



## Frequently Asked Questions

### Using Zonal and Scan2OCR

#### Abstract

Image Access scanners always include a PC running under Linux. The Scan2Net® software includes a basic zonal OCR functionality and has now been enhanced with an OCR module based on the Tesseract OCR engine and the layout analysis software Leptonica. This enhancement, known as the Scan2OCR option can be purchased through our portal <https://portal.imageaccess.de>. The Tesseract engine is known to be one of the best OCR software engines and currently supports more than 100 languages, including many from Asia.

Many language packages can be downloaded for free from our portal and can all be installed on the target scanner. The engine works best however, if only one or a maximum of two languages are activated at the time of reading.

The OCR engine runs in the background at a low priority level and uses all remaining computing power. The multitasking software will not slow down the scanning process or any other processes. If OCR is enabled the user scans as always and depending on the size of the document and the speed of the operator, it will OCR page by page and may not keep up at certain stages. The software shows the progress on each page and at the end it may take a couple of extra seconds to complete if it did not keep up from the beginning.

This FAQ describes both the zonal OCR single scan functionality as well as the optional Scan2OCR for OCR processes in job mode. Since Scan2OCR is an option, the purchase and installation processes are also described here. The intended target audience is the operator and the administrator of the scanner at the customer site.

<b>Title</b>	Using Zonal and Scan2OCR
<b>Revision</b>	1.4
<b>Date:</b>	15.06.2016
<b>Category</b>	FAQ
<b>Owner</b>	Image Access GmbH, Germany
<b>Authors</b>	DI

## 1. Confidentiality

Status	Interested Party	Source	PDF
Public Information	Image Access Support	Yes	Yes
	Authorized Service Providers	No	Yes
	Image Access Customers	No	Yes

## 2. Revision History

Date	Rev.	Name	Description of Change	Reason of Change
01.06.2016	1.0	DI	Initial Version	
03.06.2016	1.1	DI, TI	Enhanced version	
03.06.2016	1.2	DI, TI	Edits for latest firmware	
09.06.2016	1.3	DI	Final version before publishing	
14.06.2016	1.4	DI	Rewrite to include zonal OCR	Zonal OCR missing
20.07.2016	1.5	DI	Changed option name from Background OCR to Scan2OCR	
28.07.2017	2.0	TI	Minor formatting changes	Formatting

## 3. Table of Contents

<b>1. Confidentiality .....</b>	<b>2</b>
<b>2. Revision History .....</b>	<b>2</b>
<b>3. Table of Contents .....</b>	<b>2</b>
3.1. Table of Figures .....	3
<b>4. Purpose .....</b>	<b>4</b>
<b>5. Scope .....</b>	<b>4</b>
<b>6. Terms and Definitions .....</b>	<b>4</b>
<b>7. Introduction .....</b>	<b>4</b>
7.1. The Difference between Zonal OCR and Scan2OCR .....	5
<b>8. Using Zonal OCR .....</b>	<b>5</b>
8.1. Add Additional OCR Languages .....	6
8.2. Upload Language Packages .....	7
8.3. Working with OCR Templates .....	8
8.4. Selecting the OCR Language Template for Zonal OCR .....	11
8.5. Masking and Starting the OCR Process .....	11
<b>9. Purchasing and Using Scan2OCR .....</b>	<b>13</b>
9.1. Purchasing an Option Key for Scan Wizard's Scan2OCR .....	13
9.1.1. Log on to the Customer Service Portal .....	13
9.1.2. Initiate Purchase of OCR Software Key .....	13
<b>10. Setting Up the Scanner for Scan2OCR .....</b>	<b>14</b>
10.1. Log On to the Scanner .....	14
10.2. Activate Software Key .....	14

10.3. Upload Language Packages .....	15
10.4. Working with OCR Templates .....	16
<b>11. ScanWizard Required Settings .....</b>	<b>19</b>
<b>12. OCR Process.....</b>	<b>20</b>
<b>13. License &amp; Disclaimer .....</b>	<b>22</b>

### 3.1. Table of Figures

Fig. 1: See installed OCR Languages .....	5
Fig. 2: Creating OCR language packages.....	6
Fig. 3: Select OCR settings .....	7
Fig. 4: Upload the language zip file .....	7
Fig. 5: Working with OCR Templates.....	8
Fig. 6: Select OCR in Templates Menu.....	9
Fig. 7: Template Setup.....	9
Fig. 8: List of configured OCR templates .....	10
Fig. 9: Selecting a Language for Zonal OCR .....	11
Fig. 10: Masking and starting the OCR process .....	11
Fig. 11: Include text in PDF file .....	12
Fig. 12: Purchase an Option Key for Scan2OCR.....	13
Fig. 13: Activating installed options.....	14
Fig. 14: Entering and applying the option key.....	14
Fig. 15: Select OCR settings .....	15
Fig. 16: Upload the language zip file .....	15
Fig. 17: Working with OCR templates.....	16
Fig. 18: Template setup .....	17
Fig. 19: List of configured OCR templates .....	18
Fig. 20: OCR transfer window showing selected language template.....	19
Fig. 21: Start the OCR process .....	20
Fig. 22: OCR pages with their status.....	21
Fig. 23: Selecting an output device for the PDF file .....	21

## 4. Purpose

The purpose of this document is to answer frequently asked questions about how to purchase, install and use the Scan2OCR option in ScanWizard.

## 5. Scope

The scope of the document includes all Scan2Net scanners like Bookeye or WideTEK scanners as well as OEM versions of these scanners. The firmware versions covered by this document are 6.50 and higher. The Windows versions covered in the documentation are Windows 7 and higher.

## 6. Terms and Definitions

Term	Description, Meaning
Scan2Net	Technology from Image Access implemented in many scanners. More at: <a href="http://www.imageaccess.de/?page=SoftwareScan2Net">www.imageaccess.de/?page=SoftwareScan2Net</a>
OCR	Optical Character Recognition More at: <a href="http://www.imageaccess.de/?page=Glossary&amp;lang=en">http://www.imageaccess.de/?page=Glossary&amp;lang=en</a>
Bookeye	Book scanners. More at: <a href="http://www.imageaccess.de/?page=ScannersBookscanner">www.imageaccess.de/?page=ScannersBookscanner</a>
WideTEK	Wide format scanner. More at: <a href="http://www.imageaccess.de/?page=ScannersWideformat">www.imageaccess.de/?page=ScannersWideformat</a>
Tesseract	<a href="https://en.wikipedia.org/wiki/Tesseract_(software)">https://en.wikipedia.org/wiki/Tesseract_(software)</a>

## 7. Introduction

The target audience for this FAQ document is the operator of a Scan2Net scanner who would like to use either the single scan zonal OCR function or the Scan2OCR function or both.

