

Automating Tasks in ABBYY FlexiCapture 8.0 Professional

This document provides a general survey of the most common scenarios involving task automation in ABBYY FormReader 6.5 and their support in the successor product, ABBYY FlexiCapture 8.0 Professional.

Typical customization and integration tasks

Customization and integration tasks can be broken down into the following four groups:

1. **Automating routine operations**
2. **Customizing the GUI**
3. **Adding new functionality:**
 - **Custom data validation rules**
 - **Support for multi-page documents**
 - **Customized export**
 - **Customized import**
4. **Integration with other systems**

Of course, all of the above may be closely interrelated within a given project. For example, integration with another system may mean exporting extracted data into this system, which will require changes to the GUI (e.g. adding a button that calls an export settings dialog box).

Below we will take a closer look at each group of tasks.

Customization/integration in ABBYY products

1. Automating routine operations

This includes automated launch of required document processing steps and automating certain button clicks.

In ABBYY FormReader 6.5, you can use the API to automatically launch any document processing step.

In ABBYY FlexiCapture 8.0, the following operations can be automated:

- Loading images from a folder using so-called import profiles (Project → Image Import Profiles...)
- Launching OCR immediately after the images have been loaded (Tools → Options → Document Processing → Recognize added pages automatically)
- Removing processed documents after the data have been exported (Tools → Options → Document Processing → Delete processed documents after export)

2. Customizing the interface

Customizing the data window. Sometimes you may find it inconvenient to verify recognized field values displayed as one continuous list (especially in the case of questionnaires or test papers, which often contain many checkmarks). It may be more convenient to view the recognized fields either in the order in which they are arranged on the forms or by grouping certain fields together and supplying them with captions.

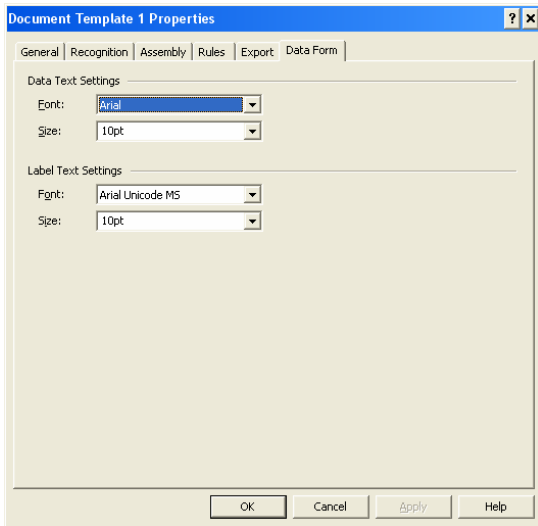
In ABBYY FormReader 6.5, the data window can be customized using a DLL created for a given template in Microsoft Visual Studio.

In ABBYY FlexiCapture 8.0, there is a built-in visual editor that changes the appearance of the data window. You can use drag&drop to rearrange the fields and supply them with text captions.

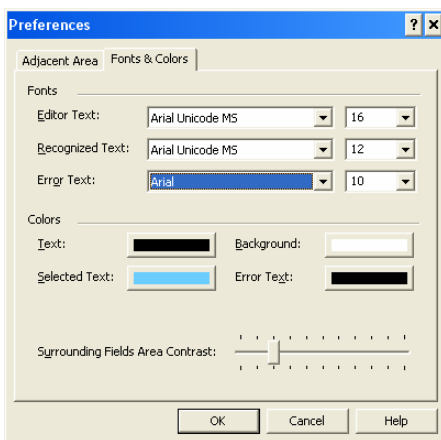
Useful tip

This release of FlexiCapture 8.0 does not support OCR of hieroglyphic scripts, but you already can use Chinese, Korean, Japanese and some other scripts in captions for elements in the data window. Simply select a Unicode font in the following two dialog boxes:

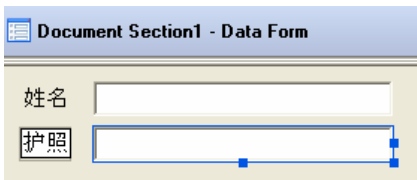
Document Template Editor → Template → Document Template Properties → Data Form:



and in the properties of the verification window. When verifying any text field, in the top menu select View -> Preferences -> Fonts and colours



To edit a checkmark, select the desired field and double-click the checkmark:



Similarly, you can create checkmarks:

1. In the context menu, create Label.
2. In the Label's properties, select a font whose name includes the word Unicode.
3. Type the required hieroglyphic characters.

Modifying the GUI for operator needs. A user-friendly GUI speeds up document processing and prevents incorrect operator actions. For example, you can modify the GUI so that a particular batch is automatically opened for the operator. Or you can disable batch creation and opening operations and restrict access certain administrative features. In still other cases, it may be desirable to create new buttons for one-click launch of certain recurrent operations.

In ABBYY FormReader 6.5, you can use the API to show/hide buttons and add your own toolbars, menus, and shortcut keys. You can even create your own GUI, run FormReader in invisible mode, send data from your own GUI to FormReader for processing, and then display the results in your own GUI.

In ABBYY FlexiCapture 8.0 release 1, you cannot yet create your own buttons. But already available are and Administrator Station and a Verifier Station with modified GUI.

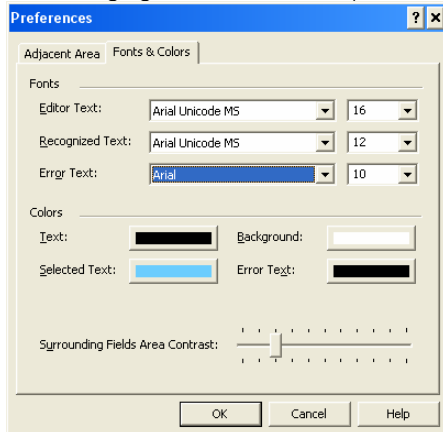
The modifications are as follows:

- The Operator Station displays only work batches (Project -> Batch List), whereas the Administrator Station displays both work batches and test batches.
- On the Operator Station, a command is available which switches the station into administrator mode (Project -> Switch to Admin mode).
- On the Operator Station, the project creation command is disabled. This command is only available on the Administrator Station (File -> New Project...).
- On the Operator Station, the template editing command is disabled. This command is only available on the Administrator Station (Project -> Document Templates).

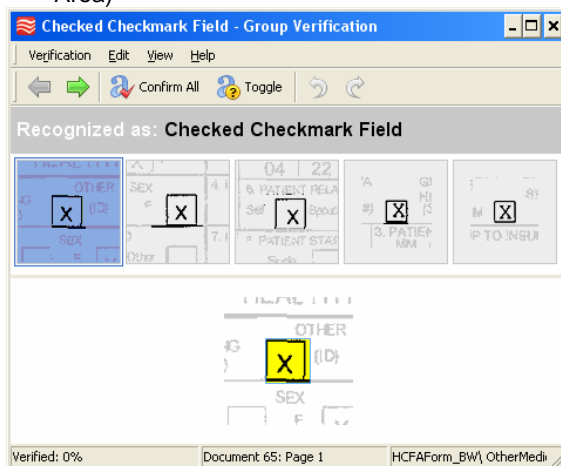
- On the Operator Station, import profile editing is disabled. This command is only available on the Administrator Station (Project -> Image Import Profiles).
- On the Operator Station, the password protection command is disabled. This command is only available on the Administrator Station (Project -> Protect with password).

