicking **Batch>Switch to Test Batch…**.

Scripting rules can be checked directly from the Script Editor – click **Tools>Check Script**. This function can be used to check script activity with different field values.

If rule errors occur during testing, or if invalid field property values are found, edit the Document Definition to correct these errors. When all errors are corrected, you can publish the Document Definition and proceed to mass document input.

### 3.1.9.    Editing and Publishing a Document Definition

After creating a Document Definition and successfully testing it on several images, publish it to make it accessible for recognition of working batches.

To publish a Document Definition, click **Publish** in the **Document Definitions** window (**Project > Document Definitions...**).

To go back to editing the Document Definition, select it in the **Document Definitions** dialog box and click **Edit…**. The Document Definition will not be available for editing to other users, and the latest published version will be used for recognition. Thus, when you edit a Document Definition, you will not be in the way of Operators' work. Publish the Document Definition when you are finished editing it, and the new version will become available to the users. If you don't want to publish the edited version, you can discard all changes and return to the latest published version. To do it, click **Discard Changes** in the **Document Definitions** dialog box.

When a new version of the Document Definition is published, it will be used to recognize documents. For separate documents that were previously recognized, you can change the used Document Definition version without re-recognition (only if the version changes didn't involve changing field locations and recognition properties (export settings, for example, can be changed)). To do it, click **Tools > Update to Latest Version** on the main menu.

## 3.2.   Creating a FlexiLayout

Configuring FlexiLayout processing is a more elaborate task. Invoices, orders and other similar documents are not structured (the location, size, and number of fields are different on different copies), so it is impossible to create a common layout with fixed field positions. In the case of such documents, fields are searched after prerecognizing the whole page, and the Document Definitions are created using keywords and relative field locations (these definitions are called FlexiLayouts). A special program, ABBYY FlexiLayout Studio 10 was designed to create FlexiLayouts. For more information on FlexiLayout Creation, see the ABBYY FlexiLayout Studio 10 help file.

The FlexiLayout you created will be exported to an .afl file and then used in ABBYY FlexiCapture 10. The FlexiLayout will be used as one of the Document Definition sections. Thus, a Document Definition can include both structured and non-structured sections. FlexiLayouts can consist of multiple pages.

FlexiLayouts can be attached to a Document Definition when creating the latter. To do it, at the second step of the Document Definition creation (Create New Document Definition Wizard), add the document image, and then select **Load FlexiLayout** and enter the path to the .afl file with the FlexiLayout.

You can also load a FlexiLayout from the Document Definition Editor using the Properties dialog box. To open it, click **Properties**... on the shortcut menu of a section. Click the **FlexiLayout** tab and then click **Load**.... Select the required .afl file with the FlexiLayout.

When you load a FlexiLayout, all fields and their geometric marking will present on the image.

The number of fields in the Document Description must correspond to the number of fields in the FlexiLayout. It is not recommended to add or delete fields (except for fields with no area on the image), the set of recognized fields can be changed only by changing the FlexiLayout itself.

Now you need to specify field properties just like for a structured document. Configure the recognition, verification, and export properties as well as rules.

### 3.2.1. Classifiers

A **Classifier** is a special ABBYY FlexiLayout Studio project designed for preliminary identification of separate pages and for labeling them as a certain type (for example, by the FlexiLayout/ FlexiLayout variant used). Classifiers are used in ABBYY FlexiCapture to classify documents at batch type level (for Default batches, the classifier will be added at project level).

Classifiers can be used to:
- Speed up the process of selecting a Document Definition. Using a classifier allows choosing a Document Definition according to classifying signs. You won't have to go through all Document Definition to choose the necessary one. An increase in speed will be especially noticeable when working with large projects containing many Document Definitions.
- To distribute the classifying features in the classifier tree. For example, at the topmost level, you can select documents sent in by a certain company based on a document section that bears the name of that company. On the next level, the documents can be classified based on their titles or based o the number of separators on a particular page.
- To manage the application of third-party FlexiLayouts. FlexiLayouts created by a third-party developer may not match properly their documents. In this case you can use a classifier to remedy the situation without contacting the developer.
- To speed up projects that contain Document Definitions in multiple languages. If a classifier is used, pre-recognition of the documents is performed only twice: when classifying the documents and when applying the selected Document Definition. Otherwise, pre-recognition has to be done for each Document Definition language in the project.