The document will provide a guideline to:

- Use the built in zonal OCR function for single scans.
- Perform masking on a document for zonal OCR.
- Purchase the Scan2OCR option on the Image Access Customer Service Portal.
- Download language packages.
- Activate the Scan2OCR software key on the scanner.
- Work with OCR templates.
- Configure ScanWizard settings for Scan2OCR

### 7.1. The Difference between Zonal OCR and Scan2OCR

There are two ways to perform OCR on a scanned document on an Image Access scanner, zonal OCR and Scan2OCR. Both methods are functionality within the Scan2Net® technological platform running in Image Access scanners, also known as the firmware.

Zonal OCR can only be operated in single scan mode. When operators scan to a PDF file, they can highlight a section of the scanned text for OCR by using a frame to cover the area they would like to OCR. This is known as “masking”. This OCR function is useful for example, when scanning an engineering drawing and using zonal OCR to capture the metadata in an information block (title, date, version etc.).

Scan2OCR is an optional software function which allows the user to scan in job mode and to OCR the entire page or multiple pages scanned, using multipage PDF mode. As the name implies, the OCR process runs in the background so that you can continue scanning subsequent pages in the job. This function is useful when scanning complete source material such as books, legal files and other multipage documents for which the operator wants all of the text OCR'd. Scan2OCR requires the purchase of a software option key.

## 8. Using Zonal OCR

By default, the languages German and English are installed on the scanner. If you would like to check your scanner to see if any other languages are installed, log in to the scanner as Poweruser and then go to Base Settings and click the OCR button. A screen will show you which languages are currently installed.

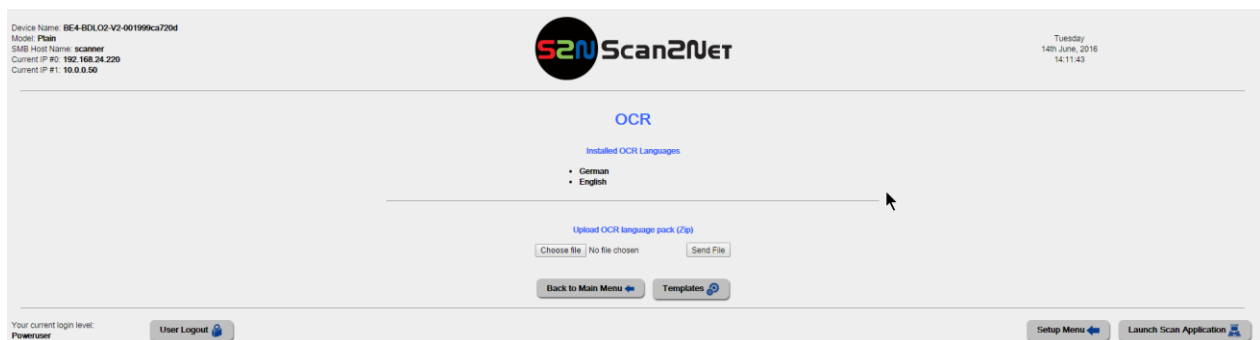


Fig. 1: See installed OCR Languages

### 8.1. Add Additional OCR Languages

**Note:** This process is the same for both Zonal OCR and Scan2OCR.

If you need to upload a different language for OCR, you first need to download the language from the Image Access Customer Service Portal. create OCR Language Packages

In the **Actions** menu of the Customer Service Portal, click the link for **OCR Package** or follow this link:

[https://portal.imageaccess.de/index.php?page=Actions\\_getOCRPackage&lang=en](https://portal.imageaccess.de/index.php?page=Actions_getOCRPackage&lang=en)

Select one or more of the languages shown in the list for use with OCR. To select multiple languages, hold the CTRL key when pressing on a language in the list. German and English are preinstalled on the scanner by default when the OCR key is activated.

**Best Practice:** We recommend that you download the English and German language files from time to time, in addition to your selected languages. They are updated periodically.

**CREATE OCR PACKAGES**

On this page, you can create and download OCR Language Packages which contain languages other than the standard English (EN) and German (DE) languages already included with the option. You can select multiple languages by holding down the control (ctrl) button to select more than one language from the list (Windows) or holding down the command button to select more than one language from the list (Mac).

**PLEASE SELECT PACKAGES:**

- Lithuanian
- Macedonian
- Malay
- Malayalam
- Maltese
- Marathi
- Math / Equation
- Middle English (1100-1500)
- Middle French (1400-1600)
- Nepali
- Norwegian
- Old Georgian
- Oriya
- Pashto
- Persian
- Polish**
- Portuguese
- Punjabi
- Romanian

**SELECTED OCR PACKAGES**

- English
- French
- German
- German (Fraktur)
- Polish

Create OCR packages

**Fig. 2: Creating OCR language packages**

Confirm your selection by clicking the button **Create OCR packages**. Shortly after clicking the button, a ZIP file containing the selected language packages will be made available to you for download.

## 8.2. Upload Language Packages

In the menu area **Base Settings**, press the **OCR** button.

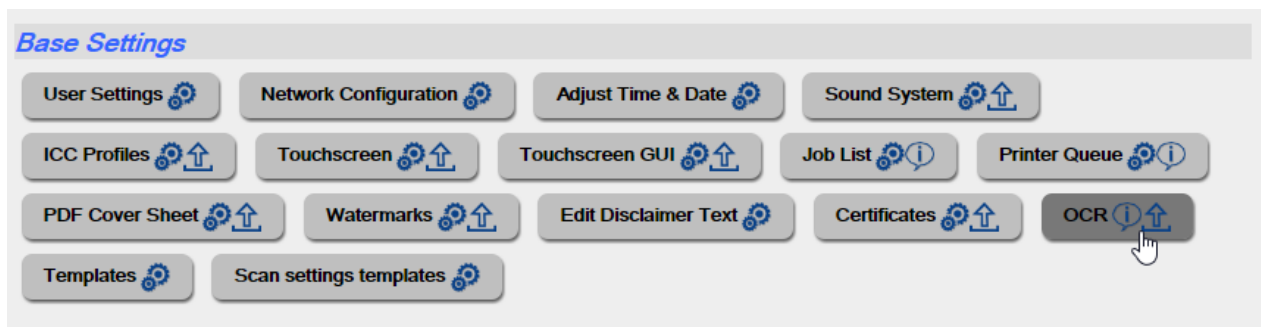


Fig. 3: Select OCR settings

Two languages – German and English – are preinstalled on the scanner by default. If you downloaded other languages in the Customer Service Portal, you can upload the ZIP file here.



Fig. 4: Upload the language zip file

After uploading your ZIP file, all languages available on the scanner for OCR will be listed here. Press the **Templates** button to add, edit, duplicate or delete OCR templates.

**Note:** At least one OCR template must be enabled to use the OCR functionality. See the next chapter “Working with OCR Templates”

**Back to Main Menu** returns you to the Main Menu.

### 8.3. Working with OCR Templates

In order to work effectively with both zonal and Scan2OCR, you will need to create and use OCR templates. OCR templates optimize processing time and deliver results with fewer errors. If you work with more than one language, you may need different templates for each language used.

