



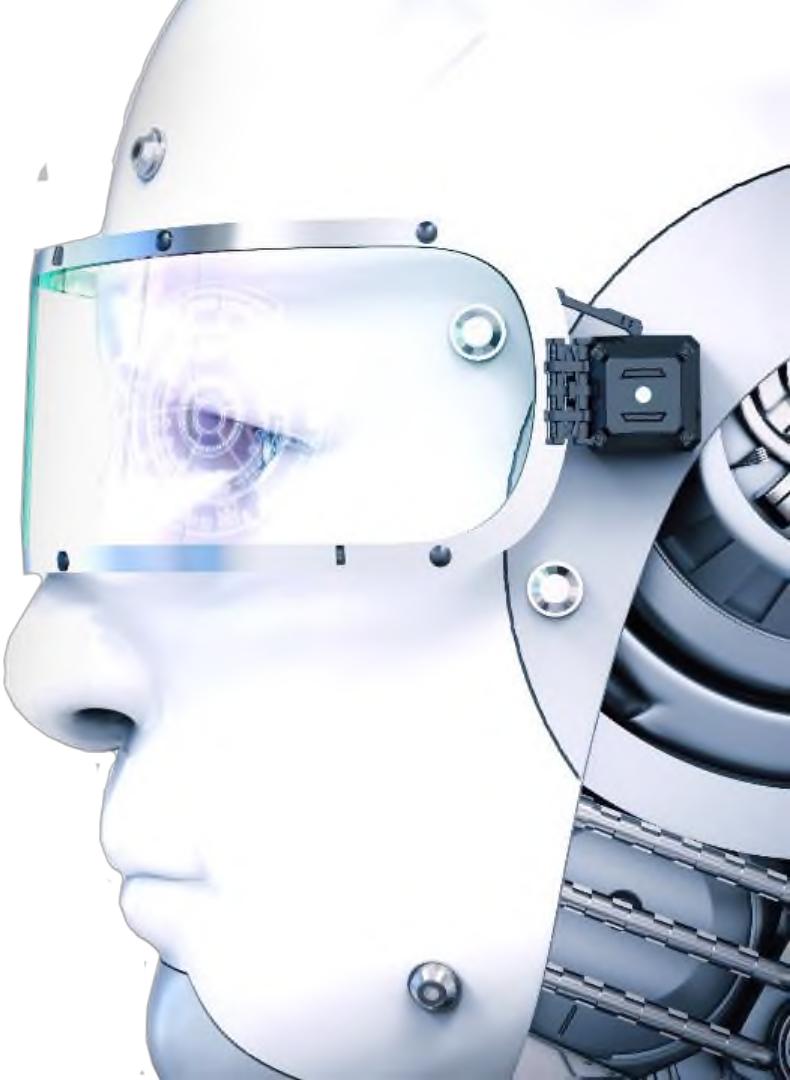
## ABBYY FLEXICAPTURE

---

### OVERVIEW



**Find, extract and action information**  
by intelligent capture, analysis  
and understanding data



# ABBYY in 30 seconds

ABBYY®

1

**Thousands of businesses worldwide** rely on ABBYY for converting unstructured data into structured information

2

ABBYY sets the global standard in **document recognition, content capture and language-based technologies and solutions** that integrate across the information lifecycle

3

ABBYY's innovative **Intelligent Capture** technology, products, solutions and services are delivered in multiple platforms - **cloud, mobile, desktop, server or SDK** - directly or via a global partner network

**40+ million users**  
**from 200+ countries**  
**and territories**

Technology innovator with  
**600+ developers, 100+**  
**patents and 100+ patent**  
**applications**

**500+ partners worldwide**

ABBYY was founded in **1989**

**1250+ people**  
in **16 offices**  
located in **13 countries**  
serve ABBYY customers  
in **200+ countries and territories**



# Intelligent document processing PEAK Matrix by Everest Group – ABBYY positioned as leader

## Everest Group PEAK Matrix®

Intelligent Document Processing (IDP) Products PEAK Matrix® Assessment 2020 |  
ABBYY positioned as Leader



Everest Group Intelligent Document Processing (IDP) Products PEAK Matrix® Assessment 2020



# A Global Business

---



**ABBYY**<sup>®</sup>

Leading Global Hardware and Software Developers Trust  
ABBYY Technologies

**Panasonic**

**TOSHIBA**  
Leading Innovation >>>

**MICROTEK**

**SAMSUNG**

  
**KONICA MINOLTA**

**KYOCERA**

**Laserfiche**<sup>®</sup>

  
**SCANTRON**<sup>™</sup>

  
**KnowledgeLake**

  
**FUJI XEROX**

**intuit.**

**FUJITSU**

**SHARP**

  
**LG**

  
**nuix**

  
**brother**  
At your side.

**RICOH**

  
**plustek**

**EMC<sup>2</sup>**

  
**FileBound**<sup>®</sup>  
Document Management Solutions

# Award-winning Recognition and Linguistic Technologies

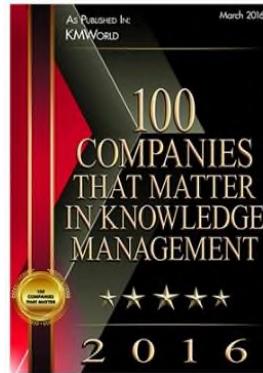
---

ABBYY OCR, data capture, mobile and linguistic products received

## 260+ awards

(more than other industry players)

Key global IT, business publications and analysts consider ABBYY to be the provider of leading document recognition technologies and solutions.



# FlexiCapture Product Line

---



# FlexiCapture Product line and delivery options

ABBYY®



Enterprises/Mid-size



SI, ISVs

## FlexiCapture

Enterprise-grade platform

## FlexiCapture SDK

SDK for complicated data capture: document types with various layouts, invoices, full FC functionality

## FlexiCapture Cloud, available via REST API

FlexiCapture, deployed in MS Azure

ONPREM

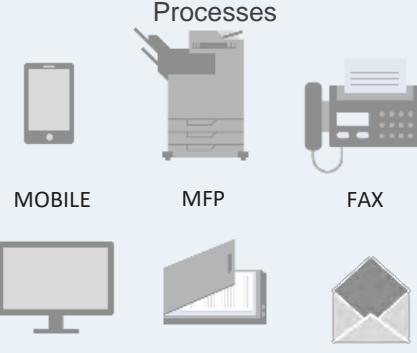
CLOUD

All delivery options are available for **Invoice processing**

# ABBYY FlexiCapture Enables Complex Automation Environments to Work

## Customer Interactions

Real-time, Content Centric Processes

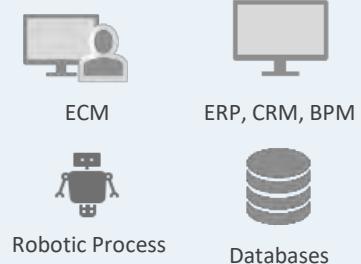


Structured & Unstructured Content

ABBYY provides the intelligence to turn unstructured content into structured, actionable information, transforming processes for a better customer experience, lower costs, lower risk and competitive advantage.