What's new in the verifier interface

- The following options are available for Group Verification:
 - Confirm All (Enter)
 - Toggle (Space)
- The following commands are available for Field Verification:
 - ✓ navigation
 - Next/Previous Field (Page Up/Down)
 - Next/Previous Uncertain Character (F4/Shift+F4)
 - ✓ viewing several fields at a time (View -> Surrounding Fields)
 - ✓ changing fonts and colours (View -> Preferences -> Fonts & Colours)



- Settings common to these two modes:
 - ✓ changing the character size for extracted data (View -> Character Size)
 - ✓ displaying the relevant part of the source image (View -> Field Image -> Show Field Image) either above (On Top) or below (On Bottom) the recognition results.
 - ✓ one-time pop up of the relevant part of the source image (View -> Show Character Image (F2))
 - ✓ full screen mode (F11)
 - ✓ size and transparency of the area surrounding the recognized character (View -> Preferences -> Adjacent Area)



- many other settings to make information display more user friendly.

3. Adding new functionality

Custom data validation and conversion rules

Data validation is the main procedure that makes the entire data capture process more reliable and error-free. ABBYY products already include a number of ready-made validation rules that can quickly detect apparent typos and help the operator to correct them.

Apart from validation, rules can be used to convert data into a desired format, merge several fields into one string, etc.

Simple rules. Very simple rules can be used to compare values, to check the number of words, to check the length of a string or to check that a data format is valid. These may prove useful for practically any type of document.

Complex rules. These rules require connection to a data storage, a database, and external library of functions, etc. One example is checking a key field (e.g. passport number) against a database with the ability to modify the data in the database.

In **ABBYY FormReader 6.5**, all of the above can be implemented using Automation and VBScript rules.

In **ABBYY FlexiCapture 8.0**, the typical data format restrictions are already built into the corresponding data types (e.g. maximum and minimum length, changing all letters to uppercase or lowercase, replacing characters and strings, etc.). One advantage of this approach is that "invalid format" messages are displayed to the operator in real time as he/she is verifying the data in the recognized fields. For other checks and tips for operators, you can use Script rules, which can be created in the build in editor in one of the standard scripting languages, VBScript or JScript.

Both products include a built-in rule which checks extracted data against a database. This rule allows you to compare the value of a field with the corresponding database records to make sure that there is indeed such a database record (which has the same or a similar format).

Multi-page validation rules. These rules check relations between items located on different document pages.

In **ABBYY FormReader 6.5**, users had to use the API to create multi-page rules, because Automation checks only work within one page.

In **ABBYY FlexiCapture 8.0**, document-level Script rules are used for the purpose.

Multi-page documents

In **ABBYY FormReader 6.5 (Desktop Edition)**, a very simple model was used for multi-page documents: they were represented as a sequence of one-page templates. In other cases, users had to use the API to group separate pages into a multi-page document.

In **ABBYY FlexiCapture 8.0**, support for multi-page documents is implemented in the product itself.

Multi-page documents in FlexiCapture 8.0 Release 1

ABBYY FlexiCapture 8.0 Professional allows you to create templates for multi-page documents. A template can include any number of **sections** (one or several pages and the fields on them, which correspond to a logically complete part of the document). Document sections can be described using a fixed or flexible template with a variable number of pages. For templates of multi-page documents, you must specify the order of the sections, their number, and the rules used to assemble the pages into documents. For a template consisting of several sections, you can enable verification of key fields in addition to the order and number of recurring sections. ABBYY FlexiCapture 8.0 Professional lets you create templates for documents with **annexes**. Annexes are additional pages that may accompany a form. No template is applied to these pages, but they are taken into account during document assembly. For example, a loan application (fixed form) is submitted along with an income statement in a free-form style. This income statement can be processed as an annex.

Apart from describing the structure of a document in template properties, you must select an appropriate method of separating incoming images into documents. The desired method is selected when creating a project. FlexiCapture 8.0 provides the following separation methods:

- Create new document for

- Every image. This mode uses document templates to separate the page flow into documents.
- Every file. Each image file with all the pages it contains will be treated as a single document.
- Images separated by
 - blank pages (blank pages themselves will not be added to the batch)
 - pages with barcodes. You can specify a barcode type and/or value. If no barcode type or value is specified, any page with a barcode on it will be treated as a separator page. Otherwise, only pages with barcodes of the specified type or value will be used. Each separator page with a barcode will be added to the batch and will become the first page of a document.