To select a previously configured OCR template, press the **Templates** button either in the OCR screen or in the menu area **Base Settings**.

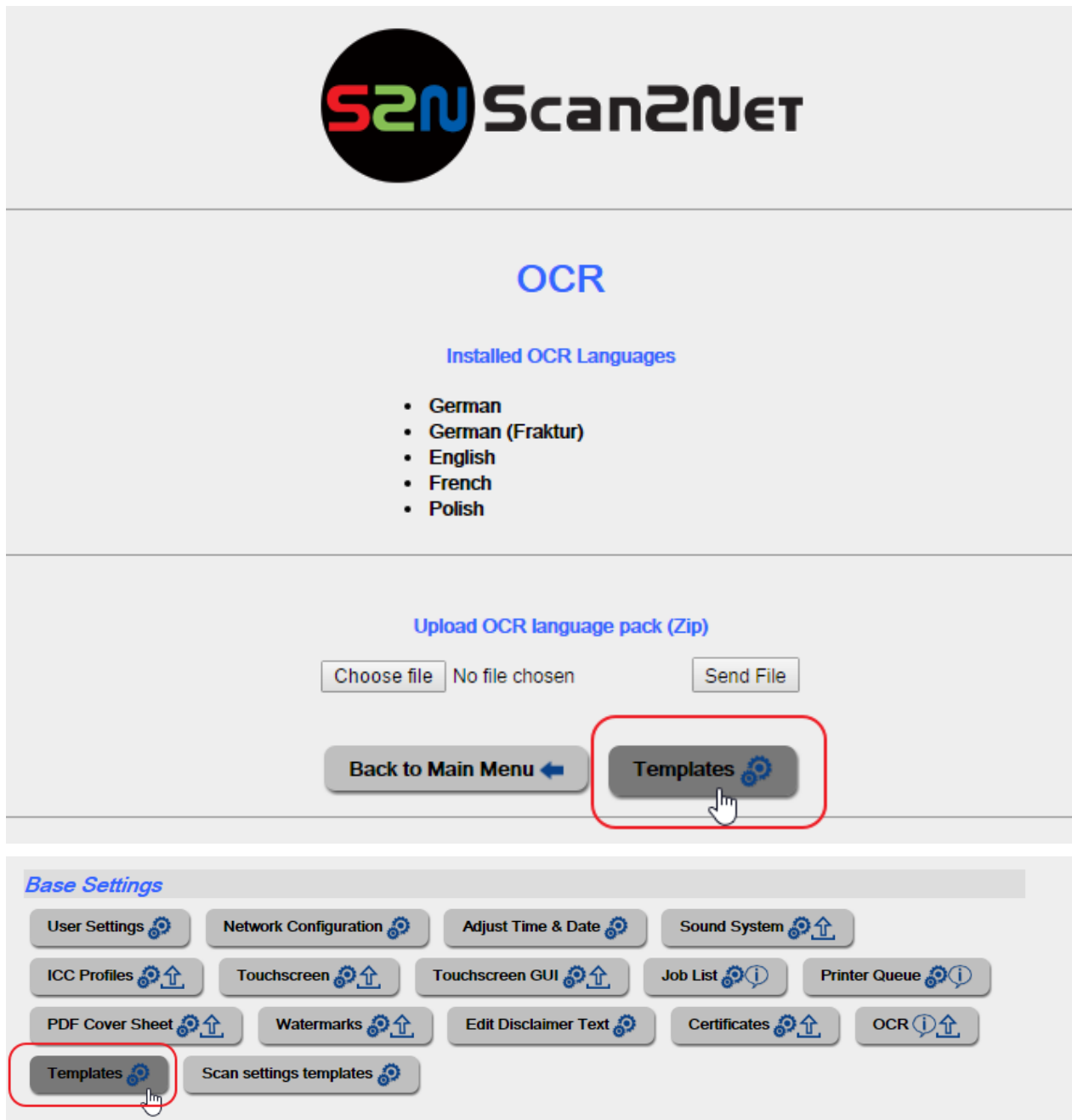


Fig. 5: Working with OCR Templates



In the **Templates** menu, select the item **OCR**.

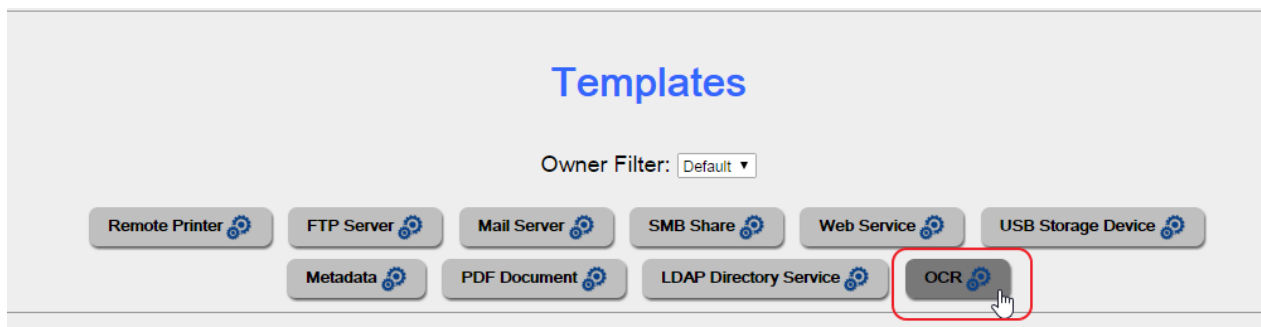


Fig. 6: Select OCR in Templates Menu

Open the available template **OCR-EN-DE** by clicking on **Setup**. Use this default template to configure your own individual template. After modifying and configuring the default template for your own requirements, each template can be assigned a maximum of two recognition languages. If you only need one language in your template, select the same language for both OCR language fields.

**Best Practice:** Ensure that the templates with which you will actively work all have the **Visibility** status set to **Public** (as shown below).

The screenshot shows the 'Setup' page for the 'OCR (OCR-EN-DE)' template. The title 'Setup' is in large blue font. Below it is a form with several fields: 'OCR Engine' with a dropdown set to 'Internal'; 'OCR Language' with a dropdown set to 'English'; 'Additional OCR Language' with a dropdown set to 'German'; 'Name' with a text field containing 'OCR-EN-DE' and a blue arrow icon; 'Default Template' with an unchecked checkbox; 'Visibility' with a dropdown set to 'Public', which is highlighted by a red arrow; and 'Owner' with a dropdown set to 'Default'. At the bottom of the form is a 'Back' button.

Fig. 7: Template Setup

To create additional individualized templates, press the **Duplicate** button one of the configured templates and follow the steps described above for setup. Save each template with a different name. Depending on your requirements, numerous templates can be created which can be selected for use with the application ScanWizard.