## Systems and Robots

Enterprise applications and digital robotic processes designed to run businesses



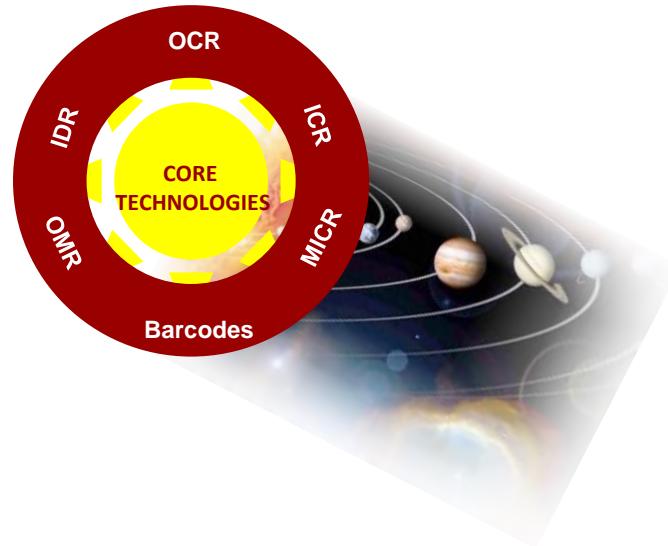
Digitization occurs at the Knowledge Worker's point of need (not IT mandate) – the platform is for making business units more effective in serving their customers, facilitating meaningful change

# ABBYY Technologies

---



- ABBYY develops a wide range of high-tech products and services that are based on document recognition and linguistic technologies (core technologies)



# What is OCR?

- OCR stands for Optical Character Recognition
- This is a set of methods and techniques used to convert text printed on paper into editable computer text (i.e. ASCII or Unicode characters, not image files)

see the snake, so ev  
ick. The sound of  
metal box, which wa



Image



Text Symbol (Code ASCII – 53)

just see the snake, so every nig  
back. The sound of metal falling  
the metal box, which was the sna  
e snake came through the air-v

# Text Recognition Principles:

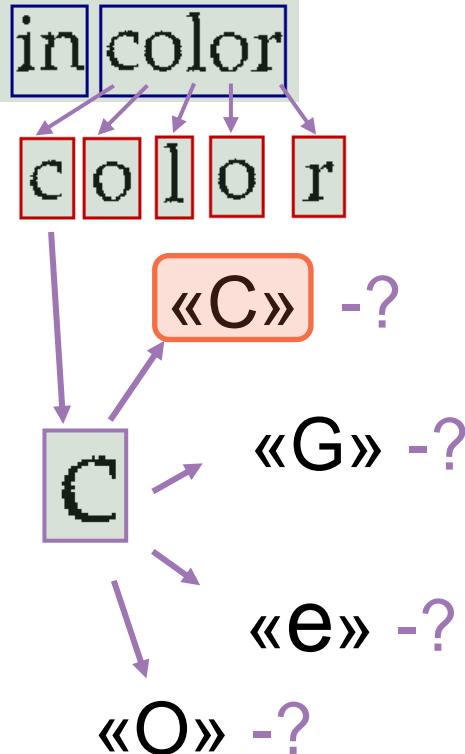
## Character Recognition

pages are displayed in color

- The system generates multiple hypotheses about the characters and words, which have different ratings (or “weights”)
- The characters are recognized based on their models



- Each character is analyzed by several classifiers for maximum accuracy



# Industries and use cases

---





- Customer onboarding
- Loan origination
- Mortgage
- KYC
- Financial Crime Compliance, OFAC
- Direct debit requests, retirement fund management, etc.



- Multichannel input and automated processing enables customers to communicate with insurance company when they want and at the speed they want
- Customer onboarding
- Insurance claims processing
- Underwriting



- Shorten the “Last Mile” in Goods Delivery:
- Shipping documents
  - Customs documents
  - Accounts payable
  - PODs (proof of delivery)
  - Technical documentation on vehicles



- Legacy patient records
- Classification clinical care documents (patient charts, admissions forms, encounter forms, lab results)
- Processing business office documents (claims, EOB, payments)
- Pharma: side-effects analyses



- Loan applications
- Lease agreements

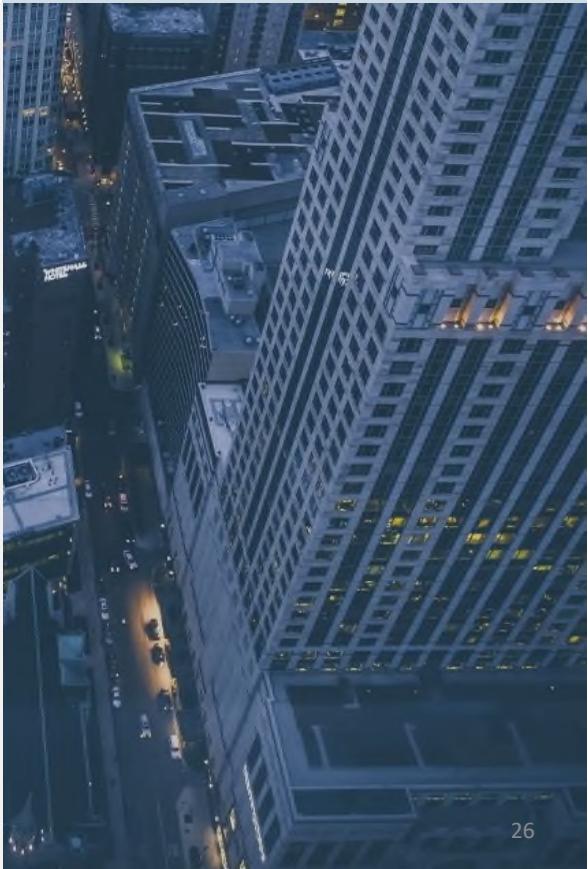


- Automatic claims classification
- Efficiently manage documents, including: tests, examinations reports, letters of reference
- Blanks, governmental forms processing
- High-volume conversion of scanned paper, images, PDFs to editable documents, searchable archives creation

## Filing Tax Returns

Intuit, an American IT company, has incorporated ABBYY technology into its mobile SnapTax application.

This application lets tax payers file their tax returns in just a few minutes. Data is transferred from one document to another by means of your smartphone's camera and ABBYY FlexiCapture Engine technology.



PRODUCTS & SOLUTIONS

EXPLORE PARTNERS SUPPORT & SERVICES COMPANY 

**EXPLORE SOLUTIONS** 

**Solutions by Industry**  
Financial Services, Healthcare, Insurance and more

**Solutions by process**  
Browse specific use cases and integrations by departments or business functions

**Solutions by technology**  
Learn how ABBYY works with other technologies

**DIGITAL INTELLIGENCE** 

**Process Intelligence**  
Easily discover, understand, and manage business process execution more effectively

**Content Intelligence**  
Content skills and understanding for digital workforces

**ENTERPRISE APPLICATIONS** 

**Timeline**  
Process Intelligence for true visibility into processes

**Vantage**  
Platform delivering Content Intelligence skills to make the digital workforce smarter

**FineReader Server**  
Deploy server-based, large volume OCR solution for document conversion

**FlexiCapture**  
Capture data from any documents, from structured forms and surveys to unstructured text-heavy papers

**FlexiCapture for Invoices**  
Ready-to-run accounts payable automation solution

**DEVELOPER TOOLS** 

**Data Capture**  
Power your document applications with content intelligence via seamless SDK integration

[FlexiCapture SDK](#) ›

[FlexiCapture Cloud SDK](#) ›

**Document Conversion**  
Enhance your applications with OCR, ICR and PDF conversion functions

[FineReader Engine](#) ›

[Cloud OCR SDK](#) ›

**Mobile Capture**  
Enhance your mobile onboarding and digital self-service offerings

[Mobile Capture](#) ›

[Mobile Web Capture](#) ›

**NeoML**  
An open-source machine learning framework

**EMPLOYEE PRODUCTIVITY** 

**FineReader PDF**   
Edit, convert, and compare PDFs and scans



Interview with Jim Sinur, Forbes Contributor, former Gartner Analyst, CEO at Flueresque