Renumbering document pages

In **ABBYY FormReader 6.5**, you can use the API to analyze and re-order pages. The basic functionality also allows you to renumber pages manually.

ABBYY FlexiCapture 8.0 provides a user-friendly visual interface and a special page preview mode where you can manually correct document assembly using drag&drop.

Custom export

Very often the standard export procedure has to be modified or an entirely new export procedure has to be created.

ABBYY FormReader 6.5 included Automation export which called an external (custom-made) data export module. However, Automation export received data page by page and could not be used for exporting multi-page documents on its own. For multi-page documents, users had to resort to the API, which allowed them to call their own export settings GUI and to access the entire document or batch.

In **ABBYY FlexiCapture 8.0**, the basic functionality has been expanded and the program now supports:

- Export of multi-page documents as single items
- Export of images together with their data, including export of recognized documents to Searchable PDF
- Export of picture fields to files
- Automated removal of pages after export (Tools -> Options -> Document Processing -> Delete processed documents after export)

Additionally, Script Export can be used for customizing the export procedure. Export scripts can be written in the same scripting languages that are used for validation rules — Jscript and VBScript. Script Export can be used to solve various tasks, for example:

- Exporting some data from each page into one storage and exporting other data into another storage
- Creating customized export formats, e.g. adding support for custom XML schema
- Customized names of resulting files, e.g. you can use the values of certain fields in the names of resulting files
- Multiple export destinations, e.g. exporting data to several storages at a time or exporting some data to one storage and other data to another
- and more

Customizing the import procedure

Document separation. Processing unstructured documents of variable sizes, separated by blank pages, separator templates, or barcodes.

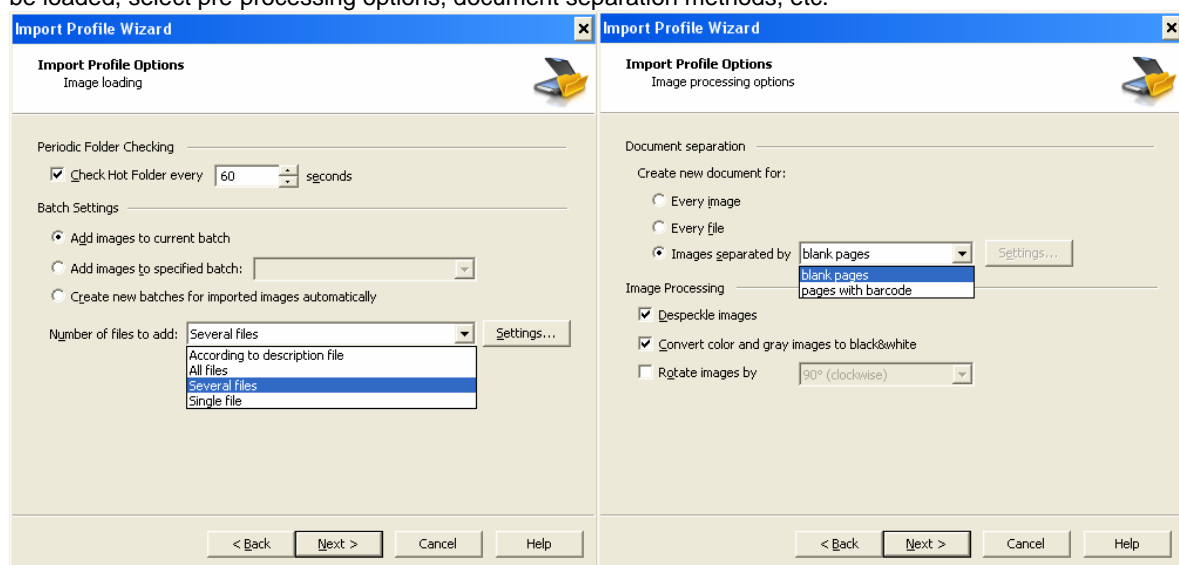
In **ABBYY FormReader 6.5 (Desktop Edition)**, a very simple model was used for multi-page documents: they were represented as a sequence of one-page templates.