For detailed information about classifiers, see the ABBYY FlexiLayout Studio help file.

Classifiers can be added at batch type level or at project level (the latter case is only applicable to batches of Default type.
To add a classifier, do the following:
- In the main window, click **Project > Project Properties…** to add classifiers at project level, or click **Project > Batch Types…**, and then **Edit…**) to add a classifier at batch type level;
- In the Batch Type Property dialog box, click the **Recognition** tab, and then click **Load…**;
- Select a \*.cfl file with a classifier.

## 3.3. Specifics of Non-structured Documents

If you need to process non-structured documents with information presented in free form, for example, contracts, letters, orders, appendices, ABBYY FlexiCapture 10 will provide a solution again. Non-structured documents with text or images, separated by empty pages or pages with a barcode, are processed and exported to searchable PDF files or files in image formats.

Processing of such documents usually includes converting them to electronic form and organizing their search based on key field values.

As a rule, key fields (for example, a contract number) are searched using a FlexiLayout created in ABBYY FlexiLayout Studio.

If automatic key field search is impossible, the values can be entered by an Operator. To allow it, create a Document Definition with one field (or several, if necessary) and, in the Recognition Properties of the field(s), select **Don't recognize (Key from Image field – will be entered manually)**. The Operator will be prompted to enter the key field value manually when verification is started.

To store documents, you need to configure the export. Key field values can be exported to files or a database, and document images can be saved in the desired format. To export images, on the **Image** tab of the **Export Settings** dialog box, click **Save Document Images** and configure the image saving parameters. You can save document images to graphic files or searchable PDF documents.

Consider page assembly into documents carefully. In the case of non-structured documents, it may be hard to define which document a page refers to. It is recommended to separate documents with empty pages or pages with barcodes to allow automatic assembly of non-structured documents. Then, when adding images to a batch (from a scanner, file, or during Import Profile creation), select **For images separated by** and, from the dropdown list, choose either **blank pages** or **pages with barcode**, depending on the pages used as separators. Pages will be assembled into documents automatically – pages will be added to the current document until a separator page is encountered.

# 4.   Configuring Project Properties

Project properties can be configured in the **Project Properties** window. To open it, click **Project > Project Properties…**

The Project Properties window has seven tabs. The **General** tab describes general project properties – the name, description, and other. The **Stage type** tab allows you to set up the processing of events at different processing stages and their appearance.

The rest of the tabs are identical to those in the **Batch Type** properties window (see Configuring Batch Types) and describe the properties of Default-type batches.

The **General** tab allows configuring the following project properties:
- **Name**;
- **Description**;
- **Export root path** — path to a folder, from which relative file export paths will stem after a project has been uploaded to the server. The path is specified in UNC format.

You can also view the project file name and its location in the folder tree (or for projects on the Application Server, its GUID).

The rest of the tabs are identical to those in the Batch Type properties window and are used to configure the Default batch type.

When configuring project properties, you can specify the order list of the Document Definitions used. As opposed to the batch types, this operation is carried out from the **Document Definitions** window (**Project > Document Definitions…**).

FlexiLayouts are matched in the same order they are listed in this window (top to bottom). To change the order of a document, use the **Up** and **Down** buttons.

As a rule, the recognition process consists of consecutive matching of all added FlexiLayouts and checking their quality, after which best one is selected. To speed up Document Definition matching, select **Use first**

**acceptable flexible layout**. In this case, if a FlexiLayout is matched, the rest won't be considered. This option may be useful when you have many different FlexiLayouts as it will decrease the overall recognition time.

A Document Definition can be temporarily disabled. To do it, clear the **Enabled** option in the list of Document Definitions. To enable the Document Definition, select this option.

# 5.  Configuring Batch Types

ABBYY FlexiCapture 10 supports document processing in batches: a batch is created before processing, and document images are added to the batch to be processed together. Batches can be created automatically or manually during image import (from a scanner or image file). Batch processing settings are configured in the Batch type properties. You can create several batch types with the most commonly used processing settings. In this case, the Operator will only have to specify the batch type.

Batch types define the following document processing parameters:
- Image preprocessing method;
- List and matching order of the Document Definitions, and the Classifier used;
- Batch integrity check;
- Default settings for manual export;
- Export settings for unrecognized documents;
- Batch processing.