## ABBYY FLEXICAPTURE

---

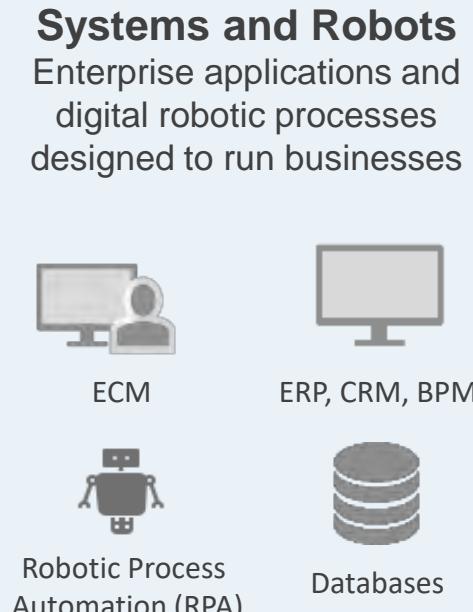
### PRODUCT OVERVIEW



# ABBYY FlexiCapture Enables Complex Automation Environments to Work



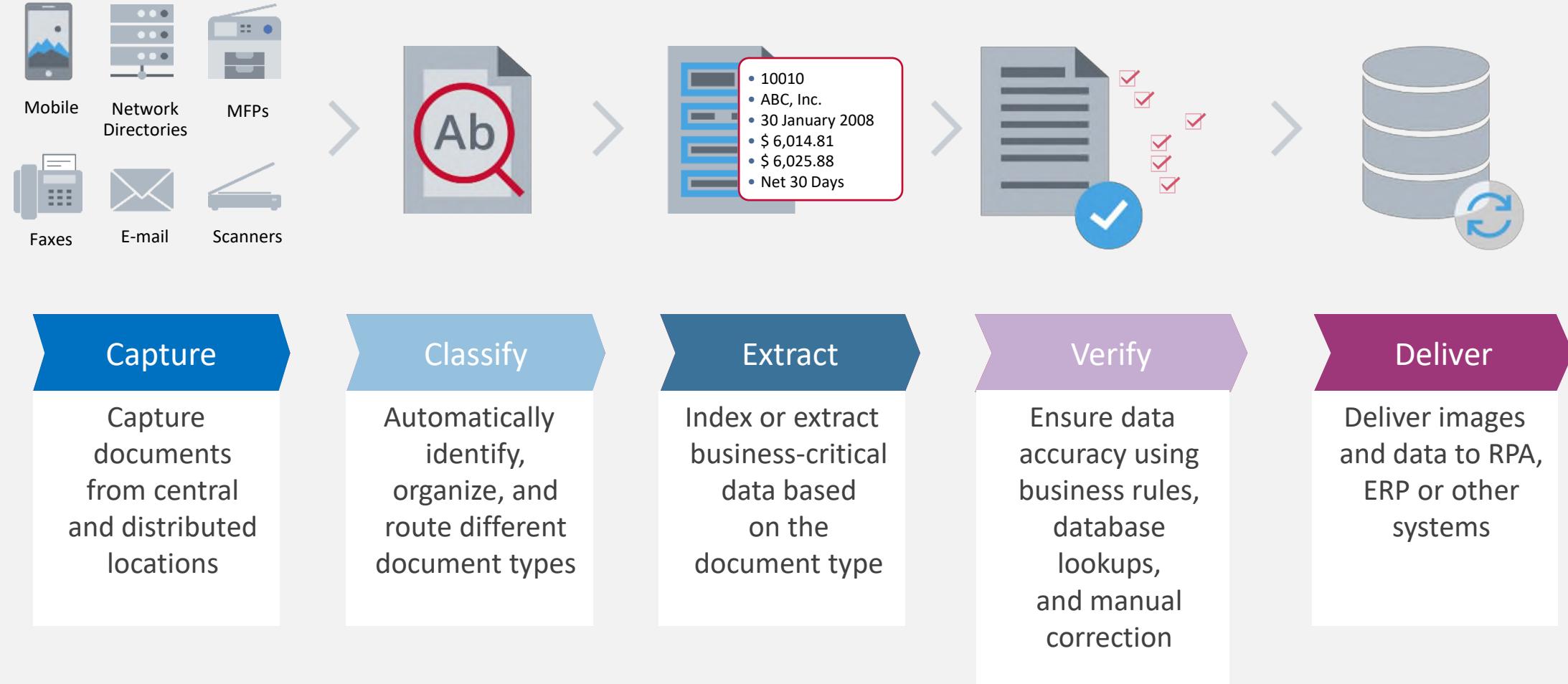
ABBYY provides the intelligence to turn unstructured content into structured, actionable information, transforming processes for a better customer experience, lower costs, lower risk and competitive advantage.



Digitization occurs at the Knowledge Worker's point of need (not IT mandate) – the platform is for making business units more effective in serving their customers, facilitating meaningful change

# Typical Capture Flow

ABBYY®



# ABBYY FlexiCapture - Key benefits

ABBYY®



Accelerate transactions, decisions and actions



Ensure control, predictability and compliance



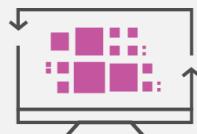
Maximise resources



Take advantage of high-quality data output



Rapidly build and run multiple applications



Scale and customize according to needs

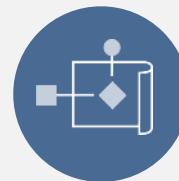
FlexiCapture speeds up your administrative and customer-facing processes and enables you to focus on your core business

# ABBYY FlexiCapture - Key usage scenarios

ABBYY®



TAX/VAT, Blanks,  
Education forms  
processing



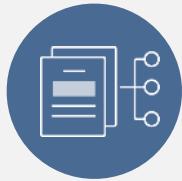
Invoice  
processing  
for Accounts  
Payable



Account  
Receivable &  
Order  
Processing



Proof  
of delivery



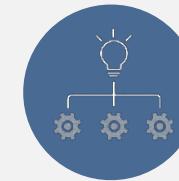
Documents Sets  
for Mortgage &  
Insurance



Unstructured  
documents/  
Contract  
processing



Mailroom  
Classification



Classification  
for Archiving  
and Records  
management



Back-end systems integration

# ABBYY FlexiCapture - Key features

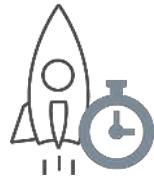
Powered by AI

ABBYY®

- Image Enhancement
- Digitally-born documents
- Advanced Classification
- Recognition – OCR/ICR
- OOTB Invoice, PO, Receipt processing based on NN
- ML for fields extraction at user side
- NLP for unstructured documents processing
- Autolearning

