
Aquaforest Searchlight Quick-Start Guide



Version 1.31
September 2019

Contents

| | | |
|-------|--|----|
| 1 | Creating a Library | 3 |
| 1.1 | Library Settings..... | 3 |
| 1.2 | Document Settings..... | 4 |
| 1.3 | Document Archive Settings..... | 5 |
| 1.4 | OCR Settings..... | 5 |
| 1.4.1 | Extended OCR Engine Settings..... | 5 |
| 1.4.2 | Aquaforest OCR Engine Settings | 6 |
| 1.5 | Scheduler..... | 6 |
| 1.6 | Alert Settings..... | 7 |
| 1.7 | Finish | 9 |
| 2 | Updating a Library | 10 |
| 3 | Import settings from an existing Library | 11 |
| 4 | Audit and Conversion Status..... | 12 |

1 Creating a Library

Creating Document Libraries in Aquaforest Searchlight is managed by a wizard. This wizard can be launched by clicking the **Add new library** button on the Dashboard

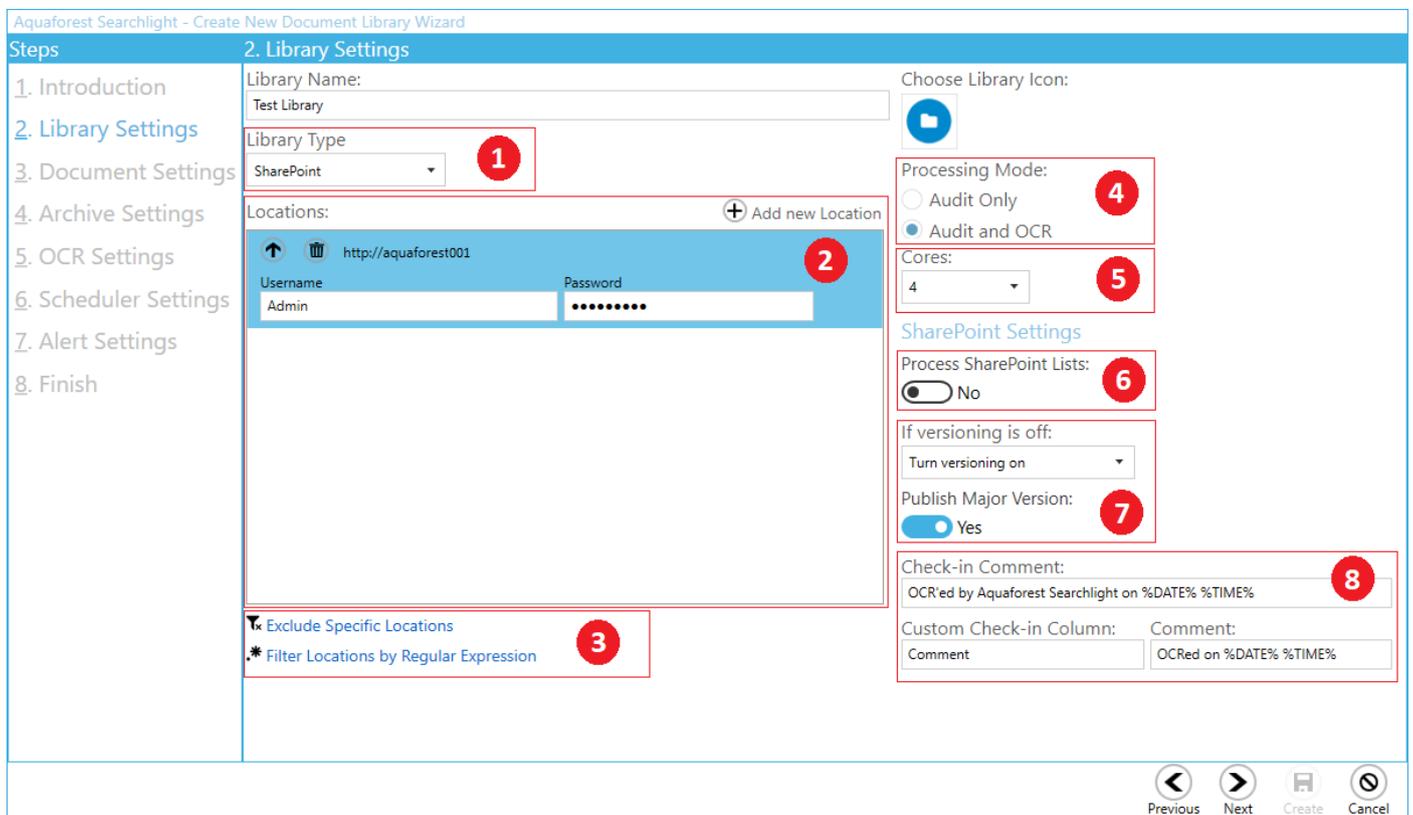
Dashboard Library Settings



| NAME | LIBRARY TYPE | LAST |
|------|--------------|------|
|------|--------------|------|

The wizard provides helpful information throughout the different stages of the document library creation process which aids in better understanding the various steps and settings involved. Refer to section 6.3 of the Reference Guide for detailed descriptions of each of the settings in each page.