**Best Practice:** For clarity, we recommend that you save the newly created templates to a file name which includes the language, such as **OCR-FR** for the French template.

The screenshot displays the 'Templates' management interface. At the top, there's a title 'Templates' and an 'Owner Filter' set to 'Default'. Below this, a row of buttons represents different template types: Remote Printer, FTP Server, Mail Server, SMB Share, Web Service, USB Storage Device, Metadata, PDF Document, LDAP Directory Service, and OCR. The 'OCR' section is active, showing a table of configured templates. The table has columns for Name, Owner, Visibility, Setup, Duplicate, Delete, and Export. Three templates are listed: OCR-DEFraktur, OCR-EN-DE, and OCR-FR. The 'Duplicate' button for the first template is highlighted with a red box. The 'Name' column for the first three rows is also highlighted with a red box. At the bottom, there are buttons for 'Choose file', 'No file chosen', 'Import', 'Back to Main Menu', and 'Export'.

Name	Owner	Visibility	Setup	Duplicate	Delete	Export
OCR-DEFraktur	Default	Public	Setup	Duplicate	Delete	
OCR-EN-DE	Default	Public	Setup	Duplicate	Delete	
OCR-FR	Default	Public	Setup	Duplicate	Delete	

Fig. 8: List of configured OCR templates

## 8.4. Selecting the OCR Language Template for Zonal OCR

Once you have created a template for the desired language and set the visibility value to Public, you will see the menu item “OCR” under the Transfer menu in the ScanWizard application, as shown in the figure below.



Fig. 9: Selecting a Language for Zonal OCR

## 8.5. Masking and Starting the OCR Process

Now you can leave the Transfer menu, perform a single scan and then select the Masking button from the toolbar. Set the red masking frame to cover the area you would like to OCR and select the item **OCR Text** to start the zonal OCR process.

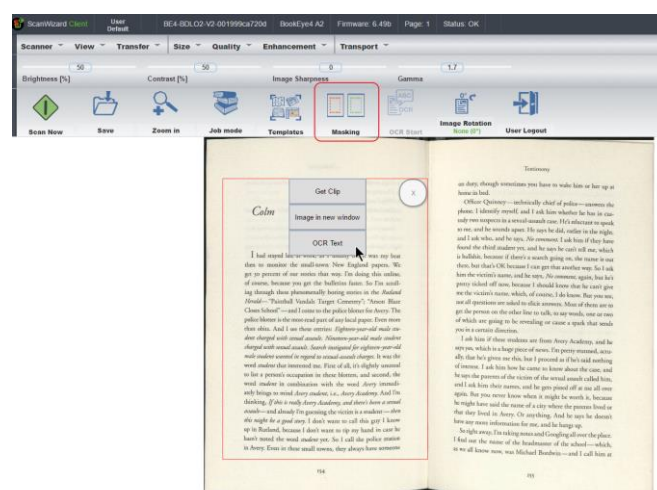


Fig. 10: Masking and starting the OCR process

When the OCR process has completed, the captured text appears in an editor window. Make any necessary corrections to the text and press the **Include Text** button to include the OCRred text in the PDF file.

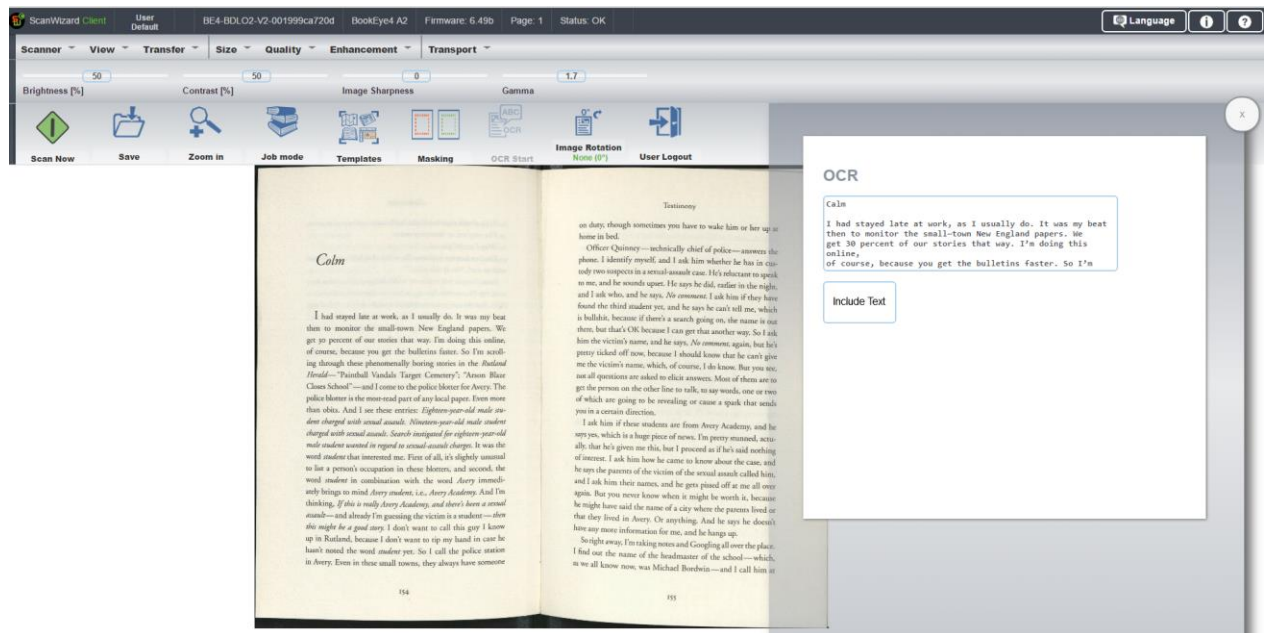


Fig. 11: Include text in PDF file

Your zonal OCR process is complete. The generated PDF file contains the OCR'd text.

## 9. Purchasing and Using Scan2OCR

### 9.1. Purchasing an Option Key for Scan Wizard's Scan2OCR

#### 9.1.1. Log on to the Customer Service Portal

In order to use Scan2OCR in ScanWizard, you must first log on to the Image Access Customer Service Portal at <https://portal.imageaccess.de> and purchase a key to activate the software on your scanner.

Select the menu item **Actions** .

#### 9.1.2. Initiate Purchase of OCR Software Key

Select the item **Software Options**.

Enter the serial number of the scanner.

Select **Scan2OCR** from the list of available options by checking the checkbox .

<input checked="" type="checkbox"/>	21	Journal/Billing	Journal/Billing	0,00EUR
		5Qw9Kg6.w4A54JiPh2UvZo		
<input checked="" type="checkbox"/>	23	Multiple User Management	Multiple User Management	0,00EUR
		I6QATahkhRMNXfkPioLgNg		
<input type="checkbox"/>	25	OCR	Background OCR	750,00EUR

**Fig. 12: Purchase an Option Key for Scan2OCR**

After confirming the purchase, the key and an invoice for the purchase will be sent to the email registered on the Customer Service Portal.