## Delivery Options:

- Enterprise-grade platform
- SaaS
- RESTful service



Faster, straight-through processing



Smooth transactions, smart decisions, and rapid action



Control, predictability, and compliance



## ABBYY FLEXICAPTURE

---

### SYSTEM ARCHITECTURE



## Solution architecture

- Local installation
- Distributed installation

## Components

## Fault Tolerance. Cluster Support

## Servers:

- Application Server
- DB Server and File storage
- Protection Server
- Processing server

## Workstations

- Project Setup Station
- Scanning Station
- Data Verification station
- Verification station

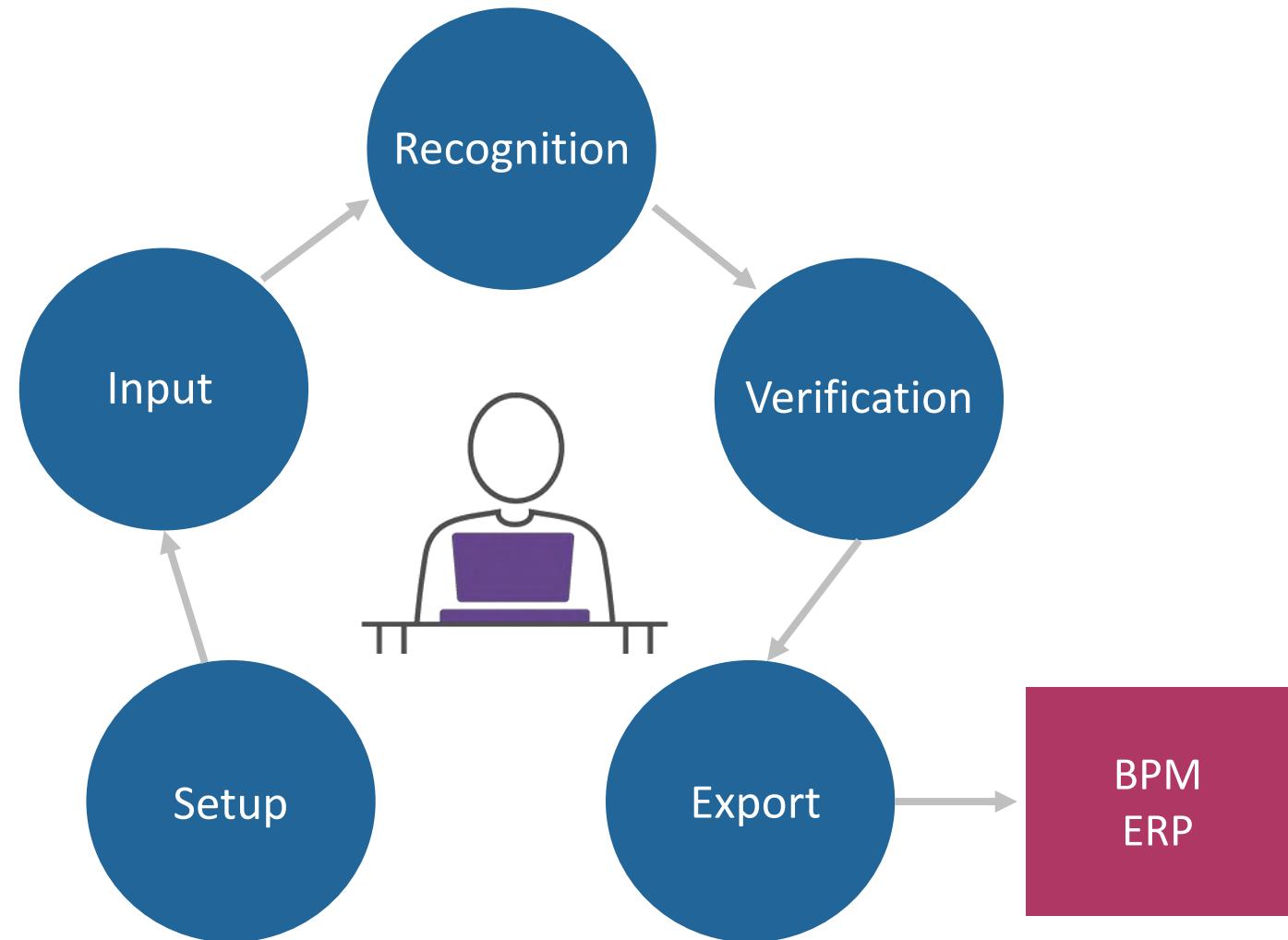


**ABBYY FlexiCapture** is powerful data capture and document processing solutions designed to efficiently transform streams of documents of any structure and complexity into business-ready data