Batch types also define the list of batch registration parameters, i.e. textual information characterizing a given batch.

Batch types are created within projects. To view the list of batch types, click **Project > Batch Types…** on the main menu.

To create a new batch type, click **New…**. Then, configure the necessary parameters in the Create New Batch Type window.

To edit a batch type, select it and click **Edit…**. The **Batch Type Property** dialog box will open.

Batch type settings are divided into six groups with a separate tab for each of them. Each group also has a corresponding stage during batch type creation.

**Note:** If there are several independent workflows and Document Definitions for which monitoring and collecting statistics should be performed separately, it is recommended that different projects should be created. Otherwise, it is recommended that different batch types should be created within a project.

## 5.1.  General

The general properties are as follows:
- **Name**;
- **Description**;
- **Registration parameters**. The registration parameters are textual properties allowing identification of a given text. The scan date, the Scanning Operator's name or other information can be used as such properties. The registration parameters are specified when creating a batch of the selected type. Registration parameters can be either required or optional.

35

## 5.2.  Image Preprocessing

The following document creation and image processing parameters can be configured for a batch type:
- **Create new document**. A new document can be created automatically when Document Definition is matched, for every image file, or for image files separated by blank pages or pages with a barcode (of certain type or value). If the separator pages are not to be processed, they can be deleted by selecting **Delete separator pages**.
- **Image Processing**. The following image processing parameters can be specified:
  - ➢ **Despeckle images**;
  - ➢ **Convert color and gray images to black and white**;
  - ➢ **Rotate images by** (90º (counter)clockwise or 180º (upside down));
  - ➢ **Use black separators to correct skew;**
  - ➢ **Use black squares to correct skew;**

These settings will be used by default when an Operator adds documents to a batch of this type without using import profiles.

## 5.3.  Recognition

The following recognition parameters can be configured:
- **Document Definitions**. You can use all Document Definitions in the project, or select the ones to be used for the current batch type. If you use a list of Document Definitions, you can specify their matching order for the batch type: the definitions will be matched in the same order in which they are presented on the list (top to bottom). To change the order of Document Definitions, use the **Up** and **Down** buttons.
- **FlexiLayout**. As a rule, all loaded Document Definitions are matched consecutively. Then, matching quality is checked, and the one with the best quality is chosen. To speed up Document Definition matching, select **Use first acceptable flexible layout**.  In this case, if a FlexiLayout is matched, the rest won't be considered.
- **Classification Tree**. A classifier can be loaded for batch type or project to select FlexiLayouts. To select a classifier, click **Load…** and, in the dialog box that opens, select a classifier file. To detach the classifier, click **Detach**.
- **Delete blank pages during recognition** – select this option if you want blank pages to be deleted.

## 5.4.  Event Handlers

You may create your own scripts that will perform an action for an event. Simply select an event in the list and click the **Edit Script…** button. In the dialog box that opens, type the code of the script and debug it. Select **File > Save** or press Ctrl+S to save the script.

The script will be disabled by default. To enable the script, select the option to the left of its name.

## 5.5.  Export

The export parameters of recognized documents are specified for Document Definitions. For a batch type, you can specify the export parameters of unrecognized documents. To export the images of unrecognized

documents (i.e. those for which no matching Document Definitions have been found), select the **Export unrecognized documents** option and specify the export parameters.

You can also set up the program to delete documents upon export. Simply specify how long documents should be stored after export.

## 5.6. Workflow

### 5.6.1. Standalone

The following workflow parameters can be configured for the Standalone version:
- Automatic batch export, if there are only few uncertainly recognized characters. For the batches to be exported automatically, select **Export batch automatically if there are no errors or warnings, and percentage of uncertain symbols is equal or less than** and enter the maximum allowed percent of uncertainly recognized characters for the export to be carried out automatically.
- Document deletion after export. To delete documents after export automatically, click **Delete processed documents after export**.