1.1 Library Settings

The screenshot shows the "2. Library Settings" step of the wizard. On the left is a "Steps" sidebar with 8 steps: 1. Introduction, 2. Library Settings (selected), 3. Document Settings, 4. Archive Settings, 5. OCR Settings, 6. Scheduler Settings, 7. Alert Settings, and 8. Finish. The main area contains several settings: "Library Name" (Test Library), "Library Type" (SharePoint, marked with a red circle 1), "Locations" (http://aquaforest001, marked with a red circle 2), "Exclude Specific Locations" (Filter Locations by Regular Expression, marked with a red circle 3), "Choose Library Icon" (video icon), "Processing Mode" (Audit and OCR, marked with a red circle 4), "Cores" (4, marked with a red circle 5), "SharePoint Settings" (Process SharePoint Lists: No, marked with a red circle 6; If versioning is off: Turn versioning on, marked with a red circle 7; Publish Major Version: Yes, marked with a red circle 7), "Check-in Comment" (OCR'ed by Aquaforest Searchlight on %DATE% %TIME%, marked with a red circle 8), and "Custom Check-in Column" (Comment: OCR'ed on %DATE% %TIME%). At the bottom right are "Previous", "Next", "Create", and "Cancel" buttons.

1. Do the document to be processed reside on a file system, SharePoint or Office 365?
2. Enter the Location(s). It can consist of the following:
 - o One or more SharePoint site collections, SharePoint sites, SharePoint document libraries and/or SharePoint lists.
 - OR
 - o One or more File System paths
3. There are 2 ways to filter locations:
 - a. Excluding specific location – locations that match specified URL(s) are excluded
 - b. By regular expressions – locations (URLs) that match the specified regular expressions are included

This is useful if you are processing a whole site collection and want to exclude specific locations and/or include only specific sites or libraries.

4. Do you only want to **Audit Only**, or **Audit and OCR**? Audit means that Searchlight will analyse the searchability of the documents and report how many searchable, partially searchable and image-only documents are found in the location(s) specified, while Audit and OCR will find the non-searchable documents, and then make them searchable.
5. The number of cores to use to process documents in parallel. For instance, if 8 cores is specified, Searchlight will process 8 documents simultaneously, which will significantly reduce the total processing time.
6. Choose whether to process SharePoint Lists or not. If this is turned on, Searchlight will process the attachments in each list item. Note, however, that processing SharePoint lists can be extremely time consuming if they are very large.
7. Turn versioning on if you want to have a 'backup' of the original documents, otherwise the documents will be overwritten with new searchable ones.
8. You can choose to add a check-in comment to the OCR'd files once they are uploaded to SharePoint. Optionally, you can also add a custom comment to a custom SharePoint column. However, the custom SharePoint column must be either of 'Text' or 'Date' type.

1.2 Document Settings

This page enables the user to specify rules and criteria for the selection of documents to be processed.

The screenshot shows the '3. Document Settings' step of the wizard. The left sidebar lists steps 1 through 8. The main content area is divided into several sections:

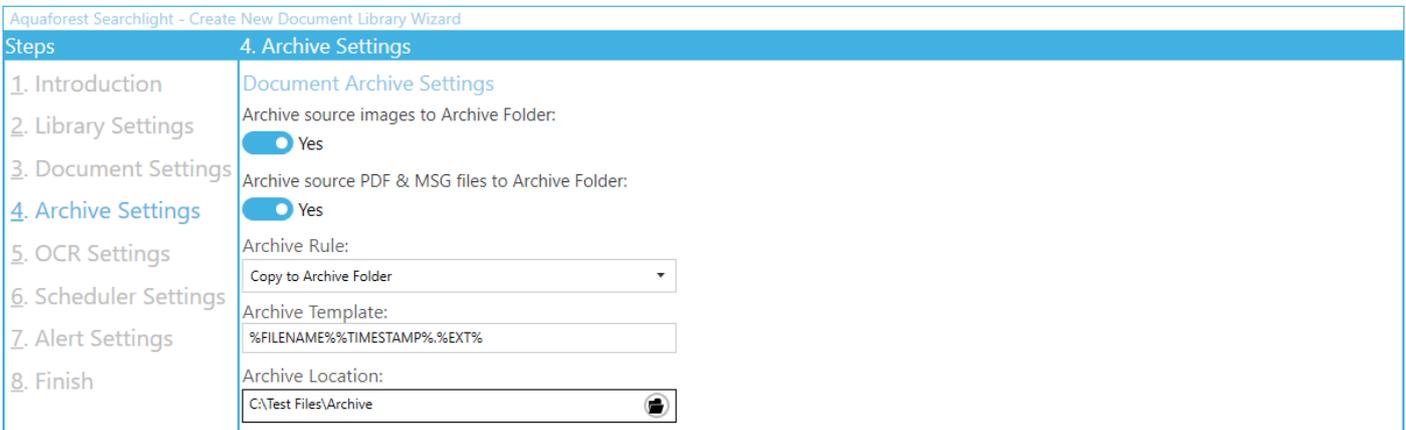
- PDF Selection:** Process PDF Documents (Yes), Image Only PDFs (Yes), Partially Searchable (Yes), Fully Searchable (No), Hidden Text (No).
- BMP Selection:** Process BMP Files (No), Delete Original BMP (No).
- JPEG Selection:** Process JPEG Files (No), Delete Original JPEG (No).
- PNG Selection:** Process PNG Files (No), Delete Original PNG (No).
- MSG Selection:** Process PDF Attachments (No).
- TIFF Selection:** Process TIFF Files (No), Delete Original TIFF (No).
- Temp Folder Location:** C:\Aquaforest\Searchlight\temp
- Filter Settings:** Date Filter: No Filter, From: 09/09/2019, To: 09/09/2019. Options: Exclude Specific Documents, Filter Documents by Regular Expression.
- Document Error Settings:** Document Error Rule: Take no Action, Document Error Location: (empty).
- Advanced Settings:** Retry On Error (No), OCR Document Limit (0), Retain Creation Date (No), Retain Modified Date (No), Retain Created By (No), Retain Modified By (No).