## 10. Setting Up the Scanner for Scan2OCR

### 10.1. Log On to the Scanner

Log on to the scanner as the **Poweruser**.

### 10.2. Activate Software Key

In the menu area **Updates & Uploads** click the button **Installed Options**.



Fig. 13: Activating installed options

Enter the key for the OCR software which was sent to you via email. You may copy and paste it out of the PDF but be aware that no extra blanks are copied. Confirm it by clicking the **Apply** button. The text in the line will change from red to green.

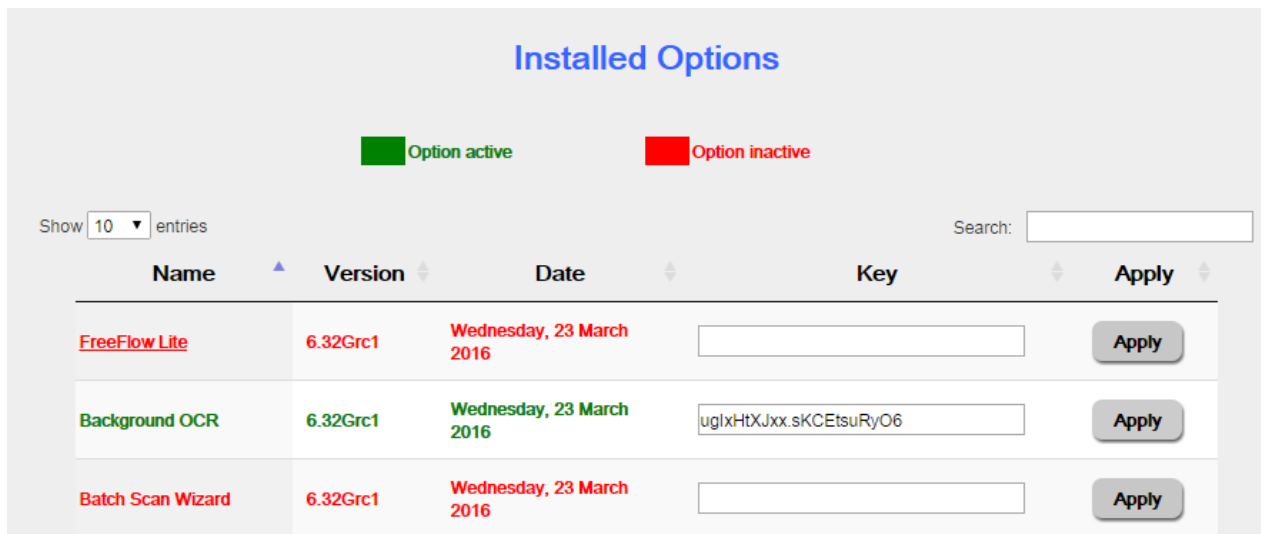


Fig. 14: Entering and applying the option key

**Back to Main Menu** returns you to the Main Menu.

### 10.3. Upload Language Packages

In the menu area **Base Settings**, press the **OCR** button.

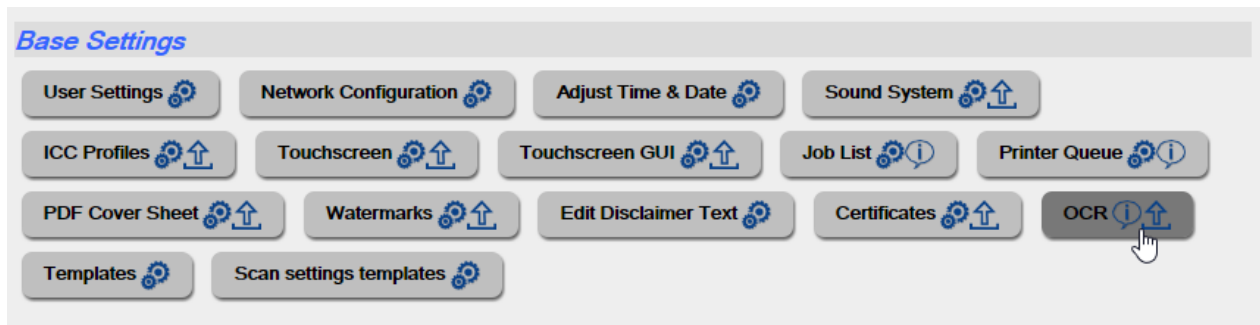


Fig. 15: Select OCR settings

Two languages – German and English – are preinstalled on the scanner by default. If you downloaded other languages in the Customer Service Portal, you can upload the ZIP file here.



Fig. 16: Upload the language zip file

After uploading your ZIP file, all languages available on the scanner for OCR will be listed here. Press the **Templates** button to add, edit, duplicate or delete OCR templates.

**Note:** At least one OCR template must be enabled to use the OCR functionality. See the next chapter “Working with OCR Templates”

**Back to Main Menu** returns you to the Main Menu.

**Templates** directs you to the OCR templates section described in the following chapter.

#### 10.4. Working with OCR Templates

In order to work effectively with Scan2OCR, you will need to create and use OCR templates. OCR templates optimize processing time and deliver results with fewer errors. If you work with more than one language, you may need different templates for each language used.

To select a previously configured OCR template, press the **Templates** button either in the OCR screen or in the menu area **Base Settings**.

