Aquaforest

an ORPALIS company

PDF Days Online 2021

Smart Processing of Millions of PDF Files

+++ NO CODE REQUIRED +++

Agenda

- A quick overview of Aquaforest's **Low Code / No Code Applications**

- **Use Case ALPHA – All that paper!**
  Record-keeping for European shipments after Brexit → **Kingfisher**
- **Use Case BRAVO – Can't email this contract!**
  Compressing PDFs for thousands of Bank employees → **Autobahn DX**

Aquaforest
an ORPALIS company

Low Code No Code Applications

Autobahn DX

Searchlight OCR

Searchlight TAGGER

Aquaforest Kingfisher

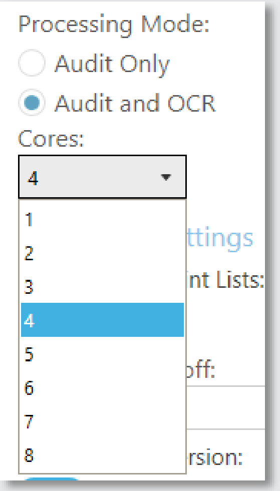
Aquaforest PDF Connector

Aquaforest
an ORPALIS company

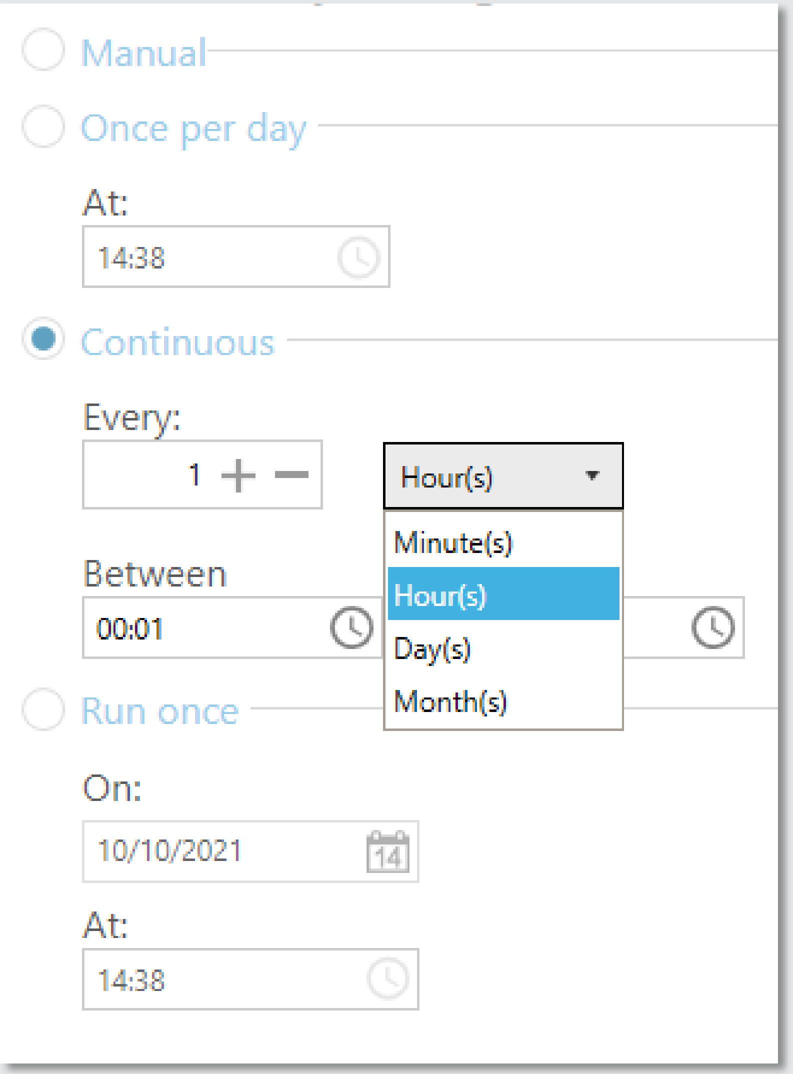
- **Autobahn** – High-performance high-volume automated OCR system
- **Searchlight OCR** – Automated in-place OCR processing tool
- **Searchlight Tagger** – Automated adding of metadata tags based on document content
- **Kingfisher** – Automated document organisation tool (split / rename / barcode / extract)
- **PDF Connector** – 10 OCR / PDF Steps in Microsoft's **Power Automate** platform

Bulk Processing Architecture

- All products run as a **background service**
- All products are **scalable by number of CPU cores used** (concurrently processed documents)
- Many of our clients have repositories of **many million documents** (Government, legal, financial, pharma…)
- Our clients predominantly use **automated scheduled processes** and only manage exceptions