### 5.6.2. Distributed

The following workflow parameters can be configured for the Distributed version:
- Automatic processing. In this mode, no human involvement is required and all operations are automatic: documents are imported, recognized, and exported. If an error occurs, failed documents are placed into the Exceptions queue.
- Regular processing. In this mode, you can select the processing stages to be used and allow or prohibit the use of Web stations at certain stages. You can also specify the following processing parameters:
  - ➢ **Process whole batches**. By default, jobs are created using a certain number of documents from one batch. If you select this option, however, each job will include an entire batch.
  - ➢ **Skip verification and data verification if percentage of characters need verification is equal or less than**. If you select this option, documents for which good recognition quality has been achieved, will not be included into verification jobs. Documents will skip verification if they contain no characters to verify or if the number of unreliably recognized characters is lower than the value you specified.
- Extended processing. In this mode, you can not only set up pre-defined processing stages, but also create new ones and change the order of the stages. You can specify various parameters for a stage, including the conditions when a document must be placed into a certain stage or routed. These conditions can be specified by logical rules or by scripts.

## 5.7. .Net Assemblies

If required, you can connect .Net assemblies to a project. Simply click the **Add…** button and provide the path to the assembly.

# 6. Configuring Image Import

In the Distributed version, images are imported in a special application – Scanning Station. However, Verification Operators and Senior Verification Operators can also add images to batches. Images can also be automatically imported from Hot Folders.

Import profiles allow specifying the import and image processing parameters. In the Distributed version, import profiles are generally used to process Hot Folders, because Verification Operators add images to correct assembly errors, and not to process streaming documents. In the Standalone version, an Operator can also use import profiles to quickly import images using one button.

To add an import profile, click **Project > Image Import Profiles…** on the main menu of the program. In the dialog box that opens, you can create new import profiles, and edit, delete and copy the existing ones (Figure 16).

To create a new import profiles, start the Import Profile Wizard by clicking **New…**

First, choose the image source. You can either choose to get an image from a scanner or a Hot Folder (the folder in which new images will be searched). If you want to add images from a Hot Folder, make sure you have the necessary rights to open it and write to it. If you choose a scanner as the image source, make sure it is connected properly.

**Creating an import profile using a scanner**

1. Select **Scanner** as the image source and click **Next**.

2. Configure the scanning and image processing options. Select the method of interacting with the scanner from the **Style of settings dialog** dropdown list: select **FlexiCapture scanning settings** to use the dialog box from ABBYY FlexiCapture or **System scanning settings** to use the dialog box from the scanner driver. Click **Next**.

3. Configure the document assembly and image preprocessing settings. For example, you can choose to despeckle images (if the image quality may be low or the image may contain rubbish), convert them to black and white, rotate, and disable skew correction. Click **Next** when you are finished.

4. Now you can change the default import profile name and add a comment.

**Creating an import profile using a Hot Folder**

1. Select **Hot Folder** as the image source and enter the path. You can choose a POP3- or FTP-server, mailbox or local or network folder as a **Hot Folder**. Then click **Next**.

2. Configure the image loading options. You can select **Check Hot Folder every** and enter the time interval between image checks. In **Batch Settings**, select one of the options to define the batch or batches the new images will be added to. In the **Number of files to add** dropdown list, select of the values to define the number of files to be added to the specified or newly created batch. Click **Next**.

3. Configure the document assembly and image preprocessing settings, just like if you were creating an import profile using a scanner. Click **Next** when you are finished.

4. The Import Profile Wizard will prompt you to configure Hot Folder purge options: successfully imported images as well as images with import errors can be deleted or retained/moved. Click **Next** when you are finished.

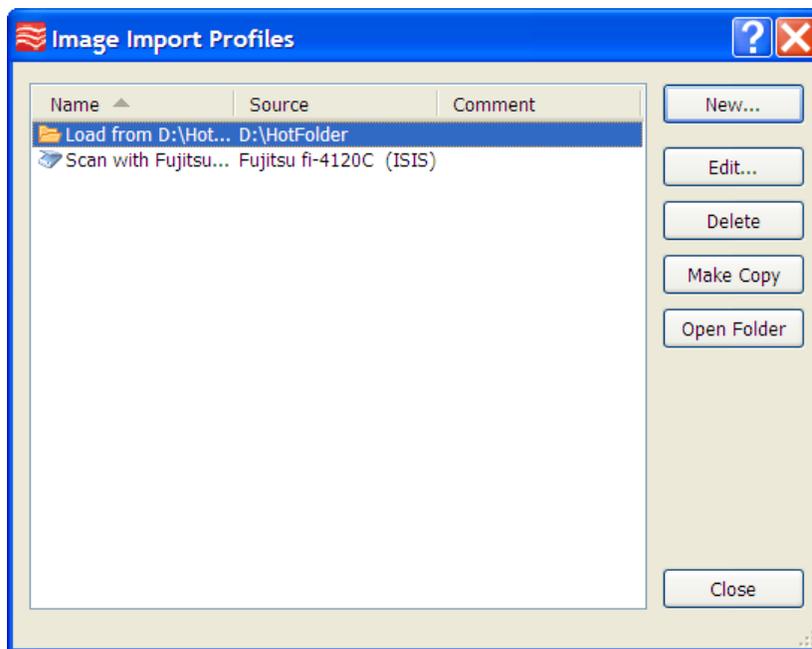5. Now you can change the default import profile name and add a comment.