1. Select the document types to process. For image files, you can delete the original images from the source location after they have been converted to searchable PDFs.
2. The **Temp Folder Location** is where Searchlight temporarily stores downloaded files as well as files created during OCR.
3. There are different options to filter documents:
 - a. By modified or creation date – documents that fall within the specified range are excluded
 - b. By document paths – documents that match the specified paths are excluded

- c. By regular expressions - documents whose properties match the specified regular expressions are included
- 4. There is also the option of retaining the original metadata on the document and in SharePoint so that even after uploading the searchable PDF these columns will not be changed.

1.3 Document Archive Settings

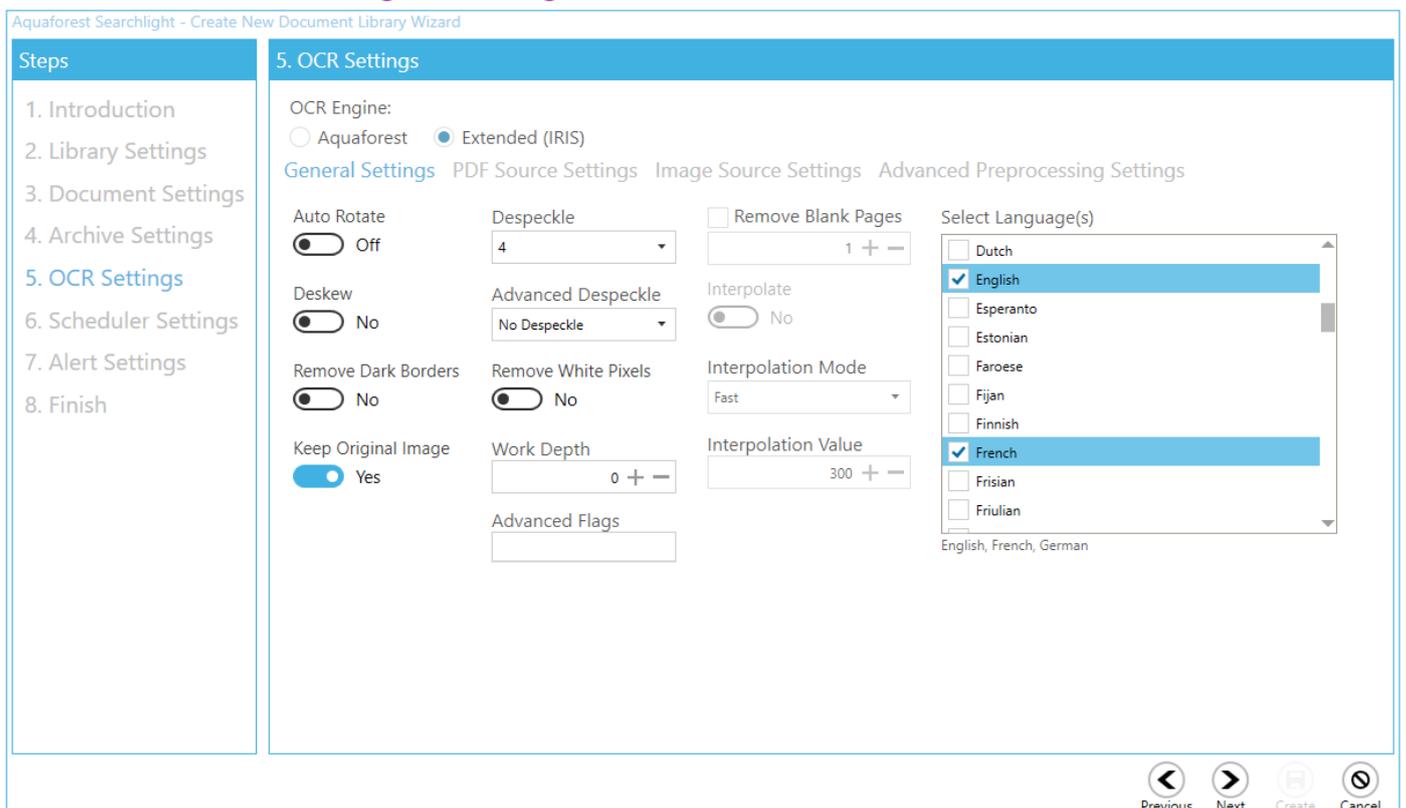
This page provides the option of archiving source files so as to have a backup before OCR is applied to them.



1.4 OCR Settings

In this section, you can set the OCR settings. Aquaforest Searchlight comes bundled with two OCR Engines: Aquaforest OCR engine and the Extended IRIS (Canon) OCR engine. The Extended OCR is the default engine and supports more languages (100+) than the Aquaforest OCR engine and also has the ability to process documents that have pages in different languages. See section 3.2 of the Reference Guide for more information about the OCR engines.

1.4.1 Extended OCR Engine Settings



1.4.2 Aquaforest OCR Engine Settings

Aquaforest Searchlight - Create New Document Library Wizard

Steps 5. OCR Settings

1. Introduction
2. Library Settings
3. Document Settings
4. Archive Settings
5. OCR Settings
6. Scheduler Settings
7. Alert Settings
8. Finish

OCR Engine:
 Aquaforest Extended (IRIS)

[General Settings](#) [PDF Source Settings](#) [Image Source Settings](#)

Auto Rotate: Off
Despeckle: 3

Deskew: On
OCR Language: English

Remove Lines: Off
Box Graphics: Treat all Graphics Areas as Text

Stamps:

Advance Flags:

Previous Next Create Cancel

1.5 Scheduler

The scheduler allows Aquaforest Searchlight to automate the running of document libraries. You can either run it manually or run periodically, every day at a particular time or every hour etc.