Processing Mode:
○ Audit Only
● Audit and OCR
Cores: 4

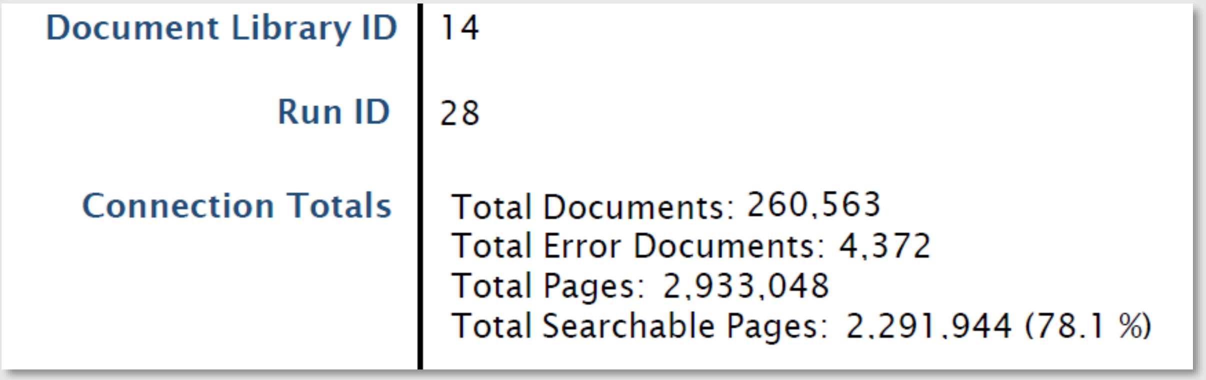


○ Manual

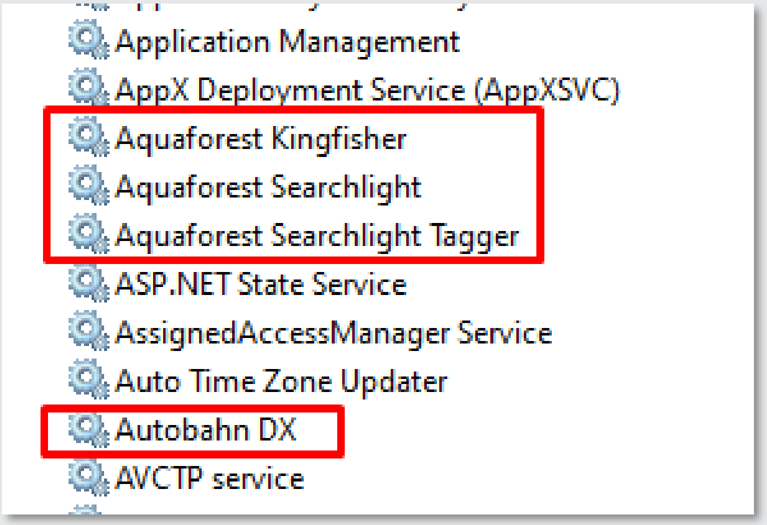
○ Once per day
At: 14:38

● Continuous
Every: 1 Hour(s)
Between 00:01

○ Run once
On: 10/10/2021
At: 14:38

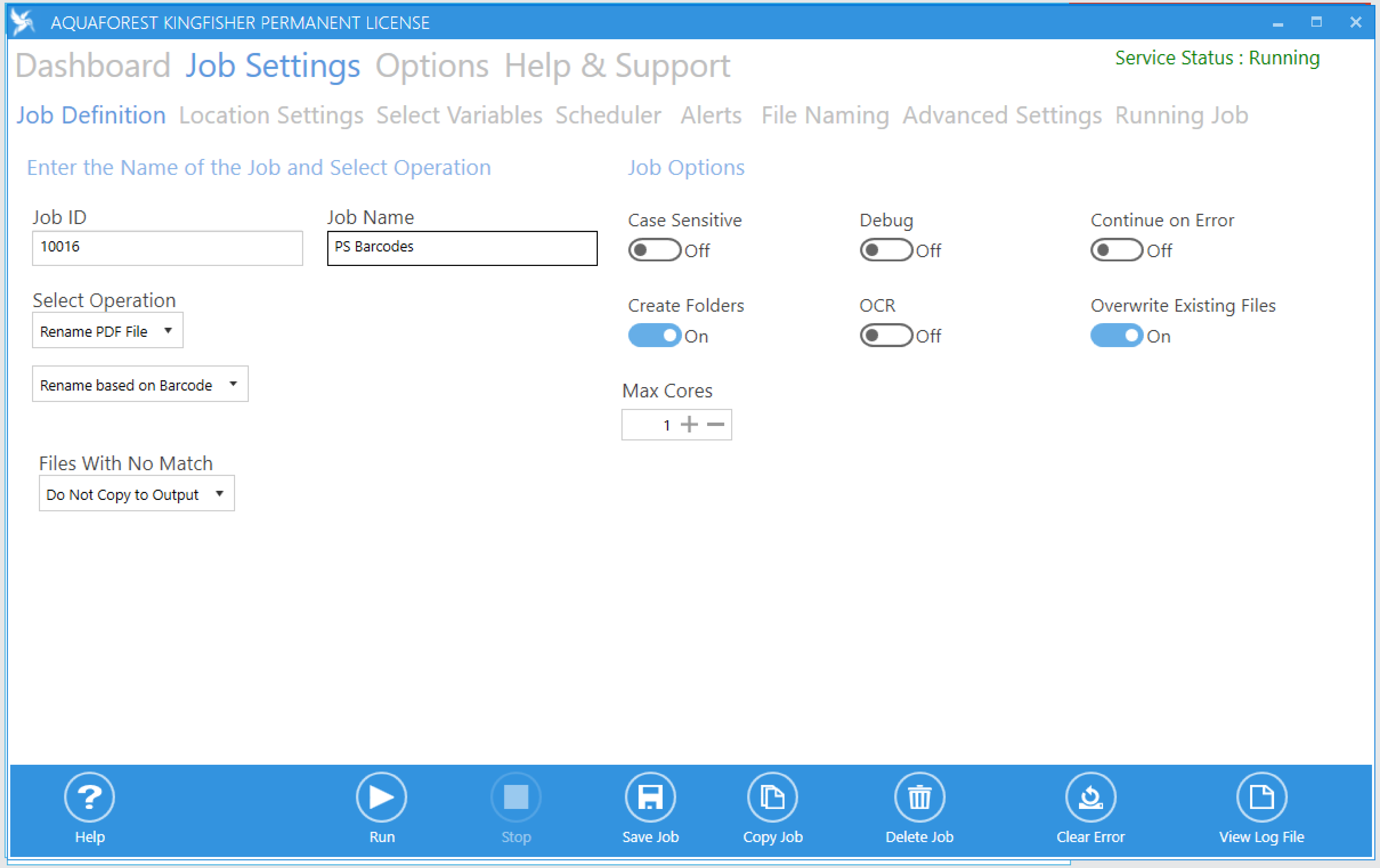


Document Library ID	14
Run ID	28
Connection Totals	Total Documents: 260,563 Total Error Documents: 4,372 Total Pages: 2,933,048 Total Searchable Pages: 2,291,944 (78.1 %)



Application Management
AppX Deployment Service (AppXSVC)
Aquaforest Kingfisher
Aquaforest Searchlight
Aquaforest Searchlight Tagger
ASP.NET State Service
AssignedAccessManager Service
Auto Time Zone Updater
Autobahn DX
AVCTP service

User Interfaces


AQUAFOREST KINGFISHER PERMANENT LICENSE
Dashboard
Job Settings
Options
Help & Support
Job Definition
Location Settings
Select Variables
Scheduler
Alerts
File Naming
Advanced Settings
Running Job
Help
Run
Stop
Save Job
Copy Job
Delete Job
Clear Error
View Log File

Service Status : Running

Enter the Name of the Job and Select Operation

Job ID: 10016

Job Name: PS Barcodes

Select Operation: Rename PDF File

Rename based on Barcode

Files With No Match: Do Not Copy to Output

Job Options

Case Sensitive: Off

Debug: Off

Continue on Error: Off

Create Folders: On

OCR: Off

Overwrite Existing Files: On

Max Cores: 1

- No coding required

- **But if you *really* want it:**
- Optional RegEx filters
- Execute scripts
- Command line
- ADX API
- SDK

Use Case ALPHA – Transport Company

Use Case ALPHA

What happens and what needed doing?

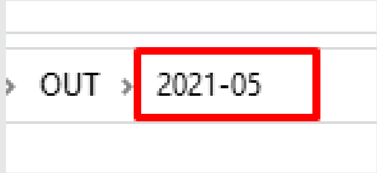
- Hundreds of pages are being scanned every day
- CRM generates cover sheet with main data and QR code
- Papers relating to the same shipment may be scanned at different times over several days
- Different shipments are scanned together in one process

- **We need to be able to find any individual shipment papers by**
  - **Console Reference**
  - **File Reference**
  - **Shipment Date**

In other words

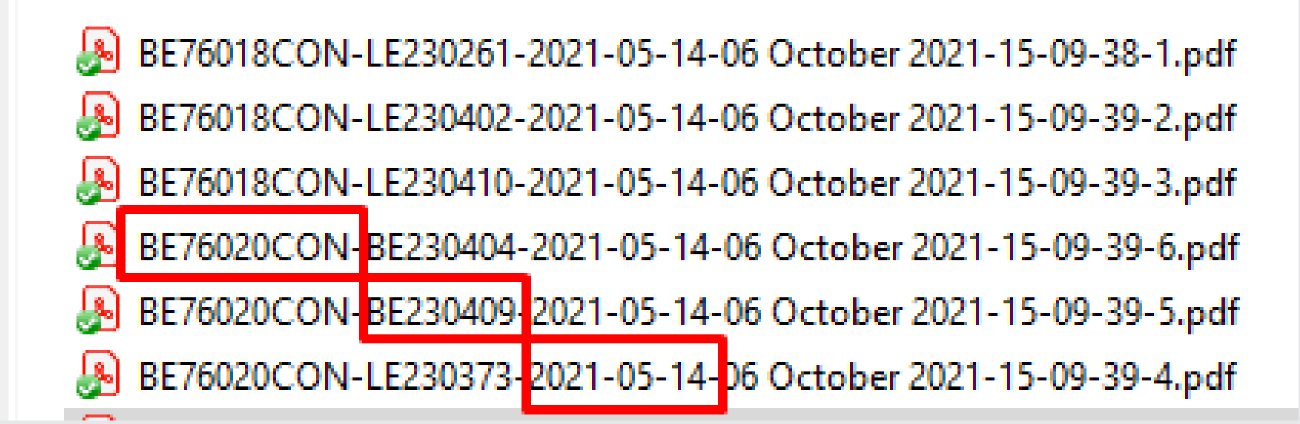
- We need to get from this

- To this for findability
  - Create folder and group by shipment year/month



OUT › 2021-05

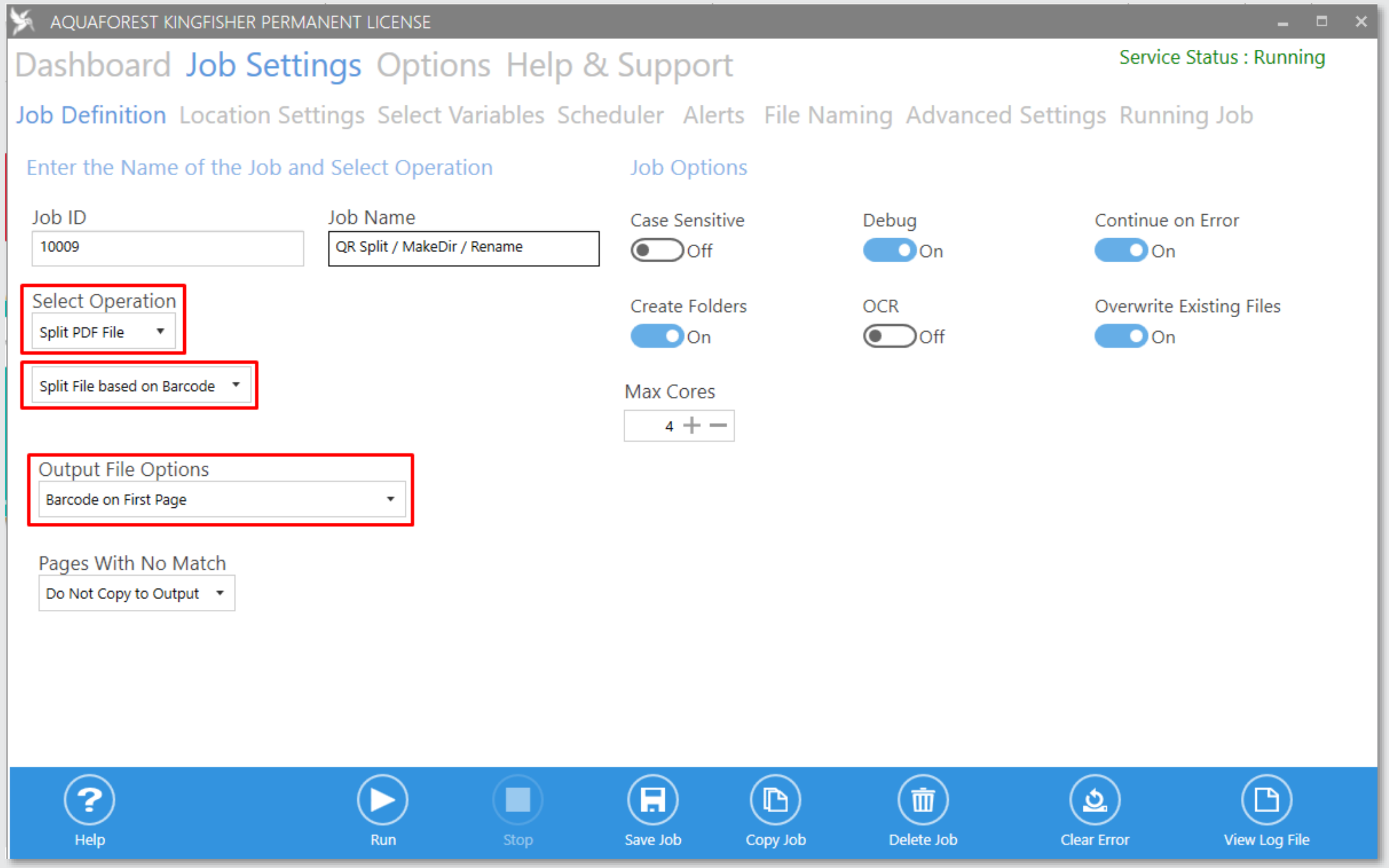
  - Feature all references in file name



BE76018CON-LE230261-2021-05-14-06 October 2021-15-09-38-1.pdf
BE76018CON-LE230402-2021-05-14-06 October 2021-15-09-39-2.pdf
BE76018CON-LE230410-2021-05-14-06 October 2021-15-09-39-3.pdf
BE76020CON-BE230404-2021-05-14-06 October 2021-15-09-39-6.pdf
BE76020CON-BE230409-2021-05-14-06 October 2021-15-09-39-5.pdf
BE76020CON-LE230373-2021-05-14-06 October 2021-15-09-39-4.pdf