**Figure 16. Image Import Profiles dialog box**

Automatic Hot Folder checking is initiated as follows:

- In Standalone version – by selecting **Automatically check Hot Folders** in the **Image Import Profiles** window.
- In Distributed version – for individual projects, using the **Processing Server Monitor**, in the **Hot Folders** section.

# 7. Uploading a Project to the Server

In the Standalone version of the system, Operators can start working with a project as soon as it is configured. In the Distributed version, the created project must be uploaded to the Application Server. To do it, click **File > Upload Project to Server…** and, in the dialog box that opens, enter the Application Server address in the http://<servername> format. Only the Application Server Administrator can upload the project to the server.

When the project has been uploaded to the server, you must assign roles to Operators using the **Administration and Monitoring Console**. If Hot Folders are used in the project, enable them using the **Processing Server Monitor**. For detailed information on administering the system, see the Administrator's Guide.

# 8. Keyboard Shortcuts
## 8.1. Main window

| | |
|---|---|
| New Project… | Ctrl+Shift+N |
| Open Project… | Ctrl+Shift+O |
| Save | Ctrl+S |
| New Batch… | Ctrl+N |

| | |
|---|---|
| Load Images… | Ctrl+O |
| Scan Images… | Ctrl+K |
| Import Images | Ctrl+I |
| Export | Ctrl+U |
| Export Data to Files… | Alt+Shift+S |
| Export to Database… | Alt+Shift+D |
| Undo | Ctrl+Z |
| Redo | Ctrl+Y |
| Cut | Ctrl+X |
| Copy | Ctrl+C |
| Paste | Ctrl+V |
| Delete | Del |
| Select All | Ctrl+A |
| Find… | Ctrl+F |
| Find Next | F3 |
| Go to Next Document | Ctrl+D |
| Go to Previous Document | Ctrl+Shift+D |
| Despeckle Image | Ctrl+Alt+K |
| Invert Image | Ctrl+Alt+V |
| Rotate Image 90º clockwise | Ctrl+W |
| Rotate Image 90º counterclockwise | Ctrl+Shift+W |
| Test Batches | Ctrl+Shift+B |
| Batches | Ctrl+B |
| Details | Alt+1 |
| Thumbnails | Alt+2 |
| Full Screen | F11 |
| Refresh | F5 |
| Properties… | Alt+Enter |
| Document Editor: Image Mode | Ctrl+1 |
| Document Editor: Data Mode | Ctrl+2 |
| Zoom In | Ctrl +Num+ |
| Zoom Out | Ctrl+Num– |
| Thumbnails: Increase Size | Ctrl+Shift+Num+ |
| Thumbnails: Decrease Size | Ctrl+Shift+Num– |

| | |
|---|---|
| Analyze | Ctrl+E |
| Match Document Definition… | Alt+Shift+E |
| Recognize | Ctrl+R |
| Run Verification | F7 |
| Next Item to Verify | F4 |
| Previous Item to Verify | Shift+F4 |
| Next Assembly Error | F9 |
| Previous Assembly Error | Shift+F9 |
| Next Uncertain Character | F8 |
| Previous Uncertain Character | Shift+F8 |
| Next Rule Error | F6 |
| Previous Rule Error | Shift+F6 |
| Document Definitions… | Ctrl+T |
| Batch Types… | Ctrl+Shift+T |
| Image Import Profiles | Ctrl+Shift+I |
| Update to Latest Version | Alt+Shift+U |
| Re-analyze | Ctrl+Alt+E |
| Re-recognize | Ctrl+Alt+R |
| Re-check Rules | Ctrl+F6 |
| Check Batch Integrity | Ctrl+Shift+E |
| Help Topics | F1 |
| Exit | Alt+F4 |