Aquaforest Searchlight - Create New Document Library Wizard

Steps 6. Scheduler Settings

1. Introduction
2. Library Settings
3. Document Settings
4. Archive Settings
5. OCR Settings
6. Scheduler Settings
7. Alert Settings
8. Finish

Manual
 Once per day
At: 11:15

Continuous
Every: 1 Hour(s)
Between: 00:01 And: 23:59

Run once
On: 10/09/2019
At: 11:15

Previous Next Create Cancel

1.6 Alert Settings

The alert settings provide you with the option of periodically sending email alerts as well as generating reports of job runs within a particular date range. Creating alerts is managed by another wizard within the library creation wizard.

1. Select the action(s) you want to perform

Aquaforest Searchlight - Create New Document Library Wizard

Steps 7. Alert Settings

| Configuration | Action |
|---------------|---|
| Action | What action(s) do you want the alert task to perform? |
| Email | Send an email <input checked="" type="checkbox"/> Yes |
| Report | Generate a CSV report <input checked="" type="checkbox"/> Yes |
| Trigger | Attach the CSV report to the email <input checked="" type="checkbox"/> Yes |
| Finish | Save Report <input type="checkbox"/> No |
| | Location: <input type="text"/> |

Navigation: Previous Next

Bottom Bar: Previous Next Create Cancel

2. Select the email settings

Aquaforest Searchlight - Create New Document Library Wizard

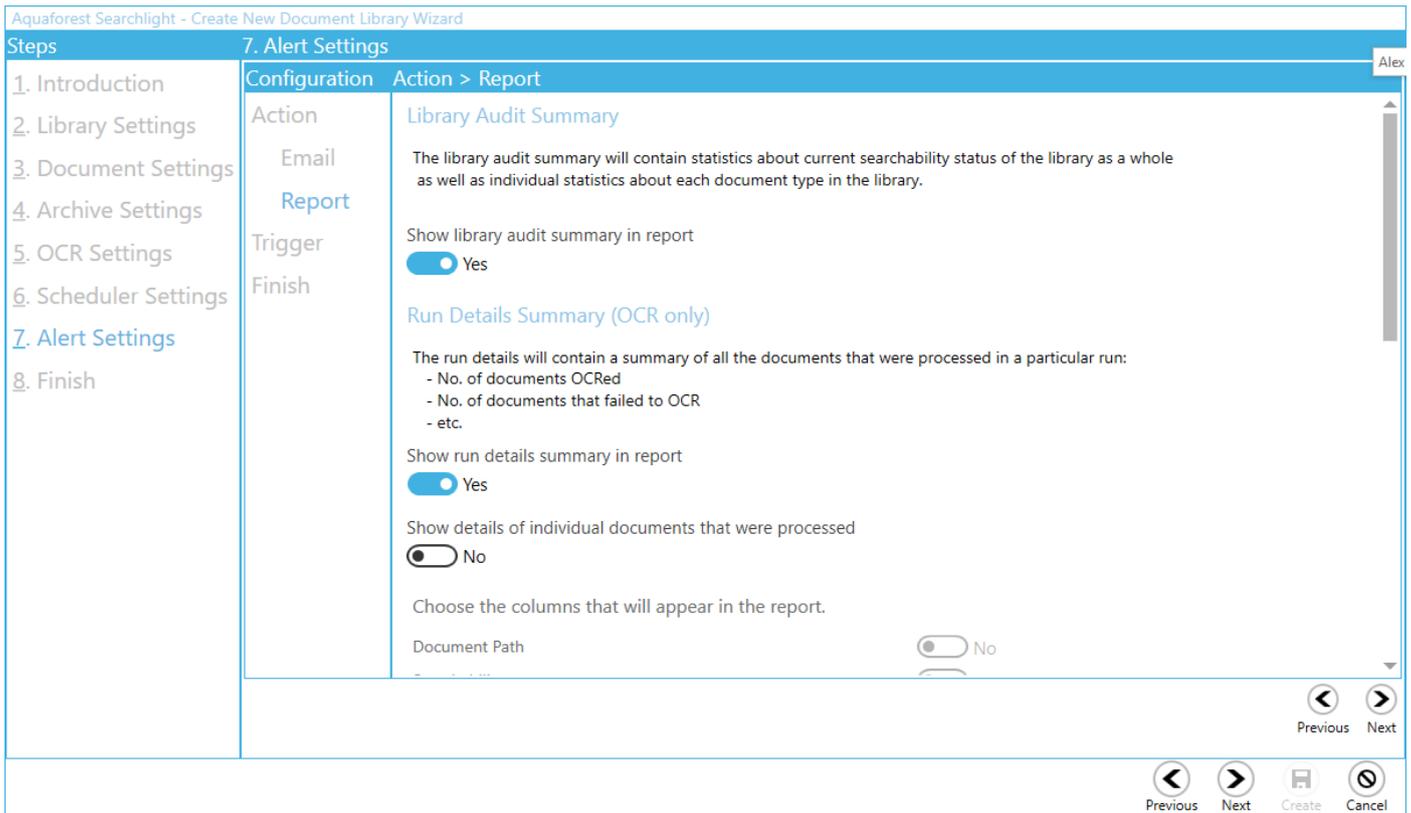
Steps 7. Alert Settings

| Configuration | Action > Email |
|---------------|---|
| Action | Email Settings |
| Email | From Email Address: <input type="text" value="support@aquaforest.com"/> |
| Report | To Email Address: <input type="text" value="support@aquaforest.com"/> |
| Trigger | Email Subject: <input type="text" value="%LIBRARYNAME% %STATUS%"/> |
| Finish | Email Message: <input checked="" type="checkbox"/> Test Email Processing of library: '%LIBRARYNAME%' completed with status: '%STATUS%' . Log file: %LOGFILEPATH% %ERRORMESSAGE% |