Kingfisher 1 - Split by barcode

Windows Server 2012 R2
Windows 10
Windows Server 2016
Windows Server 2019
Aquaforest Kingfisher
© Aquaforest Limited 2021
Aquaforest
Version 2.5


AQUAFOREST KINGFISHER PERMANENT LICENSE
Dashboard
Job Settings
Options
Help & Support
Job Definition
Location Settings
Select Variables
Scheduler
Alerts
File Naming
Advanced Settings
Running Job
Help
Run
Stop
Save Job
Copy Job
Delete Job
Clear Error
View Log File

Service Status : Running

Enter the Name of the Job and Select Operation

Job ID: 10009

Job Name: QR Split / MakeDir / Rename

Select Operation: Split PDF File

Split File based on Barcode

Output File Options: Barcode on First Page

Pages With No Match: Do Not Copy to Output

Job Options

Case Sensitive: Off

Debug: On

Continue on Error: On

Create Folders: On

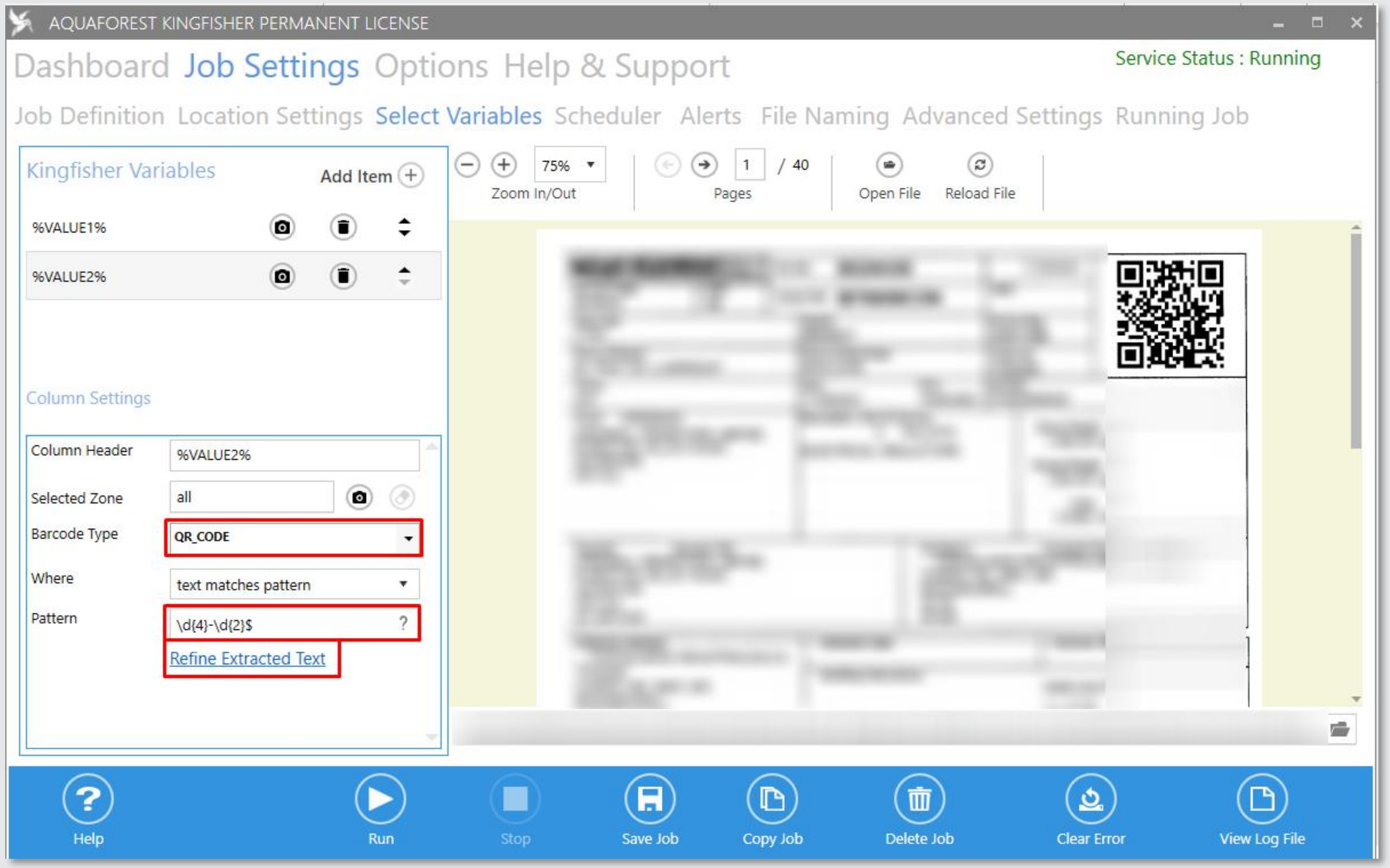
OCR: Off

Overwrite Existing Files: On

Max Cores: 4

Kingfisher 2 - Extract shipping date

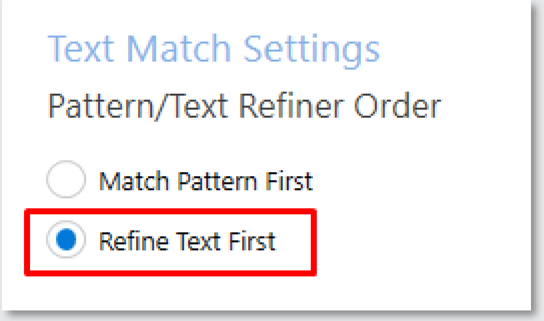
Windows Server 2012 R2
Windows 10
Windows Server 2016
Windows Server 2019
Aquaforest Kingfisher
© Aquaforest Limited 2021
Aquaforest
Version 2.5


AQUAFOREST KINGFISHER PERMANENT LICENSE
Dashboard
Job Settings
Options
Help & Support
Service Status : Running
Job Definition
Location Settings
Select Variables
Scheduler
Alerts
File Naming
Advanced Settings
Running Job
Kingfisher Variables
Add Item
%VALUE1%
%VALUE2%
Column Settings
75%
Zoom In/Out
1 / 40
Pages
Open File
Reload File
Help
Run
Stop
Save Job
Copy Job
Delete Job
Clear Error
View Log File

Column Header: %VALUE2%
Selected Zone: all
Barcode Type: QR_CODE
Where: text matches pattern
Pattern: `\d{4}-\d{2}$`
Refine Extracted Text

Kingfisher 3 - Form folder variable

Windows Server 2012 R2
Windows 10
Windows Server 2016
Windows Server 2019
Aquaforest Kingfisher
Aquaforest
© Aquaforest Limited 2021
Version 2.5

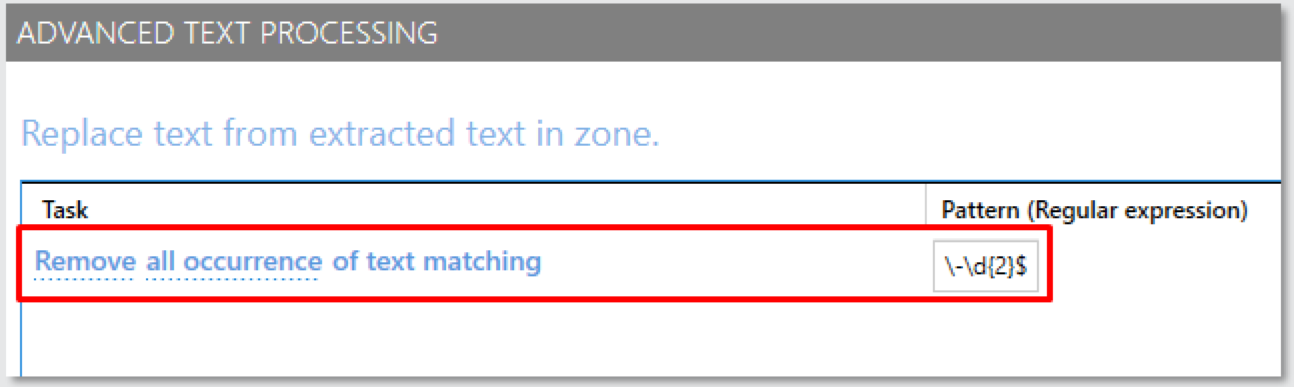


Text Match Settings

Pattern/Text Refiner Order

○ Match Pattern First

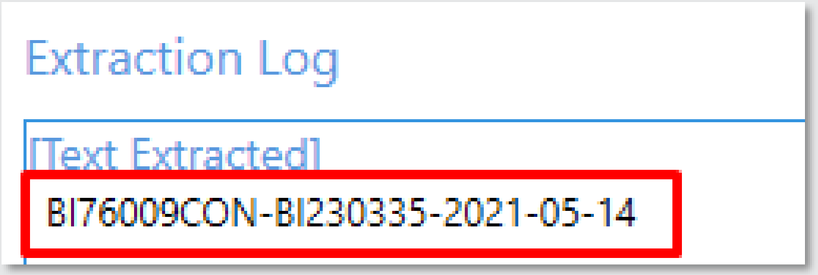
● Refine Text First



ADVANCED TEXT PROCESSING

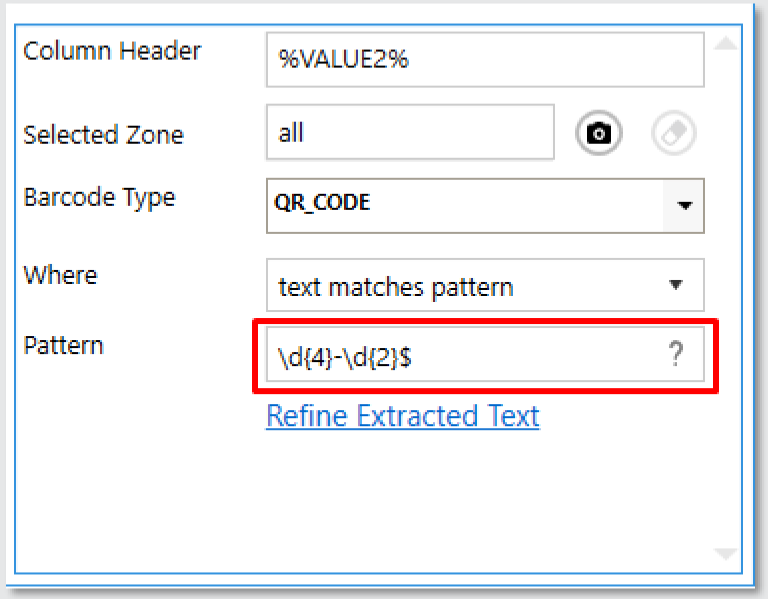
Replace text from extracted text in zone.