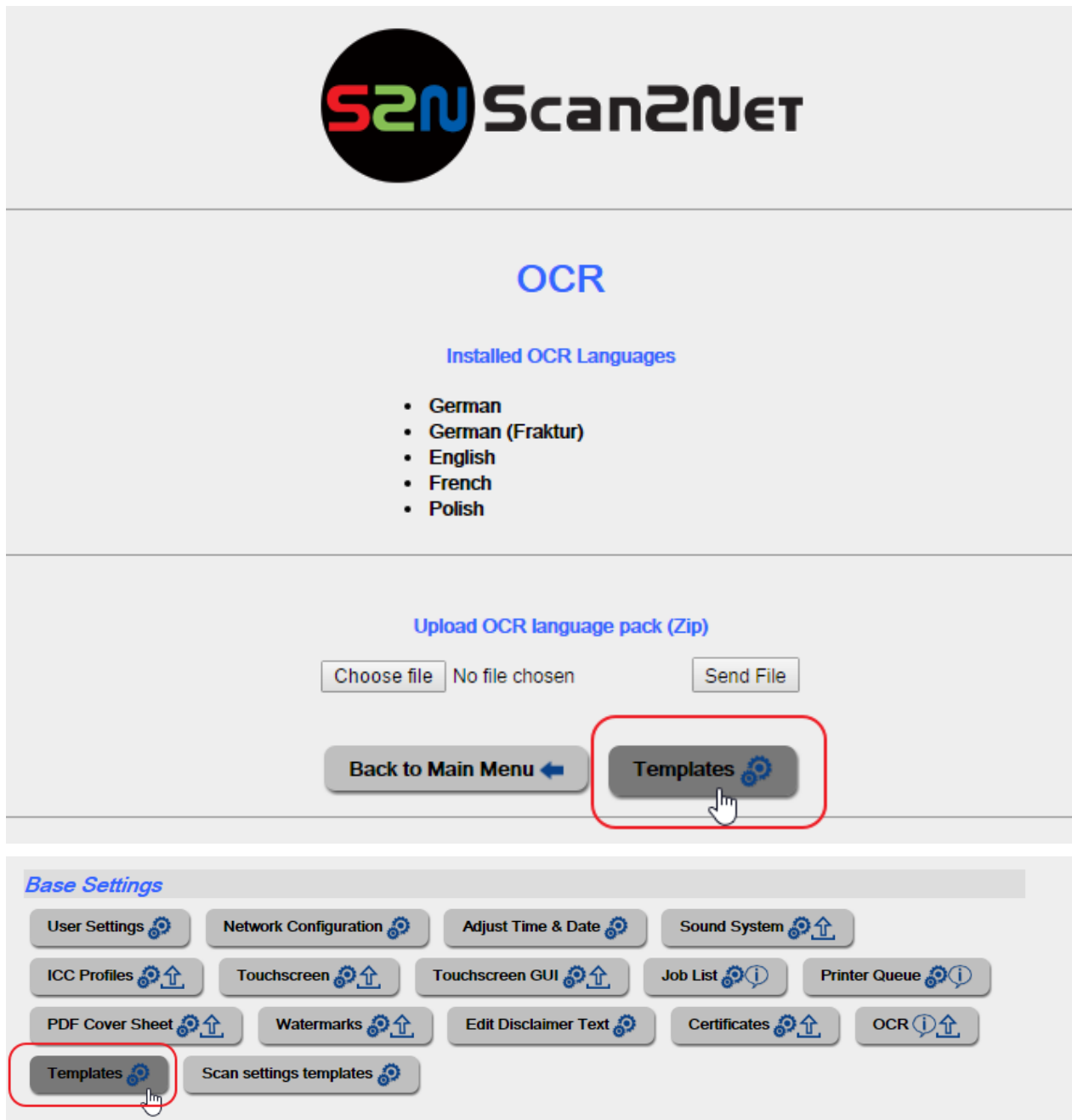
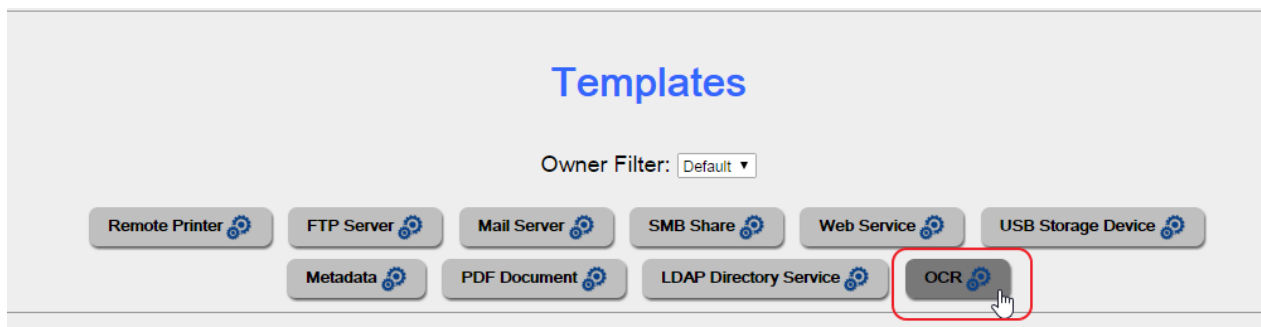


Fig. 17: Working with OCR templates



In the **Templates** menu, select the item **OCR**.



Open the available template **OCR-EN-DE** by clicking on **Setup**. Use this default template to configure your own individual template. After modifying and configuring the default template for your own requirements, each template can be assigned a maximum of two recognition languages. If you only need one language in your template, select the same language for both OCR language fields.

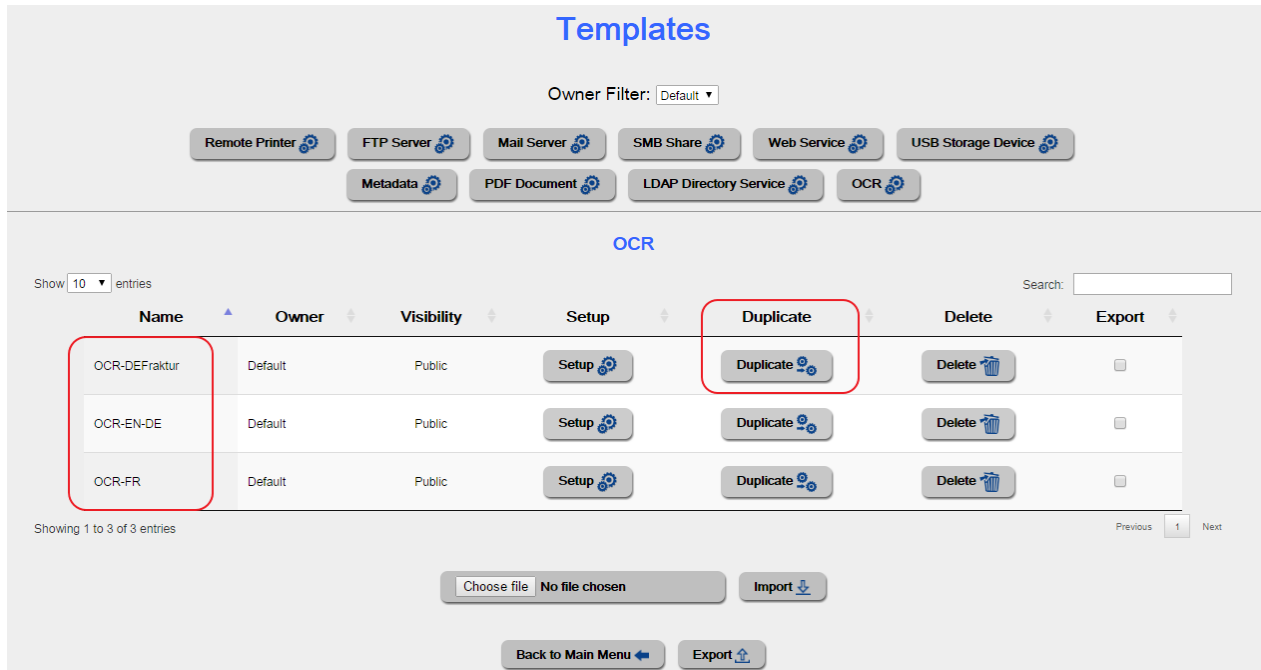
**Best Practice:** Ensure that the templates with which you will actively work all have the **Visibility** status set to **Public** (as shown below).