ABBYY FlexiCapture 8.0 supports many typical document separation scenarios, both in the course of import and while matching documents with their templates.

Different types of Hot Folder. Sometimes you may need to load pages not just from one folder, but from several folders, or load them via FTP, or process e-mail attachments. You may also need to pre-process imported images: select the correct orientation, delete blank pages, binarize the images, etc.

The latest releases of ABBYY FormReader 6.5, support automated loading of images from a Hot Folder. You can also use the API to automate the loading of pages from any source.

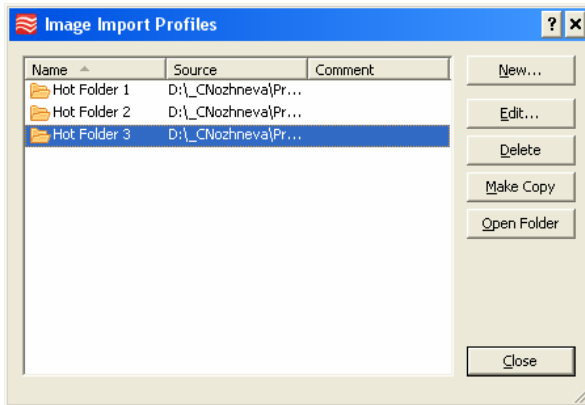
ABBYY FlexiCapture 8.0 supports so-called import profiles, which automate the loading of images from folders or from a scanner (Project -> Image Import Profiles... in Administrator mode). You can set up an import profile to access an image source at certain intervals or you can select manual loading, specify the batch into which the images should be loaded, select pre-processing options, document separation methods, etc.



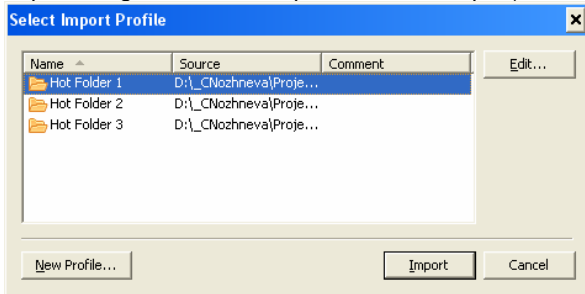
Useful tip

You can set up the program to load images from several sources in background mode. Documents will be loaded while the operator, for example, continues verifying previously loaded documents. To do this:

1. Set up the required number of import profiles with automated image loading (Check Hot Folder every...)



- The operator must consecutively select each of the profiles once (drop-down menu of Load Images button -> Import Images -> Select Import Profile -> Import).



- Once selected, A Hot Folder for which automatic image loading has been specified will be checked at regular intervals. This you can set up the program to check several Hot Folders at a time.

4. Integration with other applications

A typical integration task is to integrate forms processing into a bigger complex, where it becomes one of the stages of the entire process. In this case its is particularly important to make the import and export procedures work flawlessly with the external systems.

Other possible tasks include initializing from the outside such actions as the launch of the application, import, recognition, the launch of verification, and export. In this case the FormReader window is typically not visible and is only displayed during verification (which itself is not always required).

In ABBYY FormReader 6.5, integration is done through the API (handling events in FormReader, running the program in invisible mode, setting up export into an external system). Sometimes, a Hot Folder and the simplest Automation Export mat be sufficient for successful integration.