Task	Pattern (Regular expression)
Remove all occurrence of text matching	\-\d{2}$



Extraction Log

[Text Extracted]
BI76009CON-BI230335-2021-05-14



Column Header: %VALUE2%

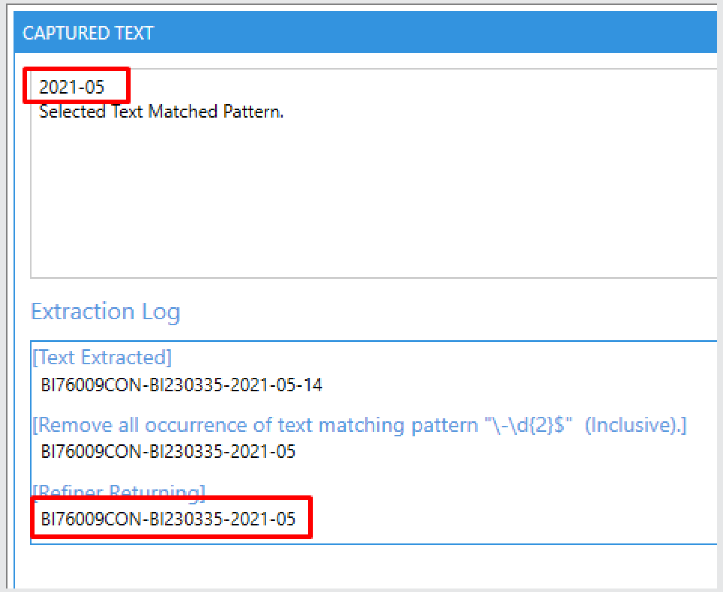
Selected Zone: all

Barcode Type: QR_CODE

Where: text matches pattern

Pattern: \d{4}-\d{2}$

?
Refine Extracted Text



CAPTURED TEXT

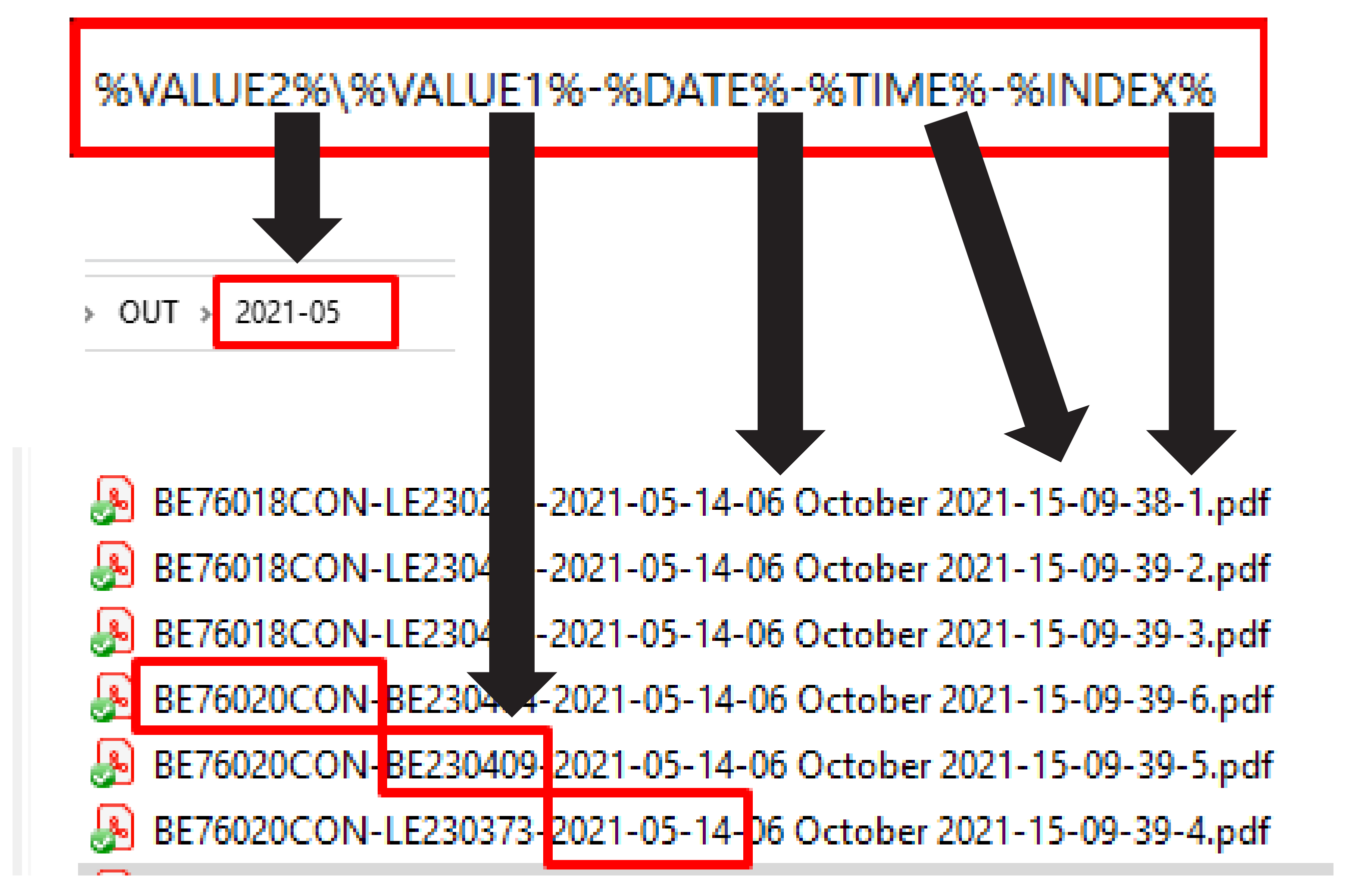
```
2021-05
Selected Text Matched Pattern.
```

Extraction Log

```
[Text Extracted]
BI76009CON-BI230335-2021-05-14
[Remove all occurrence of text matching pattern "\-\d{2}$"  (Inclusive).]
BI76009CON-BI230335-2021-05
[Refiner Returning]
BI76009CON-BI230335-2021-05
```

Kingfisher 4 - Make folder / Rename file

Windows Server 2012 R2
Windows 10
Windows Server 2016
Windows Server 2019
Aquaforest Kingfisher
Aquaforest
© Aquaforest Limited 2021
Version 2.5


%VALUE2%\%VALUE1%-%DATE%-%TIME%-%INDEX%

OUT › 2021-05

BE76018CON-LE2302 -2021-05-14-06 October 2021-15-09-38-1.pdf
BE76018CON-LE2304 -2021-05-14-06 October 2021-15-09-39-2.pdf
BE76018CON-LE2304 -2021-05-14-06 October 2021-15-09-39-3.pdf
BE76020CON-BE2304 4-2021-05-14-06 October 2021-15-09-39-6.pdf
BE76020CON-BE230409-2021-05-14-06 October 2021-15-09-39-5.pdf
BE76020CON-LE230373-2021-05-14-06 October 2021-15-09-39-4.pdf

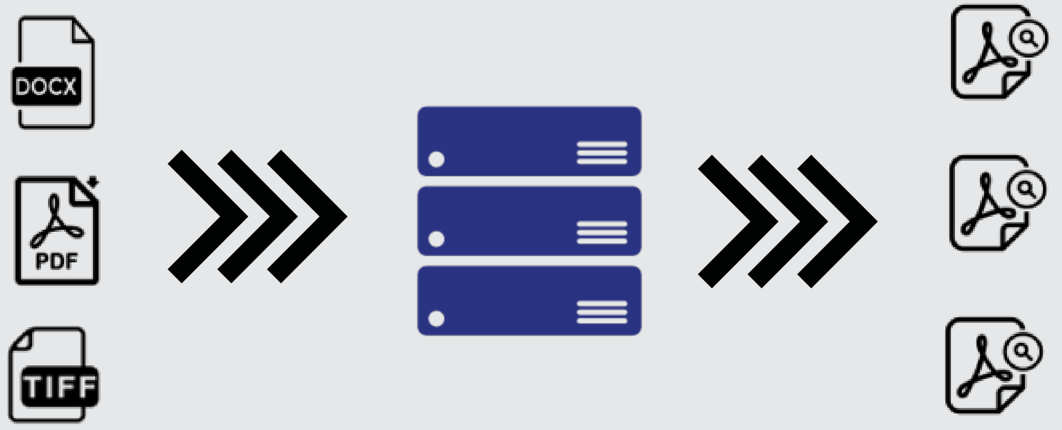
Use Case BRAVO – Bank

What happens and what needed doing?