A screenshot of the 'Setup' form for the 'OCR (OCR-EN-DE)' template. The title 'Setup' is at the top in blue. The form has a light blue header bar with the text 'OCR (OCR-EN-DE)'. Below the header, there are several input fields: 'OCR Engine' with a dropdown set to 'Internal'; 'OCR Language' with a dropdown set to 'English'; 'Additional OCR Language' with a dropdown set to 'German'; 'Name' with a text field containing 'OCR-EN-DE' and a blue icon; 'Default Template' with an unchecked checkbox; 'Visibility' with a dropdown set to 'Public' (highlighted by a red arrow); and 'Owner' with a dropdown set to 'Default'. At the bottom of the form is a 'Back' button.

Fig. 18: Template setup

To create additional individualized templates, press the **Duplicate** button one of the configured templates and follow the steps described above for setup. Save each template with a different name. Depending on your requirements, numerous templates can be created which can be selected for use with the application ScanWizard.

**Best Practice:** For clarity, we recommend that you save the newly created templates to a file name which includes the language, such as **OCR-FR** for the French template.



The screenshot displays the 'Templates' management interface. At the top, there's a section for 'Owner Filter' set to 'Default' and a row of category buttons: Remote Printer, FTP Server, Mail Server, SMB Share, Web Service, USB Storage Device, Metadata, PDF Document, LDAP Directory Service, and OCR. Below this, the 'OCR' section is active, showing a table of configured templates. The table has columns for Name, Owner, Visibility, Setup, Duplicate, Delete, and Export. Three templates are listed: OCR-DEFraktur, OCR-EN-DE, and OCR-FR. The 'Duplicate' button for the first template is highlighted with a red box, and the 'Name' column for the first three rows is also highlighted with a red box. At the bottom, there are buttons for 'Choose file', 'No file chosen', 'Import', 'Back to Main Menu', and 'Export'.

Name	Owner	Visibility	Setup	Duplicate	Delete	Export
OCR-DEFraktur	Default	Public	Setup	Duplicate	Delete	
OCR-EN-DE	Default	Public	Setup	Duplicate	Delete	
OCR-FR	Default	Public	Setup	Duplicate	Delete	

**Fig. 19: List of configured OCR templates**

## 11. ScanWizard Required Settings

To start the OCR service, the following steps are required:

- ScanWizard must be in Job Mode,
- PDF must be selected as output format

**Best Practice:** To reduce the possibility of error and speed up the processing, resolution should be set to a minimum of 300 DPI.

The transfer window shows which template is selected. The template defines the language used for the recognition process.

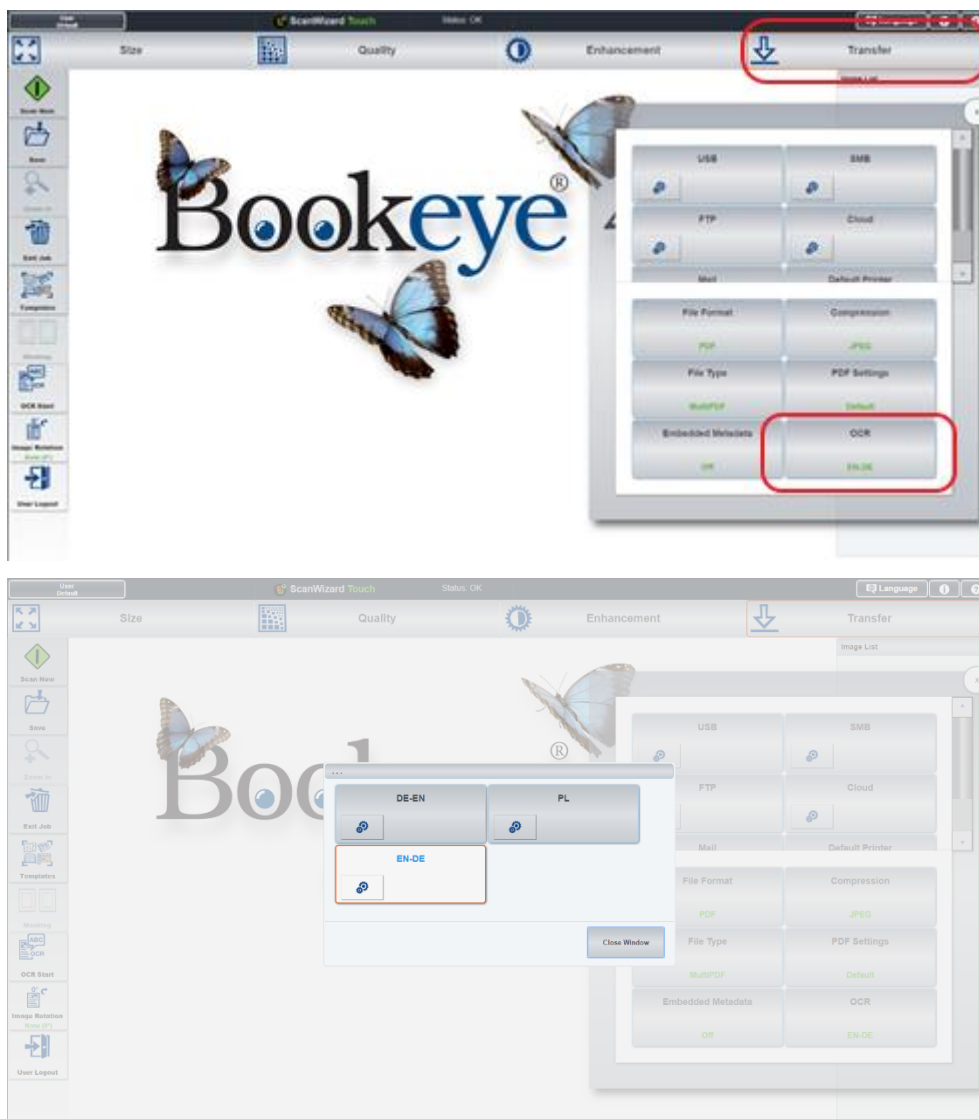


Fig. 20: OCR transfer window showing selected language template

## 12. OCR Process

When all required settings have been made, begin scanning the document on which the OCR process should be run. It is possible to start the OCR process at any time before or during scanning, even after scanning all pages in the job.