# ABBYY FlexiCapture Standalone

ABBYY®

- Small businesses
- Up to 50 K pages per month
- 1 user involved in processing



# ABBYY FlexiCapture Distributed

ABBYY®



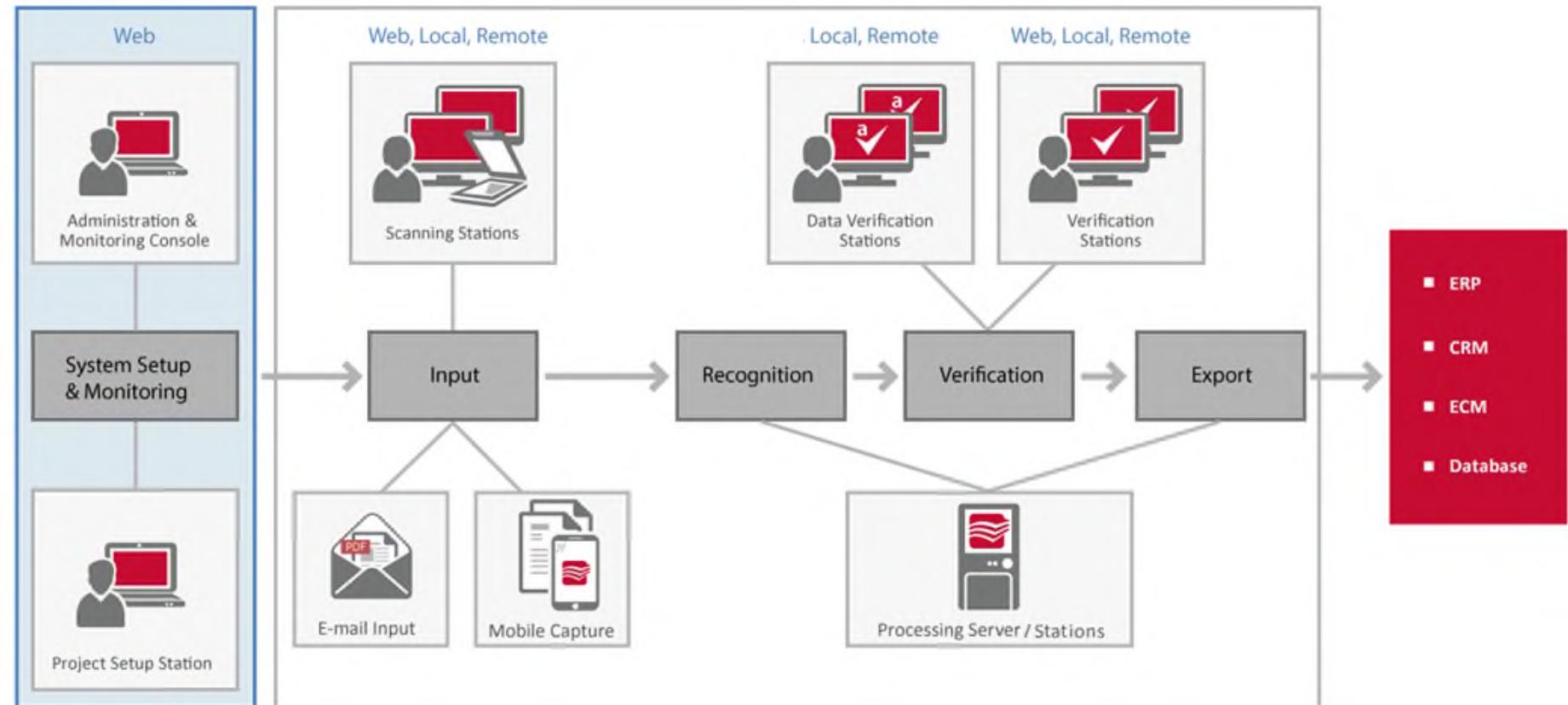
Distribution of duties



Large  
projects and enterprises



More than  
50 K pages per month



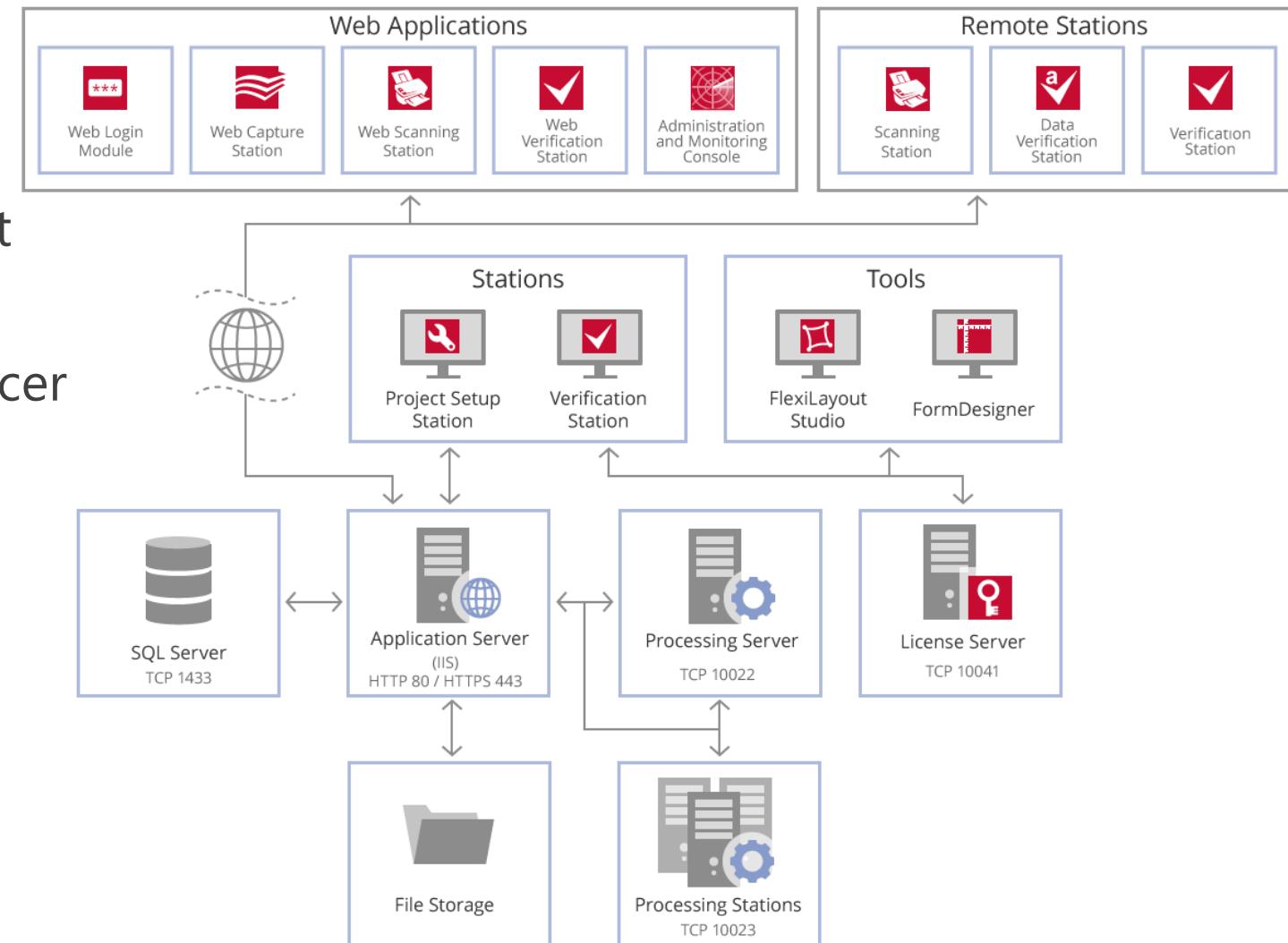
# ABBYY FLEXICAPTURE DISTRIBUTED COMPONENTS

---



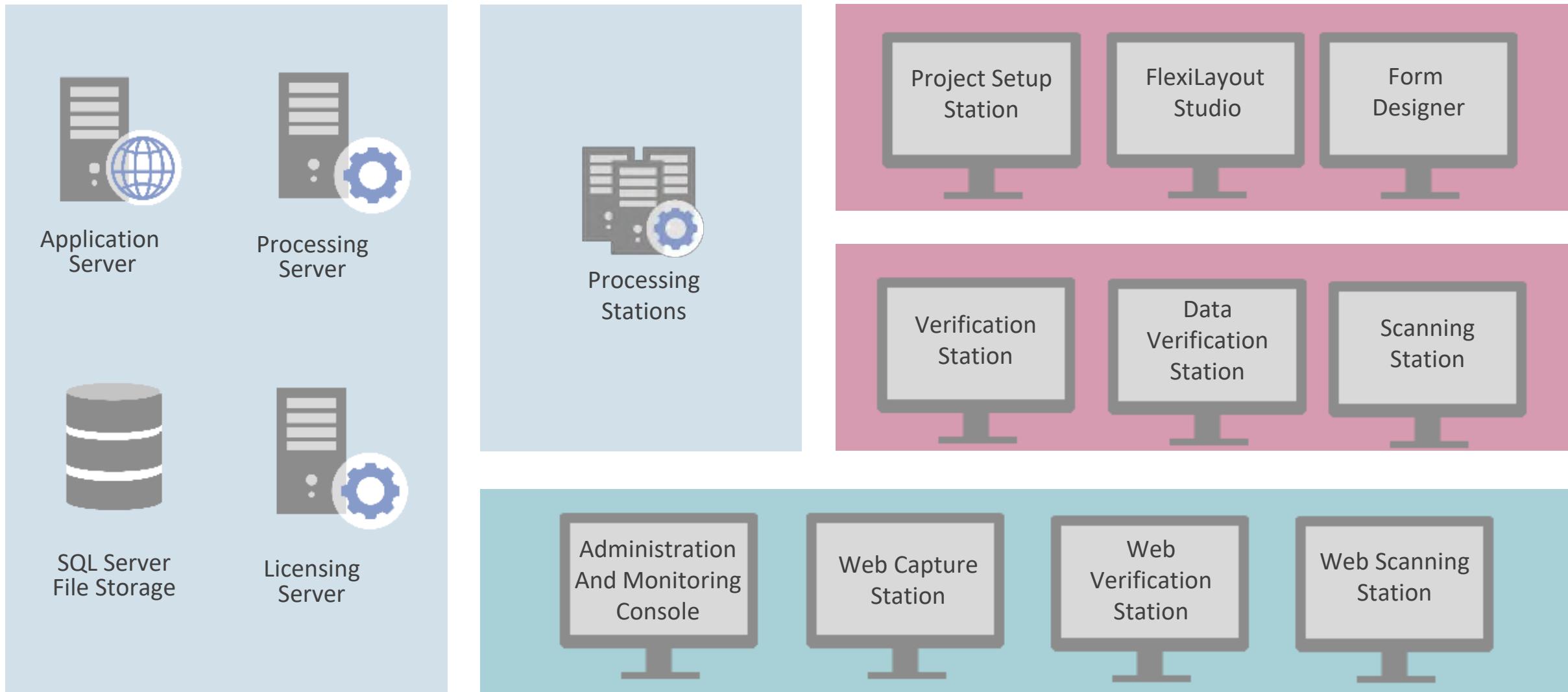
# Component Interaction of ABBYY FlexiCapture Distributed