- Regional European Bank with 3,000 employees
- Much of customer communication is done via email
- Maximum attachment size for emails is 20MB
- Many contracts get scanned and documents get combined in PDFs resulting in files larger than 20MB → cannot be emailed
- **Bank needs an effective compression tool allowing each employee individually to compress files as and when needed**
- System needs to reside within the Bank’s IT infrastructure

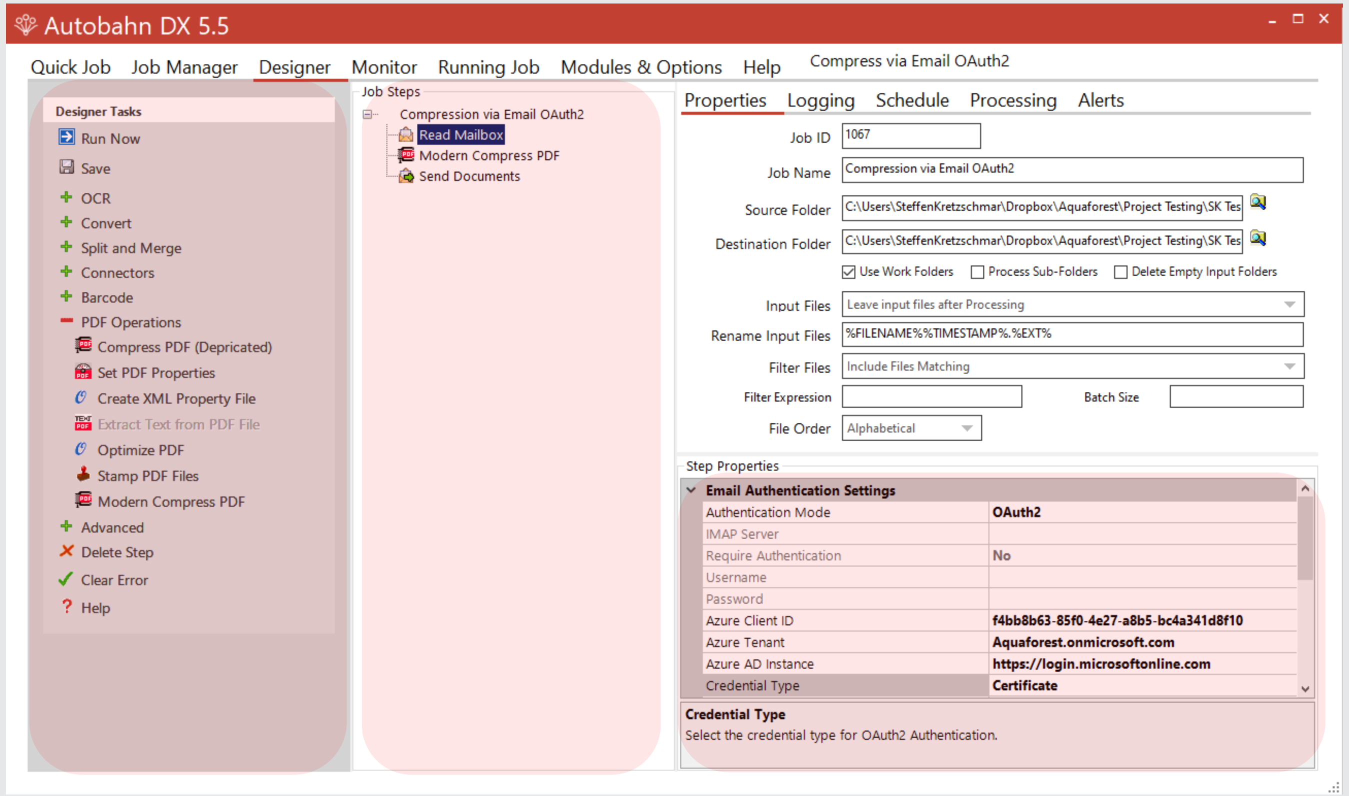
Autobahn – Overview

- **Workflow Application**
- Users **add Job Steps** from the Task Menu
- **Manage job steps order**
- Set **step properties**


DOCX
PDF
TIFF

INPUT Location

OUTPUT Location


Autobahn DX 5.5
Quick Job
Job Manager
Designer
Monitor
Running Job
Modules & Options
Help
Designer Tasks
Run Now
Save
OCR
Convert
Split and Merge
Connectors
Barcode
PDF Operations
Compress PDF (Depricated)
Set PDF Properties
Create XML Property File
Extract Text from PDF File
Optimize PDF
Stamp PDF Files
Modern Compress PDF
Advanced
Delete Step
Clear Error
Help
Job Steps
Compression via Email OAuth2
Read Mailbox
Modern Compress PDF
Send Documents
Properties
Logging
Schedule
Processing
Alerts
Step Properties

Compress via Email OAuth2

Job ID: 1067

Job Name: Compression via Email OAuth2

Source Folder: C:\Users\SteffenKretzschmar\Dropbox\Aquaforest\Project Testing\SK Tes

Destination Folder: C:\Users\SteffenKretzschmar\Dropbox\Aquaforest\Project Testing\SK Tes

☑ Use Work Folders ☐ Process Sub-Folders ☐ Delete Empty Input Folders

Input Files: Leave input files after Processing

Rename Input Files: %FILENAME%%TIMESTAMP%.%EXT%

Filter Files: Include Files Matching

Filter Expression:

Batch Size:

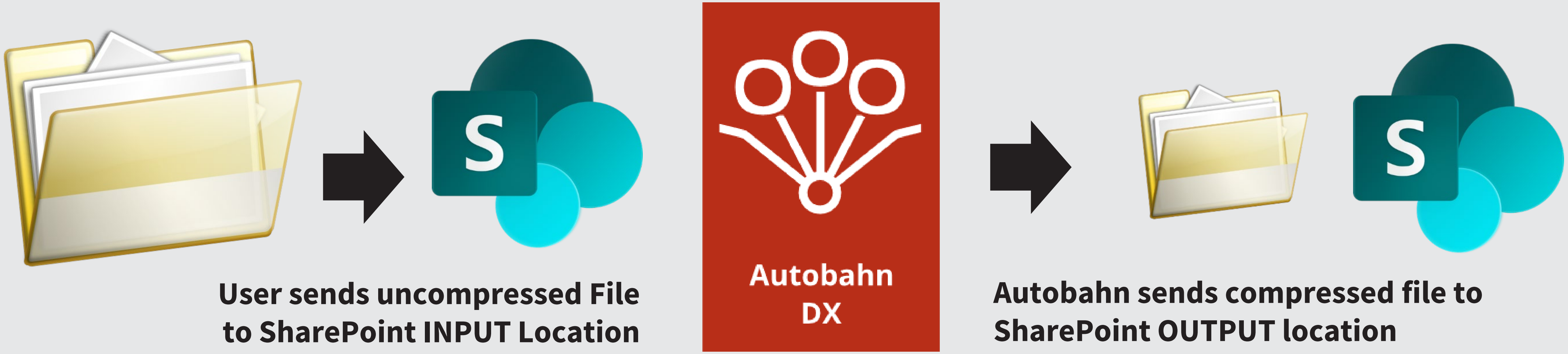
File Order: Alphabetical

Email Authentication Settings	
Authentication Mode	**OAuth2**
IMAP Server	
Require Authentication	No
Username	
Password	
Azure Client ID	**f4bb8b63-85f0-4e27-a8b5-bc4a341d8f10**
Azure Tenant	**Aquaforest.onmicrosoft.com**
Azure AD Instance	**https://login.microsoftonline.com**
Credential Type	**Certificate**

Credential Type

Select the credential type for OAuth2 Authentication.

Autobahn – Compress via SharePoint



User sends uncompressed File to SharePoint INPUT Location

Autobahn DX

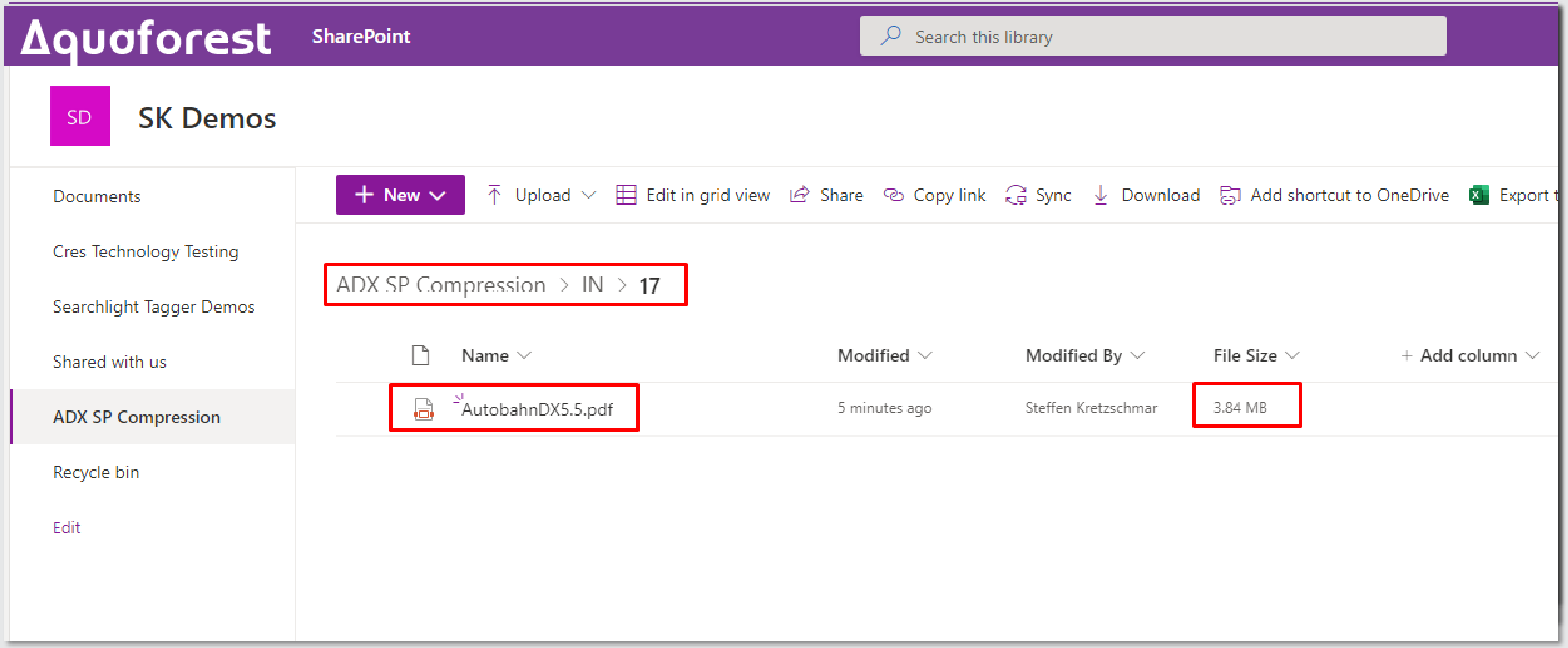
Autobahn sends compressed file to SharePoint OUTPUT location