In ABBYY FlexiCapture 8.0, you can use the following to integrate the program with other systems:

- **The Command Line Interface** can be used to run the program with various parameters. FlexiCapture.exe command line parameters:
 - /Station:Admin - Launch in Administrator Station mode
 - /Station:Operator - Launch in Operator Station mode
 - /Project:"Path to project" - Open specified project on startup
 - /Batch:"Batch name" - Open specified batch on startup. If no batch with this names exists, it will be created. This parameter can be used, for example, to create a new batch every day (month) and to use the current date as the name of the batch.
 - /ProhibitChangePath - Works only in Operator Station mode. Restricts the operator to use only the batch and the project which were specified in the "/Project:" and "/Batch:" parameters. This option disables opening another project or batch.

The Command Line Interface can be used for creating a limited Operator UI with access only to one working batch or project. For example, you can specify the name of the project "C:\Work\My_FC8_Project\" and batch *Batch2* which will be opened when the program is run in Operator mode. At the same time you can prohibit switching to a different batch/project:

```
"C:\Program Files\ABBYY FlexiCapture 8.0 Professional\FlexiCapture.exe" /Station:Operator
/Project:"C:\Work\My_FC8_Project\" /Batch:"Batch2" /ProhibitChangePath
```

Scenario 1. Several operators are working over the LAN with the same project located in a shared folder. In cases like this, you can assign a specific batch to each operator, set up an import profile for each operator, and, for the sake of convenience, create a personal shortcut icon to FlexiCapture for each operator.

Scenario 2. If FlexiCapture is a module integrated into a bigger system, it can be launched from that system via the command line at specific moments, e.g. when the Hot Folder has accumulated images to be processed. For OperatorName, you can use the log-in of the current user working with the system. You will need to make the following settings in advance:

- Set up an import profile to automatically load images into the current batch
- Enable automatic launch of recognition for loaded images
- Enable automatic removal of exported images from the batch
- Set up export to the external system (e.g. using Script Export)

Then all the operator needs to do is to load images from the Hot Folder, verify recognized images, and export the verified images.

- **Import profiles** (Hot Folder/Scanner) for loading images from an external source, e.g. using a shared folder as an intermediary.

Scenario 3. A production scanner is scanning documents and places images to a shared folder creating one multi-page file for each document. Several operators are simultaneously working with the files in this folder, loading images with the help of an import profile where the method “one file – one document” was selected and which takes not all of the documents but only a portion, e.g. 30 files at a time.

- **Export Script** with the ability to call external components for export into an external system.

Reference: List of features available in AFR DE

AFR 6.5 Developer Edition offers the following features:

- Modification of the main window:
 - Adding a toolbar from which custom or standard system commands can be selected
 - Disabling some menu commands or hiding the entire menu bar
 - Changing the appearance of the Batch window (mode, list of columns) and the arrangement of all windows
- Programmatic control over the program
 - Launching the program from an external application
 - Reading and changing the processing settings (except for template and language settings)
 - Operations with batches and pages
 - Launching any processing commands
- Accessing the data contained in the current document: operations with fields and blocks
- Intercepting events
 - Opening/closing batches, pages, windows, and the program
 - Removing, adding, and changing the status of pages
 - Starting, ending, and changing the progress of major processing steps
 - Running any menu command
- Ability to export information about batch templates

Reference: List of features available in AFC 8.0 Professional

ABBYY FlexiCapture 8.0 Professional offers the following **built-in** features:

- Automated loading of images from a specified folder into a specified batch
- Automated launch of the recognition process
- Background processes (loading, recognition, and export)
- Data view form editor
- Operator Station (simplifying the interface and restricting functionality)
- Support for multi-page documents, multi-page rules
- Export of document images and picture fields, as well as export to Searchable PDF.
- and more

ABBYY FlexiCapture 8.0 Professional offers the following **customization** features:

- Script Rules
 - Allow to create complex custom algorithms for document validation and normalization using the standard scripting languages, **JScript**[®] and **VBScript**.
 - provide an error description
 - focus field
 - suggest replacement variants
 - additional information (for example, date of last processing)
- Script Export
- Command Line Interface