- **Application Server** is at the heart of the distributed system
- **Processing server** is a load balancer for processing stations
- **License server** monitors license usage and page consumption



# ABBYY FlexiCapture Distributed. Components

ABBYY



Application Server ensures correct functioning of ABBYY FlexiCapture

Stations communicate among themselves via the Application Server

The DB Server stores internal information

- Operators
- Batches
- Operations
- Task statuses
- Processing
- Captured data

# Database Support

ABBYY®

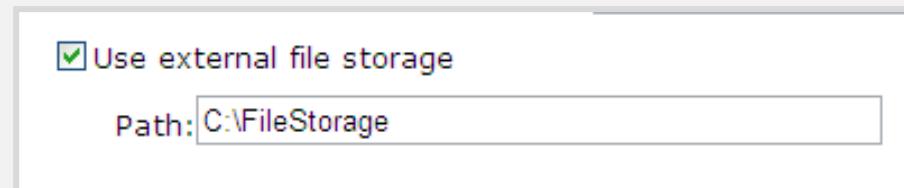


- 2012R2 +



- Is used to set up the entire complex and monitor its operation
- Can be opened from a computer with the Application Server installed
- Can be accessed from any computer using  
<http://ServerName/flexicapture12/monitoring>

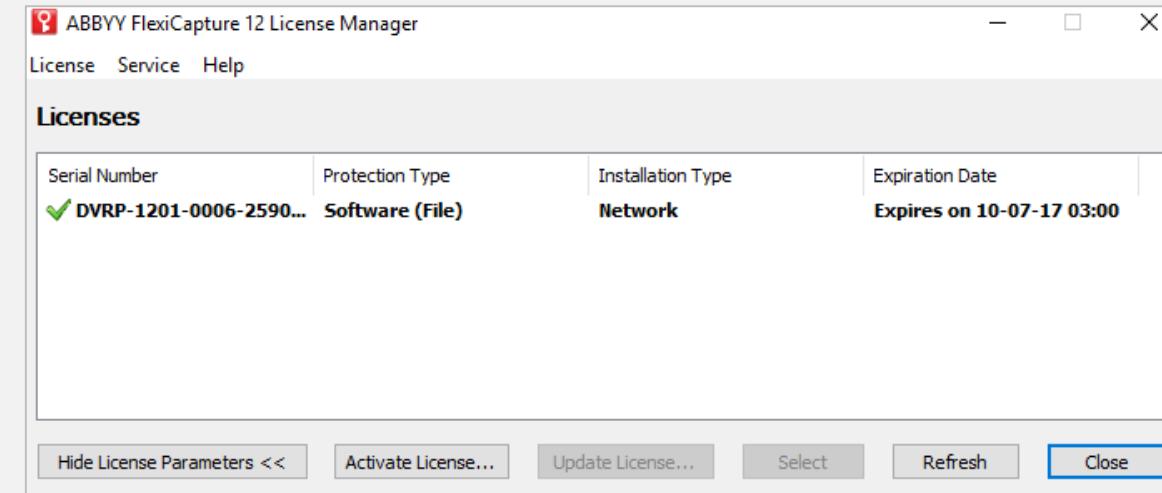
- Storage variants:
  - In a file system (dedicated external storage) – *Recommended!*
  - In the Server database
- The external file storage is specified in the DB connection settings
  - Use a local path! (otherwise, the system performance slows down)



ABBYY FlexiCapture License Manager is installed together with the Licensing Server

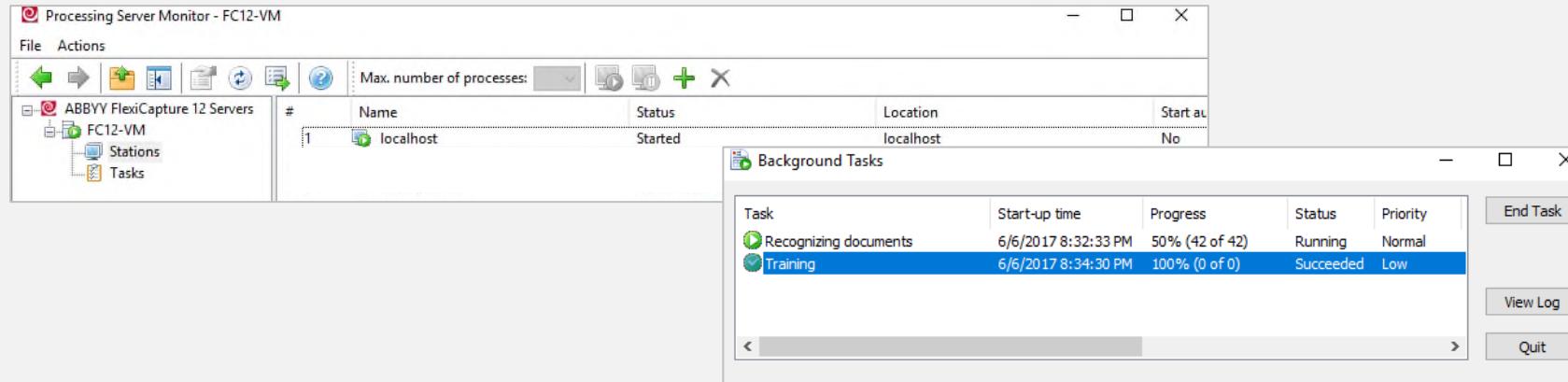
- Protection key  
USB dongle (hardware key) or Serial Number (software key)
- License control