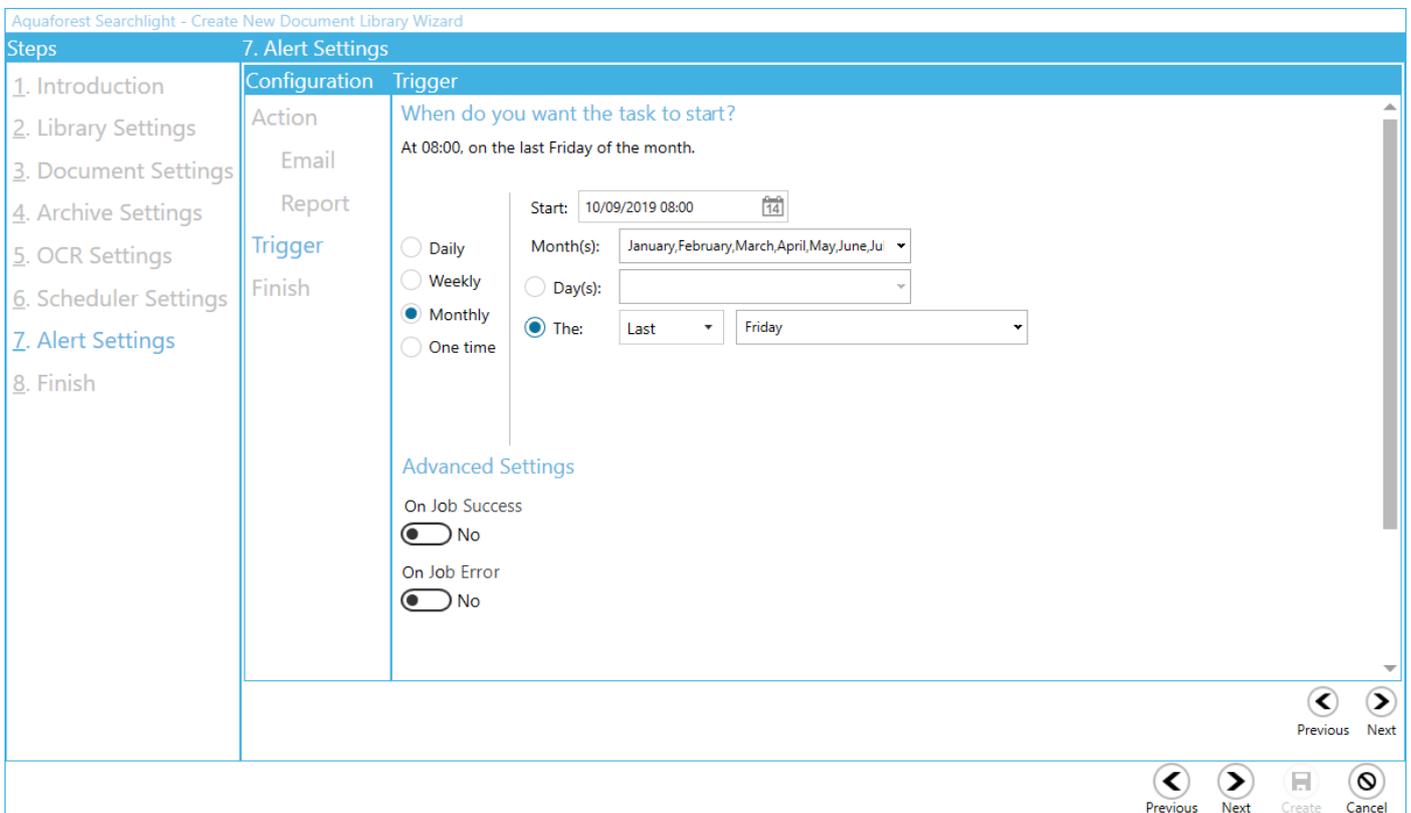
Navigation: Previous Next

Bottom Bar: Previous Next Create Cancel

3. Select the report settings. You can choose to get a summary of the library status as a whole and/or details about specific runs.