- Employees move their uncompressed files to a specific SharePoint **INPUT** location
- An Autobahn process is scheduled at certain intervals to monitor this SharePoint location
- Compressed files are sent to a specific SharePoint **OUTPUT** location (rename option)
- Input files can be archived, deleted or renamed (Power Automate)
- Emails can be sent once files arrive in OUTPUT location (Power Automate)

Autobahn – SharePoint INPUT Folder

PDF Days Online 2021

Autobahn DX


Aquaforest
SharePoint
Search this library
SD
SK Demos
Documents
Cres Technology Testing
Searchlight Tagger Demos
Shared with us
ADX SP Compression
Recycle bin
Edit
New
Upload
Edit in grid view
Share
Copy link
Sync
Download
Add shortcut to OneDrive
Export t
Add column

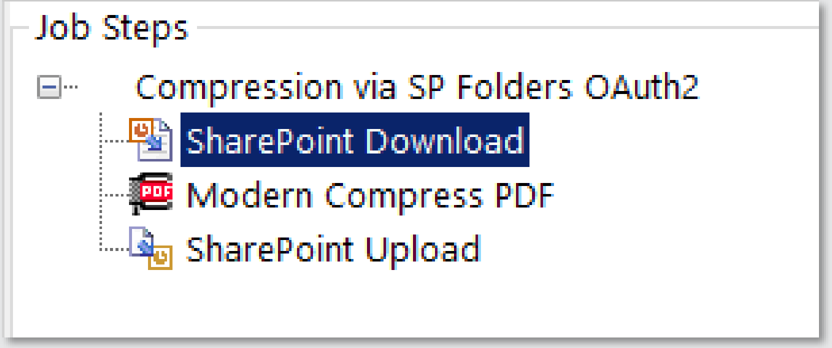
ADX SP Compression > IN > 17

Name	Modified	Modified By	File Size
AutobahnDX5.5.pdf	5 minutes ago	Steffen Kretzschmar	3.84 MB

Autobahn – ADX Process - Download

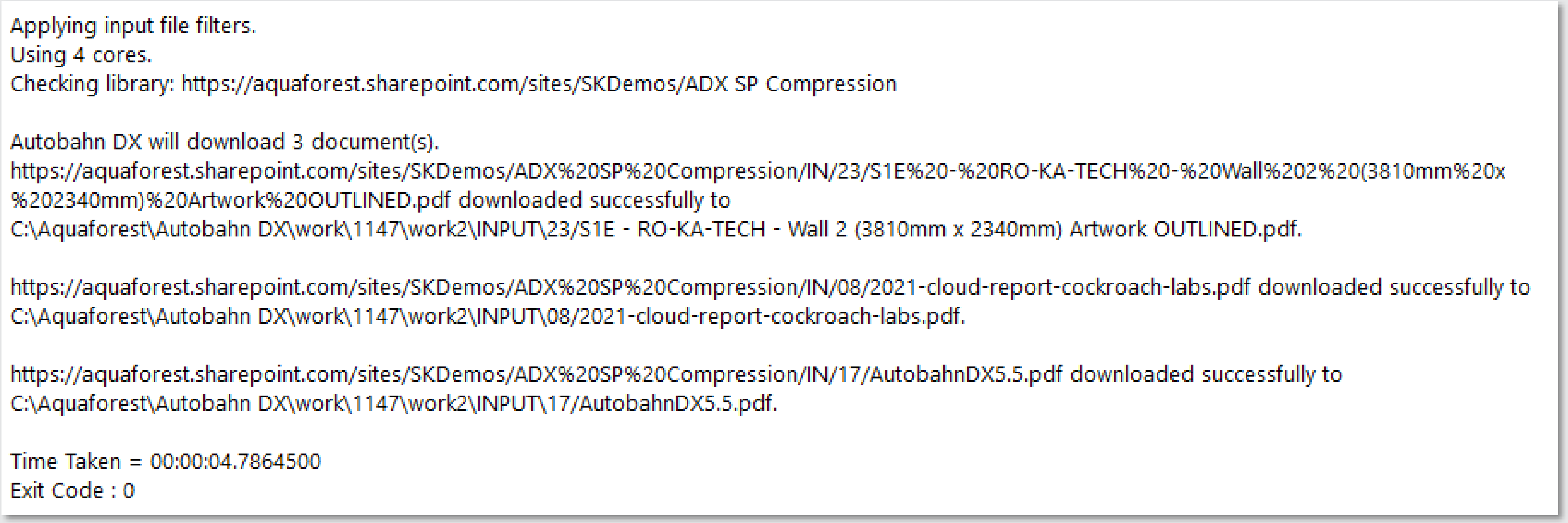
PDF Days Online 2021

Autobahn DX



Job Steps

- Compression via SP Folders OAuth2
  - SharePoint Download
  - Modern Compress PDF
  - SharePoint Upload

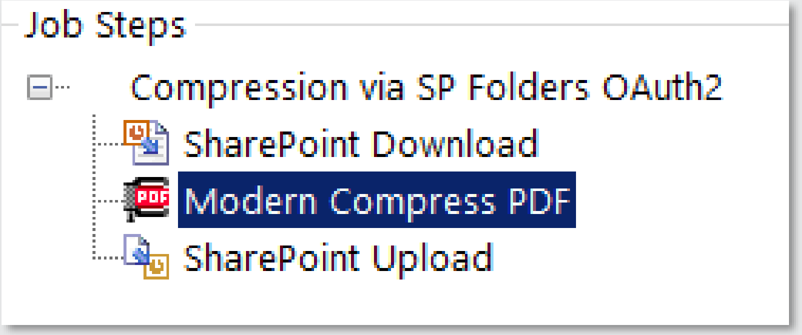


```
Applying input file filters.
Using 4 cores.
Checking library: https://aquaforest.sharepoint.com/sites/SKDemos/ADX SP Compression

Autobahn DX will download 3 document(s).
https://aquaforest.sharepoint.com/sites/SKDemos/ADX%20SP%20Compression/IN/23/S1E%20-%20RO-KA-TECH%20-%20Wall%202%20(3810mm%20x
%202340mm)%20Artwork%20OUTLINED.pdf downloaded successfully to
C:\Aquaforest\Autobahn DX\work\1147\work2\INPUT\23/S1E - RO-KA-TECH - Wall 2 (3810mm x 2340mm) Artwork OUTLINED.pdf.

https://aquaforest.sharepoint.com/sites/SKDemos/ADX%20SP%20Compression/IN/08/2021-cloud-report-cockroach-labs.pdf downloaded successfully to
C:\Aquaforest\Autobahn DX\work\1147\work2\INPUT\08/2021-cloud-report-cockroach-labs.pdf.

https://aquaforest.sharepoint.com/sites/SKDemos/ADX%20SP%20Compression/IN/17/AutobahnDX5.5.pdf downloaded successfully to
C:\Aquaforest\Autobahn DX\work\1147\work2\INPUT\17/AutobahnDX5.5.pdf.

Time Taken = 00:00:04.7864500
Exit Code : 0
```

Autobahn – ADX Process - Compress

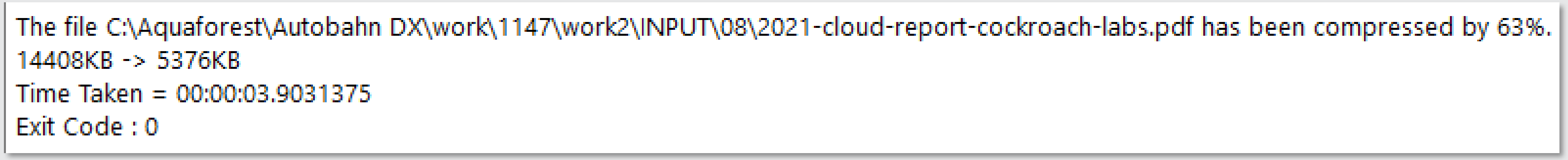
PDF Days Online 2021

Autobahn DX

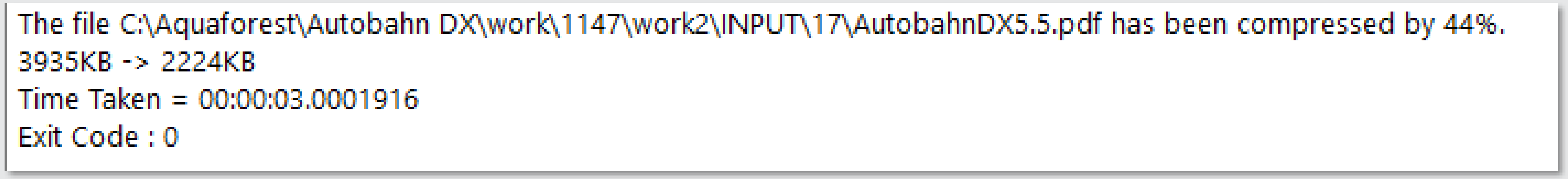


Job Steps

- Compression via SP Folders OAuth2
  - SharePoint Download
  - Modern Compress PDF
  - SharePoint Upload



```
The file C:\Aquaforest\Autobahn DX\work\1147\work2\INPUT\08\2021-cloud-report-cockroach-labs.pdf has been compressed by 63%.
14408KB -> 5376KB
Time Taken = 00:00:03.9031375
Exit Code : 0
```