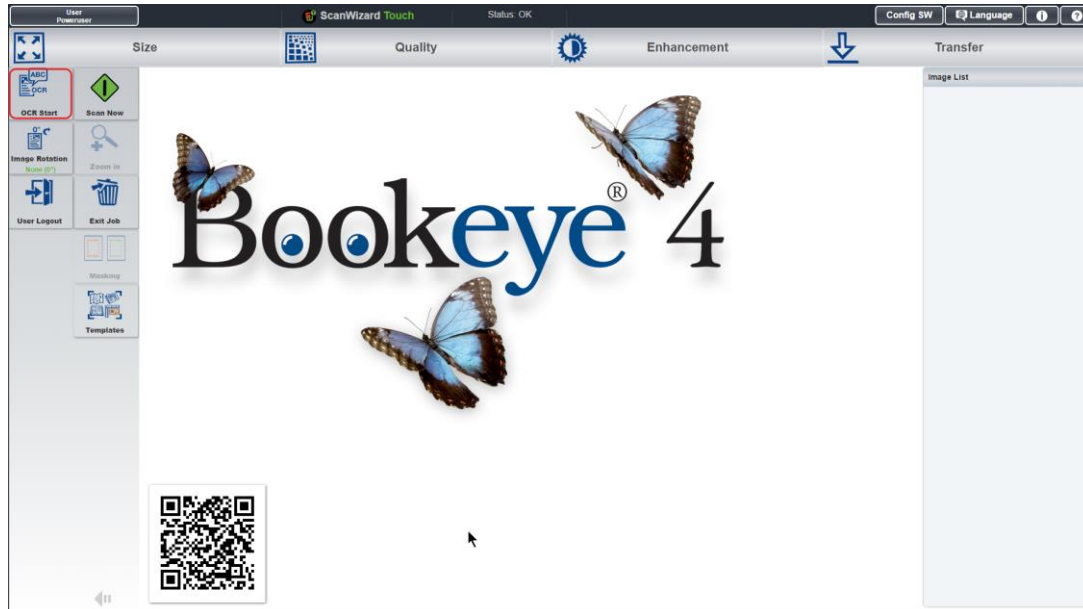


Fig. 21: Start the OCR process

After each scan, a thumbnail image will appear in the frame at the right of the screen. Each thumbnail image contains a status icon. There are four different status reports:

	The document is in the OCR queue
	OCR processing in progress
	OCR processing completed
	OCR processing cancelled or aborted

If OCR processing is interrupted before the job is complete, it may be that not all pages were successfully processed. In this case, the processing can be restarted as long as the job is still active.

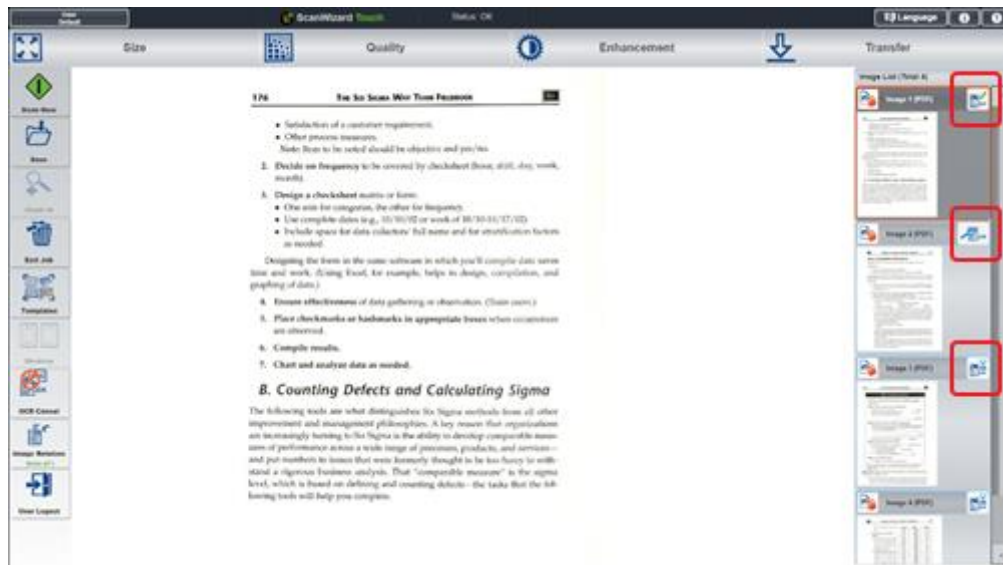


Fig. 22: OCR pages with their status

When all thumbnail images show the status icon for **Processing complete**, it is certain that all pages have been properly processed. At this point, the job should be saved as a PDF file to the desired output device.

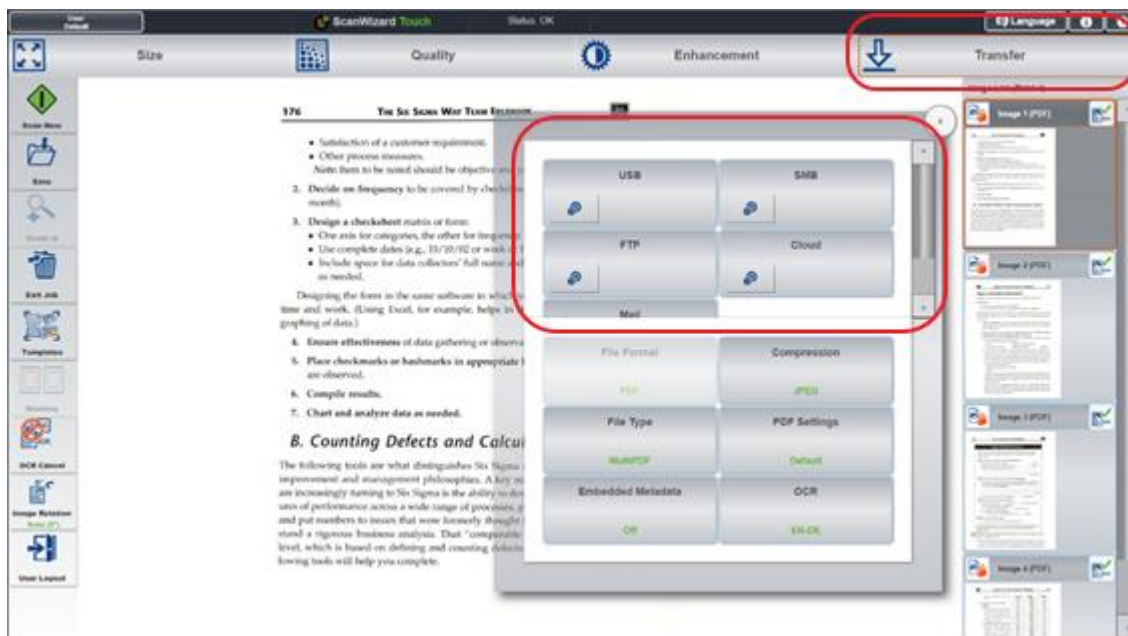


Fig. 23: Selecting an output device for the PDF file

### 13. License & Disclaimer

The OCR engine inside the Scan2Net® scanners (Bookeye® and WideTEK®) uses the Tesseract OCR engine which is used under the Apache license:

Copyright 2016 Image Access GmbH, Europe

Licensed under the Apache License, Version 2.0 (the "License");  
you may not use this file except in compliance with the License.  
You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.