Activating/Deactivating



Processing Server distributes the tasks among the Processing Stations

- Add and start the Processing Server
- Add and start the Processing Stations



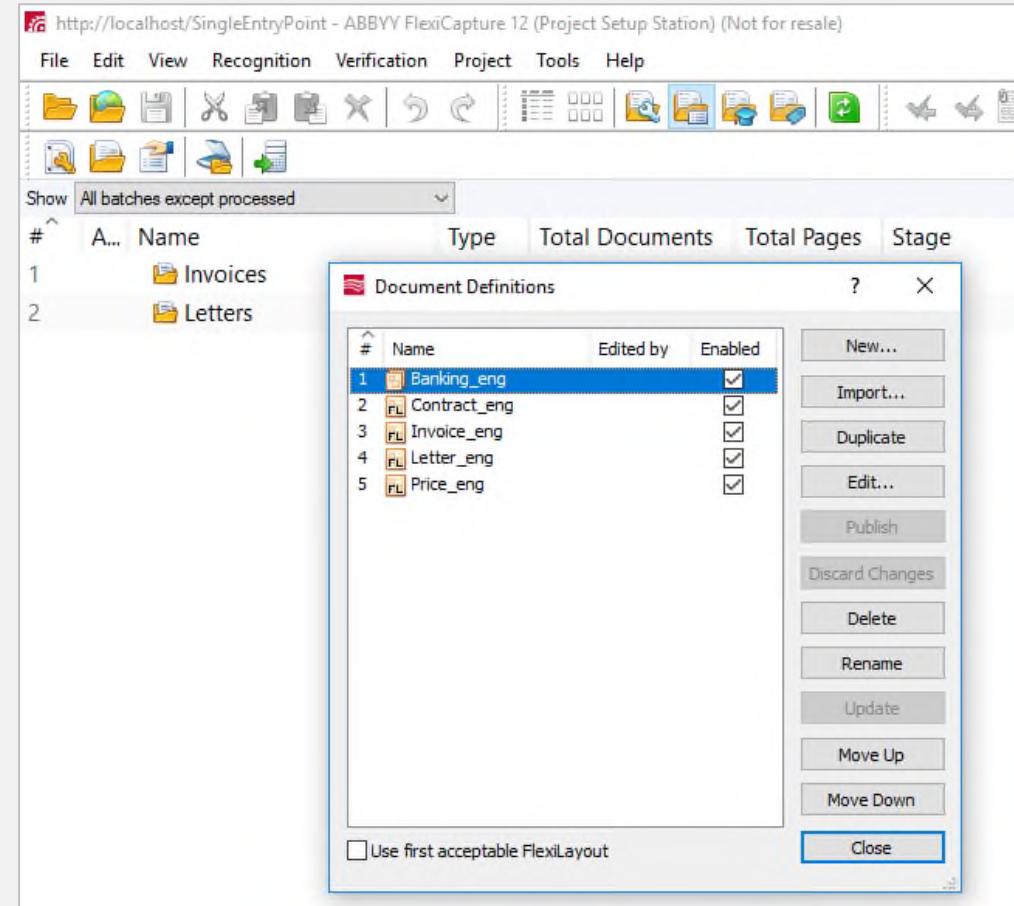
# Project Setup Station

ABBYY®

The Project Setup Station is used to create projects

Project contains

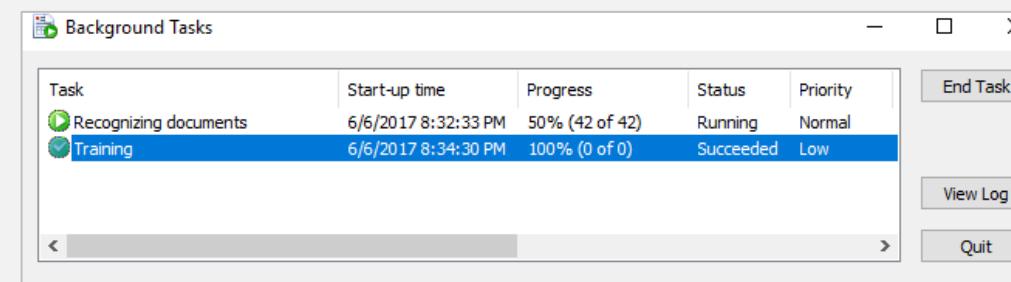
- Document definitions
- Batch types
- Image import settings
- Document export settings
- Workflow settings



## Automatic tasks:

- Loading images from Hot Folder
- Applying Document Definitions and recognition
- Export of data

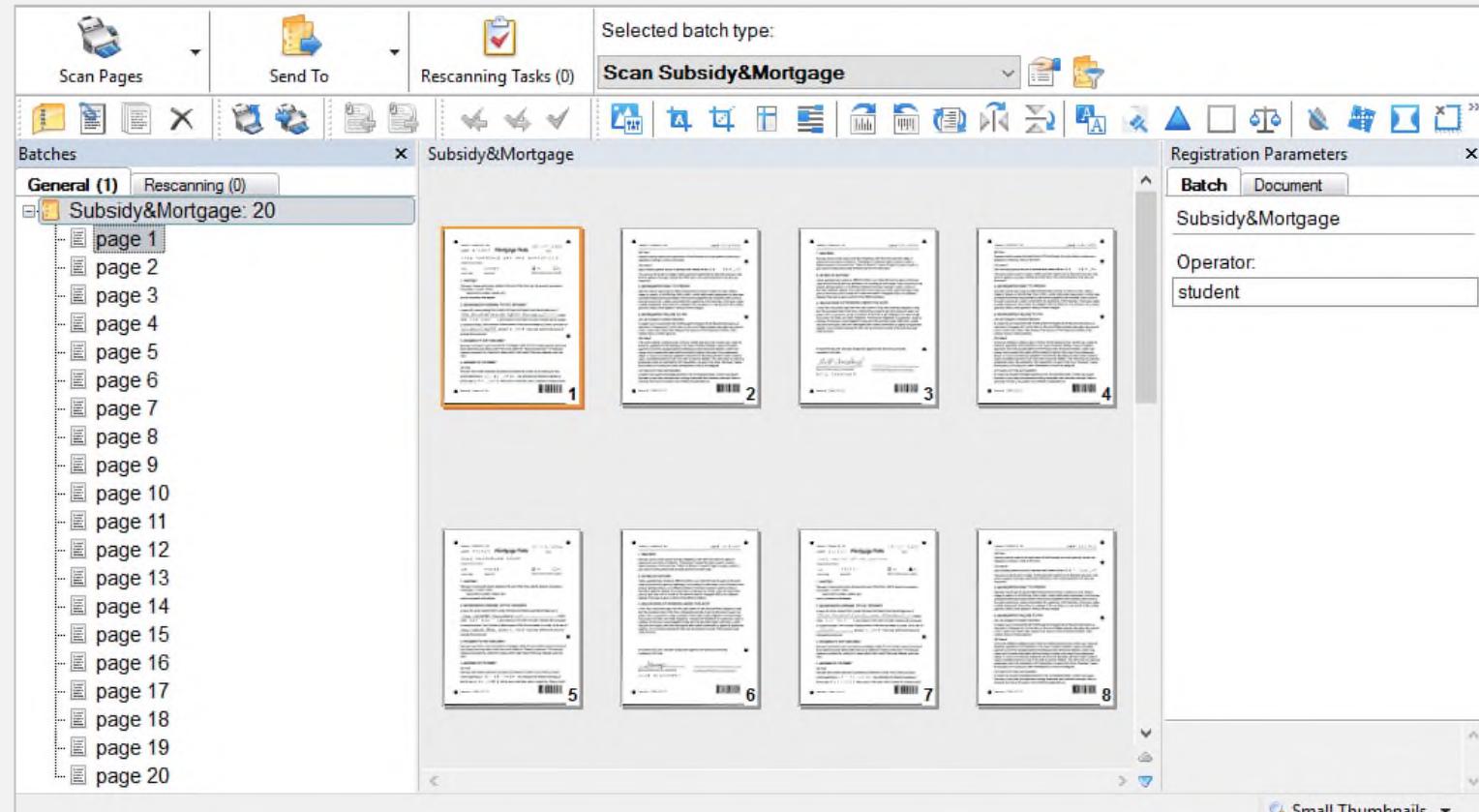
Stations receive task for processing from Processing Server