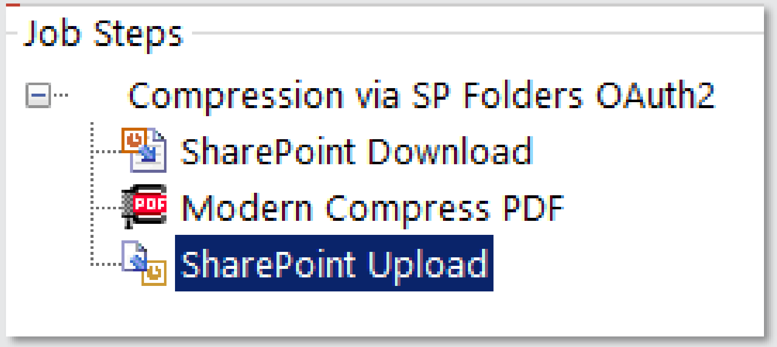


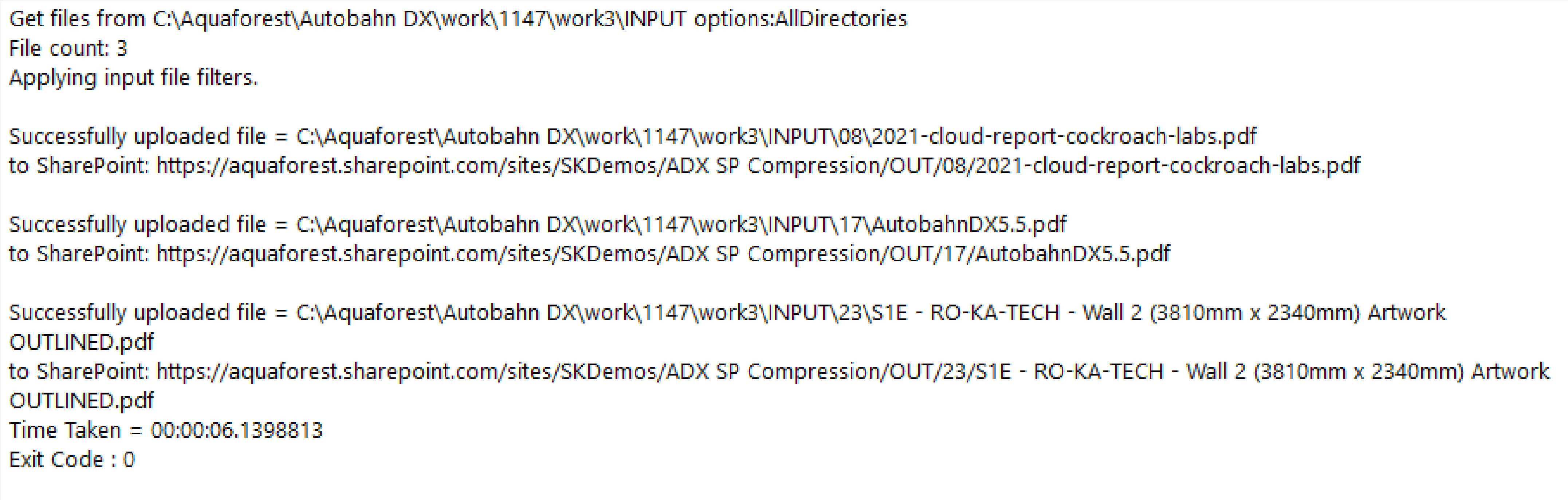
```
The file C:\Aquaforest\Autobahn DX\work\1147\work2\INPUT\17\AutobahnDX5.5.pdf has been compressed by 44%.
3935KB -> 2224KB
Time Taken = 00:00:03.0001916
Exit Code : 0
```

Autobahn – ADX Process - Upload

PDF Days Online 2021

Autobahn DX



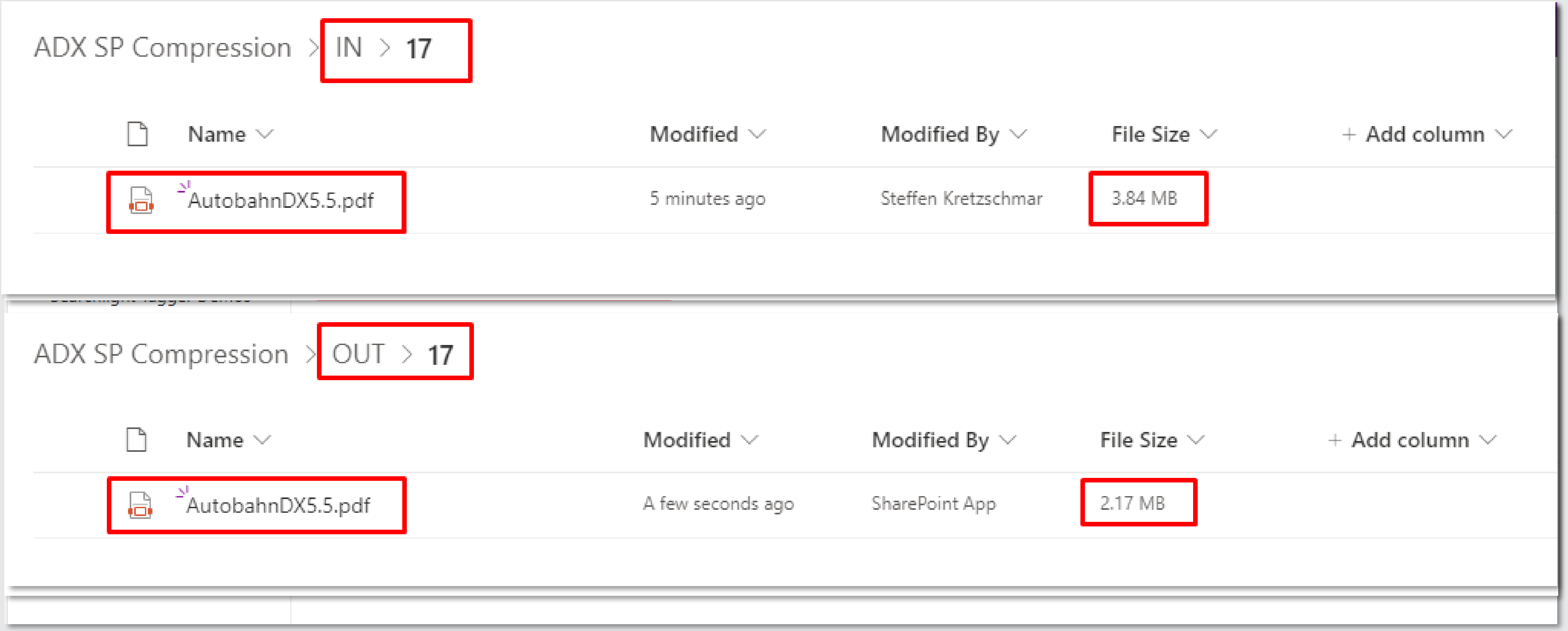


```
Get files from C:\Aquaforest\Autobahn DX\work\1147\work3\INPUT options:AllDirectories
File count: 3
Applying input file filters.

Successfully uploaded file = C:\Aquaforest\Autobahn DX\work\1147\work3\INPUT\08\2021-cloud-report-cockroach-labs.pdf
to SharePoint: https://aquaforest.sharepoint.com/sites/SKDemos/ADX SP Compression/OUT/08/2021-cloud-report-cockroach-labs.pdf

Successfully uploaded file = C:\Aquaforest\Autobahn DX\work\1147\work3\INPUT\17\AutobahnDX5.5.pdf
to SharePoint: https://aquaforest.sharepoint.com/sites/SKDemos/ADX SP Compression/OUT/17/AutobahnDX5.5.pdf

Successfully uploaded file = C:\Aquaforest\Autobahn DX\work\1147\work3\INPUT\23\S1E - RO-KA-TECH - Wall 2 (3810mm x 2340mm) Artwork
OUTLINED.pdf
to SharePoint: https://aquaforest.sharepoint.com/sites/SKDemos/ADX SP Compression/OUT/23/S1E - RO-KA-TECH - Wall 2 (3810mm x 2340mm) Artwork
OUTLINED.pdf
Time Taken = 00:00:06.1398813
Exit Code : 0
```