## 8.2. Document Definition Editor Window

| | |
|---|---|
| Add Page… | Ctrl+P |
| Add Document Section… | Ctrl+Alt+S |
| Document Definition Properties | Ctrl+T |
| Save | Ctrl+S |
| Undo | Ctrl+Z |
| Redo | Ctrl+Y |
| Cut | Ctrl+X |
| Copy | Ctrl+C |
| Paste | Ctrl+V |

| | |
|---|---|
| Delete | Del |
| Delete Region | Shift+Del |
| Select All | Ctrl+A |
| Select by Type | Ctrl+Shift+A |
| Group | Ctrl+G |
| Ungroup | Ctrl+Shift+G |
| Copy Text from Image | Ctrl+Alt+C |
| Create Field: Text | Alt+Shift+T |
| Create Field: Checkmark | Alt+Shift+C |
| Create Field: Checkmark Group | Alt+Shift+M |
| Create Field: Barcode | Alt+Shift+B |
| Create Field: Picture | Alt+Shift+P |
| Create Field: Table | Alt+Shift+L |
| Create Field: Group | Alt+Shift+G |
| Despeckle Image | Ctrl+Alt+K |
| Invert Image | Ctrl+Alt+V |
| Rotate Image 90º Clockwise | Ctrl+W |
| Rotate Image 90º Counterclockwise | Ctrl+Shift+W |
| Data Form | Alt+F2 |
| Properties | Alt+Enter |
| Static Elements | Ctrl+1 |
| Field Regions | Ctrl+2 |
| Show Document Structure | Alt+F1 |
| Fields | Alt+1 |
| Pages | Alt+2 |
| Zoom In | Ctrl+Num+ |
| Zoom Out | Ctrl+Num– |
| Run Test | Ctrl+R |
| Autodetect Objects | Ctrl+D |
| Check Document Layout | F9 |
| Match Section | Ctrl+E |
| Help Topics | F1 |
| Close | Alt+F4 |

## 8.3.  Group Verification Window

| | |
|---|---|
| Confirm All | Enter |
| Postpone All | Ctrl+Enter |
| Toggle | Space |
| Next Page | Page Down |
| Previous Page | Page Up |
| Undo | Ctrl+Z |
| Redo | Ctrl+Y |
| Select All | Ctrl+A |
| Show Character Image | F2 |
| Full Screen | F11 |
| Show Field Image | Ctrl+I |
| Show Field Image: On Top | Alt+1 |
| Show Field Image: On Bottom | Alt+2 |
| Image Scale: Zoom In | Ctrl+Num+ |
| Image Scale: Zoom Out | Ctrl+Num– |
| Help Topics | F1 |
| Exit | Alt+F4 |

## 8.4.  Field Verification Window

| | |
|---|---|
| Confirm Field | Enter |
| Postpone Field | Ctrl+Enter |
| Next Uncertain Character | F4 |
| Previous Uncertain Character | Shift+F4 |
| Next Field | Page Down |
| Previous Field | Page Up |
| Undo | Ctrl+Z |
| Redo | Ctrl+Y |
| Cut | Ctrl+X |
| Copy | Ctrl+C |
| Paste | Ctrl+V |
| Delete | Del |

| | |
|---|---|
| Delete All | Alt+Del |
| Select All | Ctrl+A |
| Insert Line Break | Shift+Enter |
| Merge Characters | Ctrl+M |
| Analogous Fields | Alt+F3 |
| Show Character Image | F2 |
| Full Screen | F11 |
| Recognized Text | Alt+F1 |
| Character Image Cutting | Alt+F2 |
| Show Field Image | Ctrl+I |
| Show Field Image: On Top | Alt+1 |
| Show Field Image: On Bottom | Alt+2 |
| Image Scale: Zoom In | Ctrl+Num+ |
| Image Scale: Zoom Out | Ctrl+Num– |
| Help Topics | F1 |
| Exit | Alt+F4 |