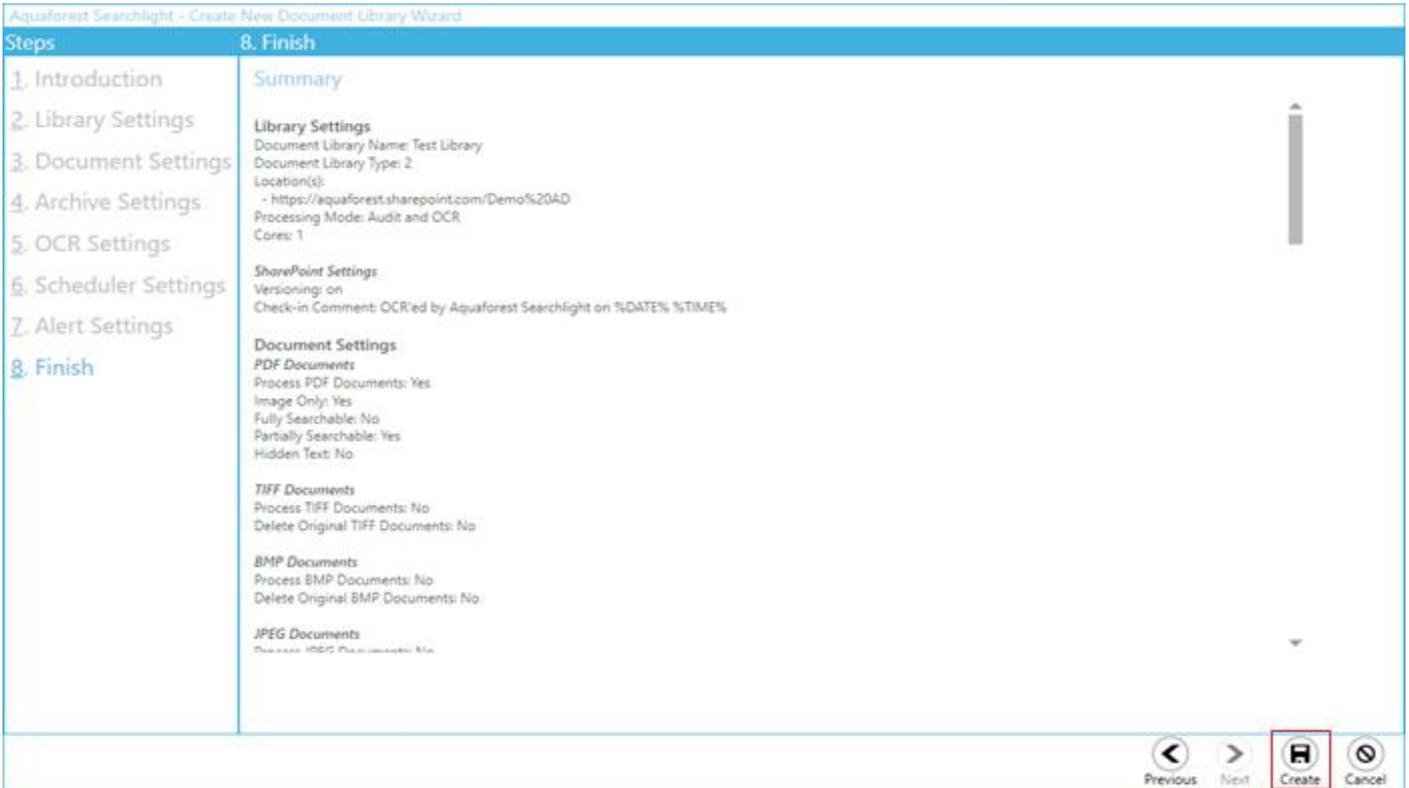


4. Select when you want the task to run. Based on the current settings, you will get an email with the report attached every last Friday of the month at 8 am.