PDF association

Autobahn – SharePoint OUTPUT Folder

Autobahn DX



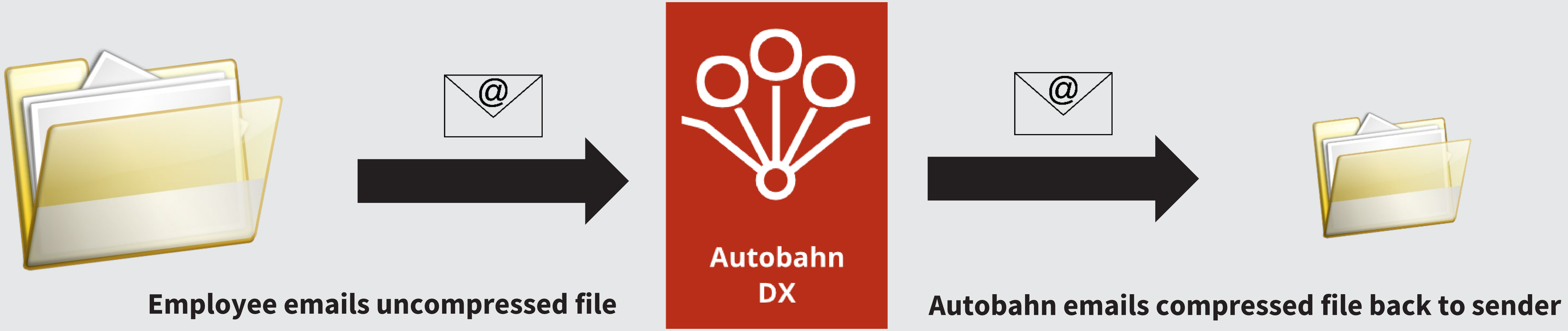
ADX SP Compression > IN > 17

	Name	Modified	Modified By	File Size	+ Add column
	AutobahnDX5.5.pdf	5 minutes ago	Steffen Kretzschmar	3.84 MB	

ADX SP Compression > OUT > 17

	Name	Modified	Modified By	File Size	+ Add column
	AutobahnDX5.5.pdf	A few seconds ago	SharePoint App	2.17 MB	

Autobahn – Compress via Email



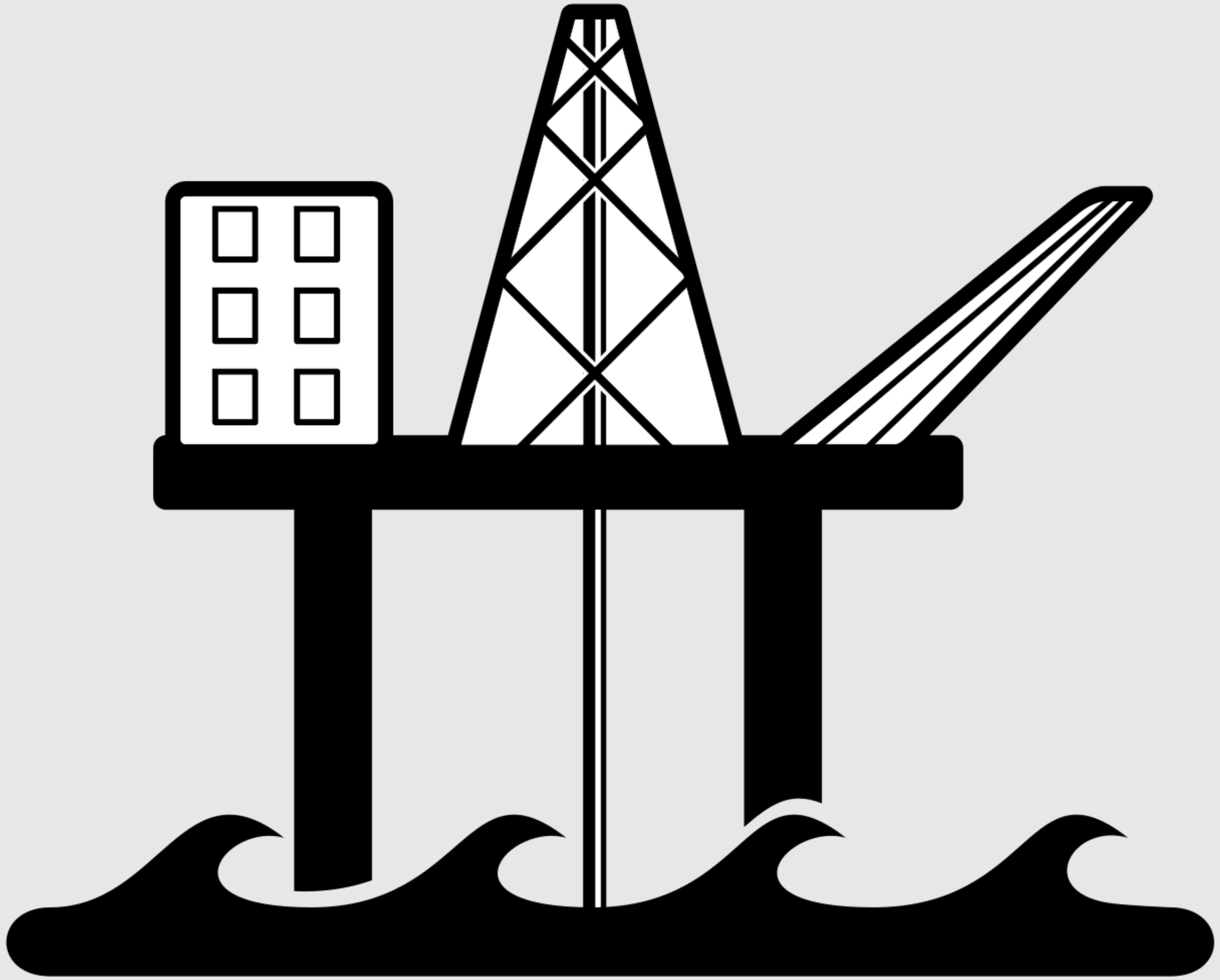
Employee emails uncompressed file

Autobahn DX

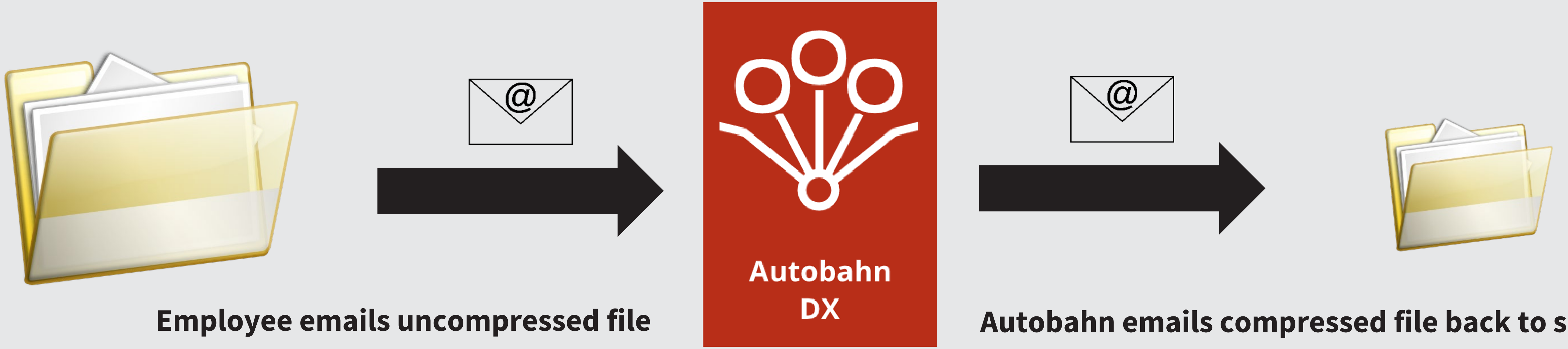
Autobahn emails compressed file back to sender

- Employees send their uncompressed to a specific Email address (e.g. compress@bank.com)
- An Autobahn process is scheduled to check that inbox at certain intervals or continuously
- Compressed files are sent back to the sender automatically once that process is complete
- That, of course, won’t work for the Bank if the uncompressed file exceeds the 20MB limit ☺

Use Case CHARLIE – Oil Authority

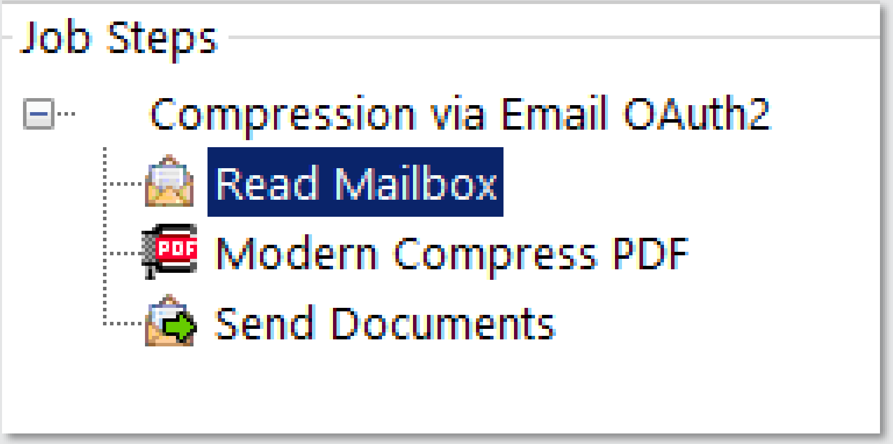


Autobahn – Compress via Email



Employee emails uncompressed file

Autobahn emails compressed file back to sender



Compression via Email OAuth2
- Read Mailbox
- Modern Compress PDF
- Send Documents

Thank you

Over 2,000 Organizations Worldwide Rely on Aquaforest Software

BBC Worldwide
BECHTEL
BlueShield of California
The Great Seal of the State of California
COMMERZBANK
Cornell University
FIDELITY NATIONAL FINANCIAL
hp
IRS
LEGO
NASA
OPENTEXT
Pfizer
SIEMENS
Time Warner Cable
Department of State United States of America

www.aquaforest.com

Aquaforest
an ORPALIS company