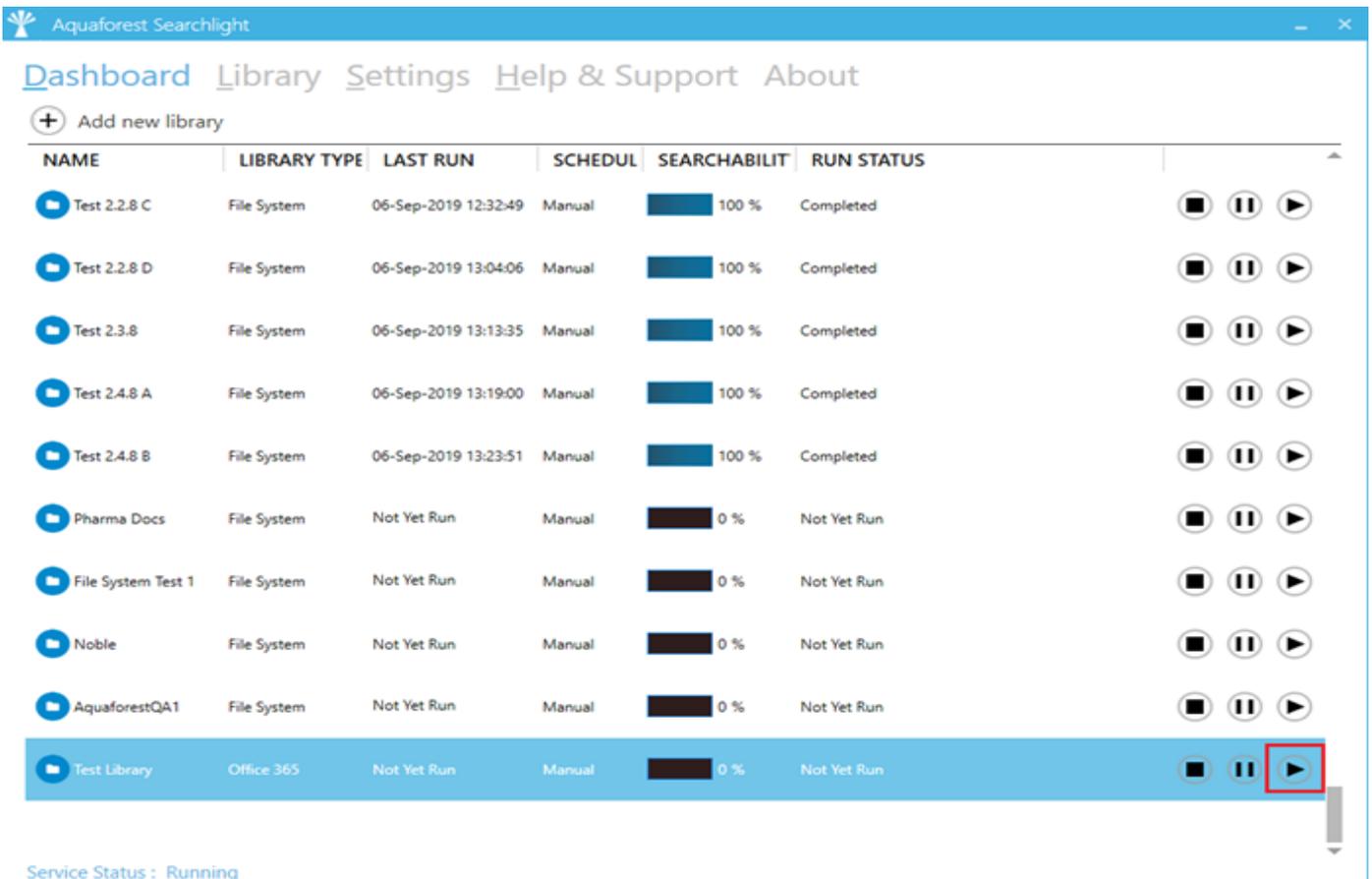


1.7 Finish

In the Finish page, you will get a summary of all the settings you selected for this library. You can review them to see if you missed anything. If not, click on the **Create** button at the bottom of the wizard to create the library.



The new library will be added to the dashboard. As the library is set to run manually, click on the **Run** button to start processing.



2 Updating a Library

1. All the settings of a library can be edited by double-clicking the library from the dashboard.
2. You can also select a library to edit by choosing the library from the combo box at the top of the page.
3. To delete the library, click on the **Delete** button at the bottom of the **Library Settings** page.

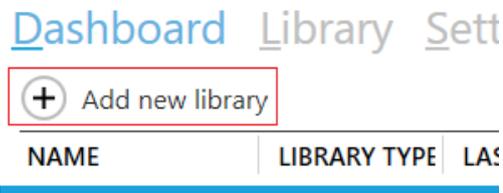
You can also, delete the library by right-clicking on the library from the dashboard and clicking on **Delete Document Library**

3 Import settings from an existing Library

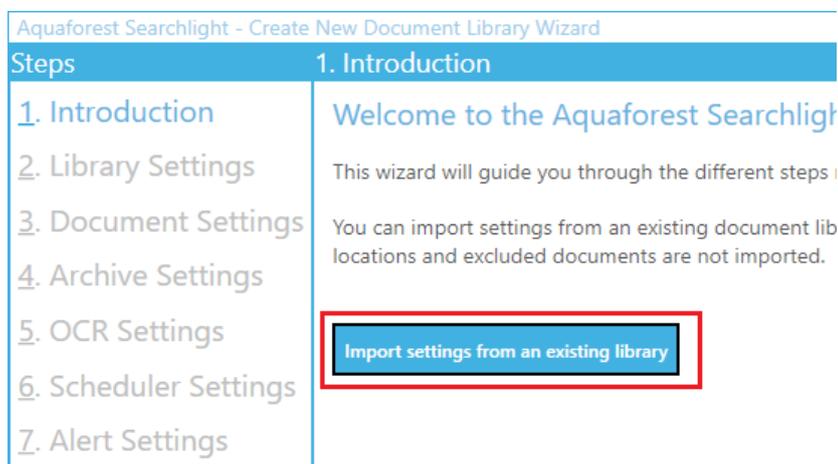
Searchlight also provides the ability to import settings from an existing library. However, “locations”, “excluded locations” and “excluded documents” are not imported because it is not allowed to have the same locations in multiple libraries.

To import settings:

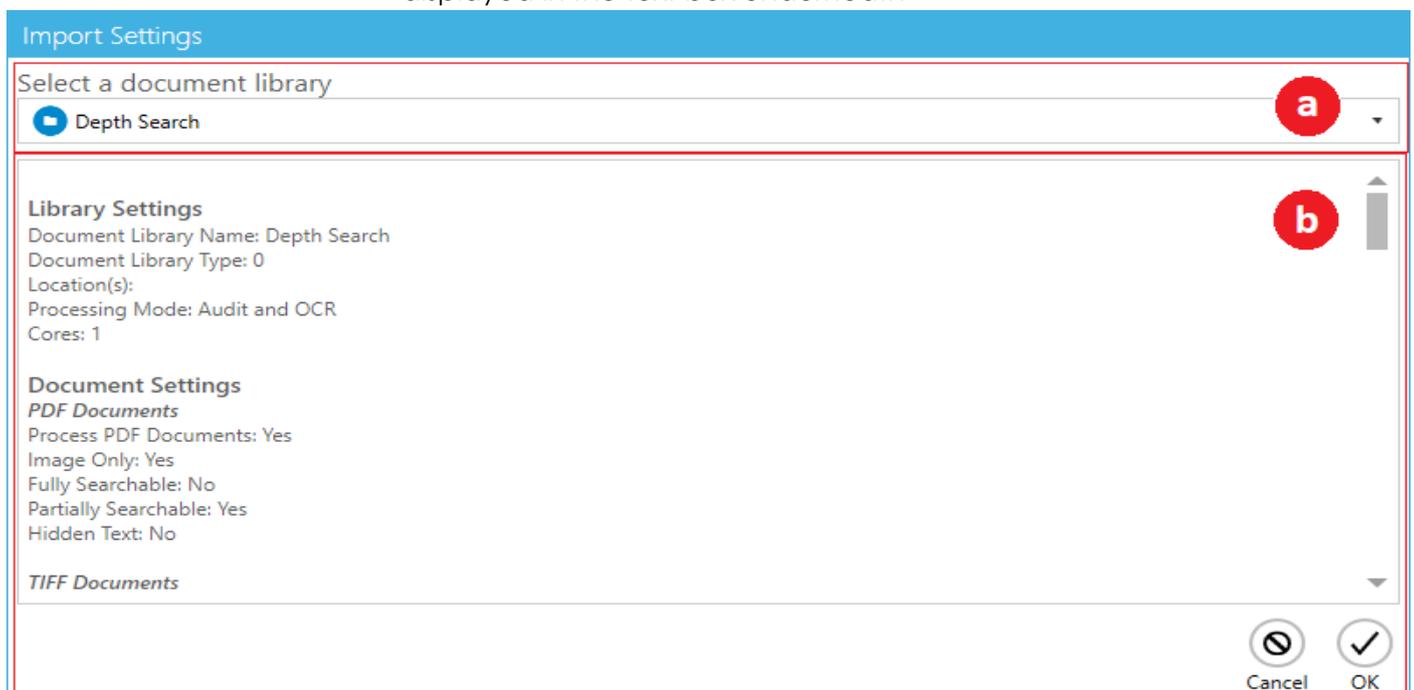
1. Click on the **Add new library** button on the Dashboard to open the wizard.



2. From the wizard select **Import settings from an existing library**



3. From the Import Settings window:
 - a. Select the document library from which the settings are to be imported
 - b. A summary of the settings of the selected document library will be displayed in the text box underneath



4. After clicking **OK** from the Import Settings window, go to **Library Settings** and add the location(s) to process. Optionally, add specific locations and documents to exclude.
5. Review all the settings in the other sections and click **Create**.

4 Audit and Conversion Status

After running a library, its current state will be summarised in the **Statistics** section of the **Status** tab as shown below:

The screenshot shows the Aquaforest Searchlight application window. The top navigation bar includes 'Dashboard', 'Library', 'Settings', 'Help & Support', and 'About'. A dropdown menu on the right shows 'Test 2.1.2 E'. Below the navigation bar, there are tabs for 'Status', 'Library Settings', 'Document Settings', 'Archive Settings', 'OCR Settings', 'Run Details', 'Scheduler', and 'Alerts'. The 'Status' tab is active, displaying two main sections: 'STATISTICS' and 'LOG OUTPUT'.

STATISTICS

- PDF Documents**
 - Total PDF Documents: 3
 - Image-only PDFs: 0 (0 %)
 - Partially Searchable PDFs: 0 (0 %)
 - Fully Searchable PDFs: 3 (100 %)
 - Error PDF Documents: 0
 - Total PDF Pages: 18
 - Image-only Pages: 0 (0 %)
 - Fully Searchable Pages: 18 (100 %)
- Image (TIFF,BMP,JPG,PNG) Documents**
- MSG Documents**
- Library Totals**
 - Total Documents: 3
 - Total Error Documents: 0
 - Total Pages: 18
 - Total Searchable Pages: 18 (100 %)

There is a 'Generate Audit Report' button at the bottom of the statistics section.

LOG OUTPUT

```

Complete with exit code: 0
Exit Code : 0

Checking document: C:\Tests 2.1.2\Test 2.1.2 E\Source\TestFile.pdf
Searchability Status : searchable,  Number of Pages 6.
Contains hidden text

02-Sep-2019 12:56:46: Finalising conversion

Re-calculating statistics after OCR...
Document library statistics after OCR:
-----
PDF Documents
Total PDF Documents: 3
Image-only PDFs: 0 (0 %)
Partially Searchable PDFs: 0 (0 %)
Fully Searchable PDFs: 3 (100 %)
Error PDF Documents: 0
Total PDF Pages: 18
Image-only Pages: 0 (0 %)
Fully Searchable Pages: 18 (100 %)

Library Totals
Total Documents: 3
Total Error Documents: 0
Total Pages: 18
Total Searchable Pages: 18 (100 %)

02-Sep-2019 12:56:46: Conversion ended
  
```

At the bottom of the window, the 'Service Status' is shown as 'Running'.

It provides a breakdown of all the documents processed grouped by the document format. For more detailed analysis of a library, go to the **Run Details** tab.

The screenshot shows the Aquaforest Searchlight application interface. At the top, there are navigation tabs: Dashboard, Library, Settings, Help & Support, and About. Below these are sub-tabs: Status, Library Settings, Document Settings, Archive Settings, OCR Settings, Run Details, Scheduler, and Alerts. The main content area is divided into two sections: Run History and Run Details.

Run History Section:

- 1. A control for "Showing last 5 runs" with plus and minus buttons.
- Table with columns: #, RUN ID, RUN DATE, PROCESSING MODE, Status, AUDIT RESULTS (Successful Documents, Error Documents), Status, CONVERSION RESULTS (Successful Documents, Error Documents).

Run Details Section:

- 2. Radio buttons for "Audit" (selected) and "Conversion".
- 3. A dropdown menu for "SEARCHABILITY" with options: Searchable (checked), Image Only (checked), Partially Searchable (unchecked), Error (checked), and Hidden Text (checked).
- 4. A "Limit" control set to 500 with plus and minus buttons.
- 5. Navigation arrows and a "1" button for page navigation.
- 6. Action buttons: Export to CSV, Generate Log, View Full Log, and Reload.

At the bottom left, it says "Service Status : Running".

1. Select the number of previous runs to show. You need to click on the **Reload** button after updating this value. Clicking on a run history will display its details in the **Run Details** section below.
2. Select whether you want to display the documents that were audited or OCRed for that particular run.
3. All columns with the ▼ icon next to them can be filtered. You can filter the Searchability status to only display documents that errored during Audit or OCR (Conversion).
4. You can limit the number of documents to display per page. You need to click on the **Reload** button after updating this value.
5. Display the next/previous 500 documents (since **Limit** is set to 500).
6. You can:
 - a. Export the current run details to a CSV file
 - b. Generate a log file of the current selected run history which will show a file by file assessment of all documents processed. The log file can be generated in a PDF, RTF or HTML format.
 - c. View the log file of the selected run (as displayed in the **Library > Status** tab).