# Scanning Station

ABBYY®

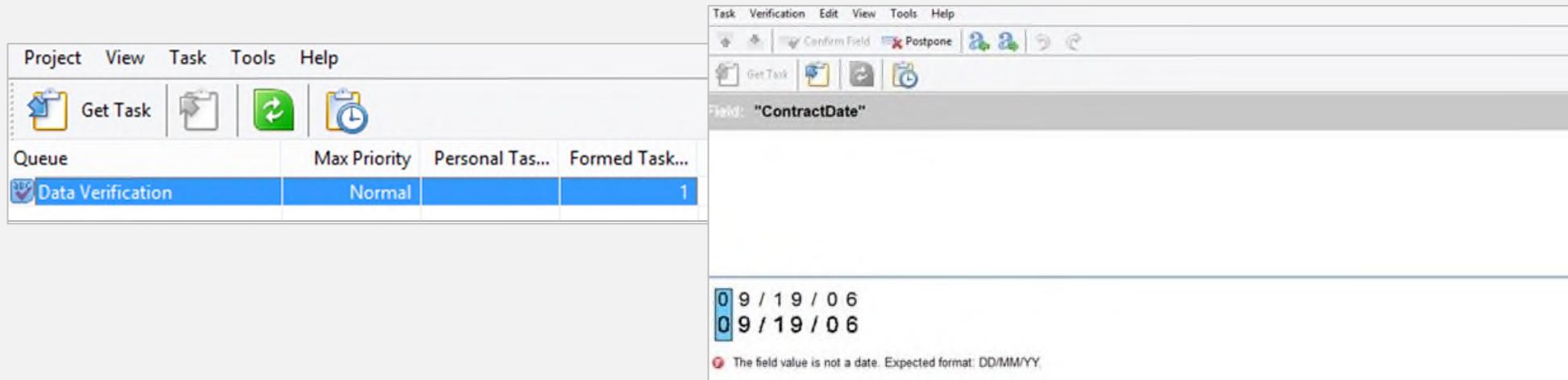
- Scan large amount of documents
- Scan new images and add the already scanned ones
- Send batches to server for automatic recognition
- Edit images



# Data Verification Station

ABBYY®

- The Data Verification Station is used to:
  - verify uncertain characters
  - check data format
- Can be local or remote



# Verification Station

ABBYY®

- Complete check of the documents before export
- Correct assembly errors
- Can be local, remote or Web-based

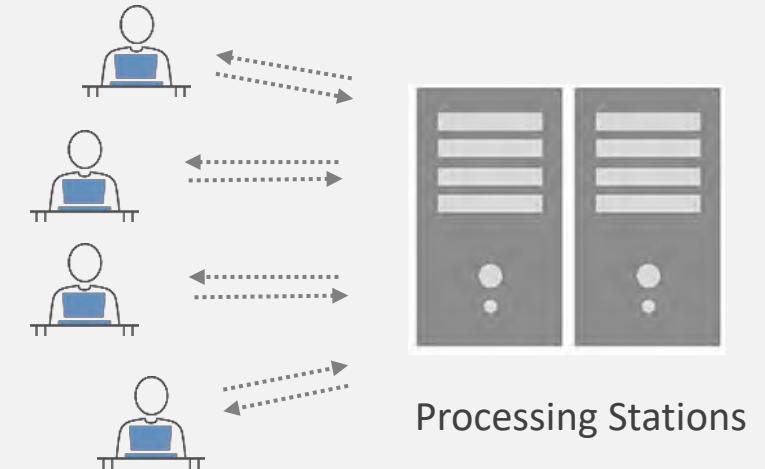
The screenshot shows the ABBYY FlexiCapture 12 (Verification Station) application window. At the top, there's a menu bar with File, Edit, View, Task, Recognition, Verification, Tools, and Help. Below the menu is a toolbar with various icons for file operations like Open, Save, Print, and a magnifying glass for search. The main area has a table titled 'http://localhost/QA\_Report - ABBYY FlexiCapture 12 (Verification Station) - batch 4 (Trial)' showing five tasks:

#	A..	E..	Name	Status	Certainly Reco...	Processing Notes	Sending Note	Previo...
1	📁	QA_Report			96% (51 of 53)			Scanning
2	📁	QA_Report			92% (49 of 53)			Scanning
3	📁	QA_Report			96% (51 of 53)			Scanning
4	📁	QA_Report			96% (51 of 53)			Scanning
5	📁	QA_Report		Verified	100% (53 of 53)			Export

Below the table, a specific task is expanded: '1. QA\_Report x'. This panel contains form fields for OrderNumber (5616553-999/N.2), OrderDate (15.07.15), Variant\_ID (Default), CompanyName (T. van LEI- I D. Heid), and Company\_ID (highlighted with a yellow checkered pattern). A message at the bottom says 'Check Company Name: Data not found in database.'

To the right of the task list is a preview of a scanned document page. The page header reads 'T. van LEFT D. Heid GmbH & Co. KG' and 'SCHRAUBEN- UND MUTTERNFABRIK UMFORMTEILE'. It includes a logo for 'LEFT' with a gear icon. The page also contains text about an inspection certificate: 'Abnahmeprufzeugnis 3.1 nach EN 56789 Inspection Certificate 3.1'. It lists details such as 'Zeugnis-Nr. / certificate No.: 1789555', 'Auftrags-Nr. / order No. + Pos: 5616553-999/N.2', 'Auftragsdatum / order date: 15.07.15', and 'Werks-Nr. / work No. + Pos: 782256.3'. Technical specifications listed include 'Komponenten / components: ISO4055 – G78x21 -10.7', 'Chargen-Nr.: 79012-2', 'Menge / quantity: 5647 Stuck / parts', and 'Herstellungsverfahren / method: Kaltumformung / vergütet cold forging / quenched'.

- Web Scanning Station
- Web Verification Station
- Web Capture Station



Easy and fast access to operator tasks from any place without installation!

# Scaled Redundant Architecture

Find online about redundancy and disaster recovery

[http://help.abbyy.com/en-us/flexicapture/12/admin\\_guide/fail-safety](http://help.abbyy.com/en-us/flexicapture/12/admin_guide/fail-safety)

<https://wiki.abbyy.com/display/DAT/Disaster+Recovery>

<https://www.youtube.com/watch?v=BUTNB4dNgul&feature=youtu.be>

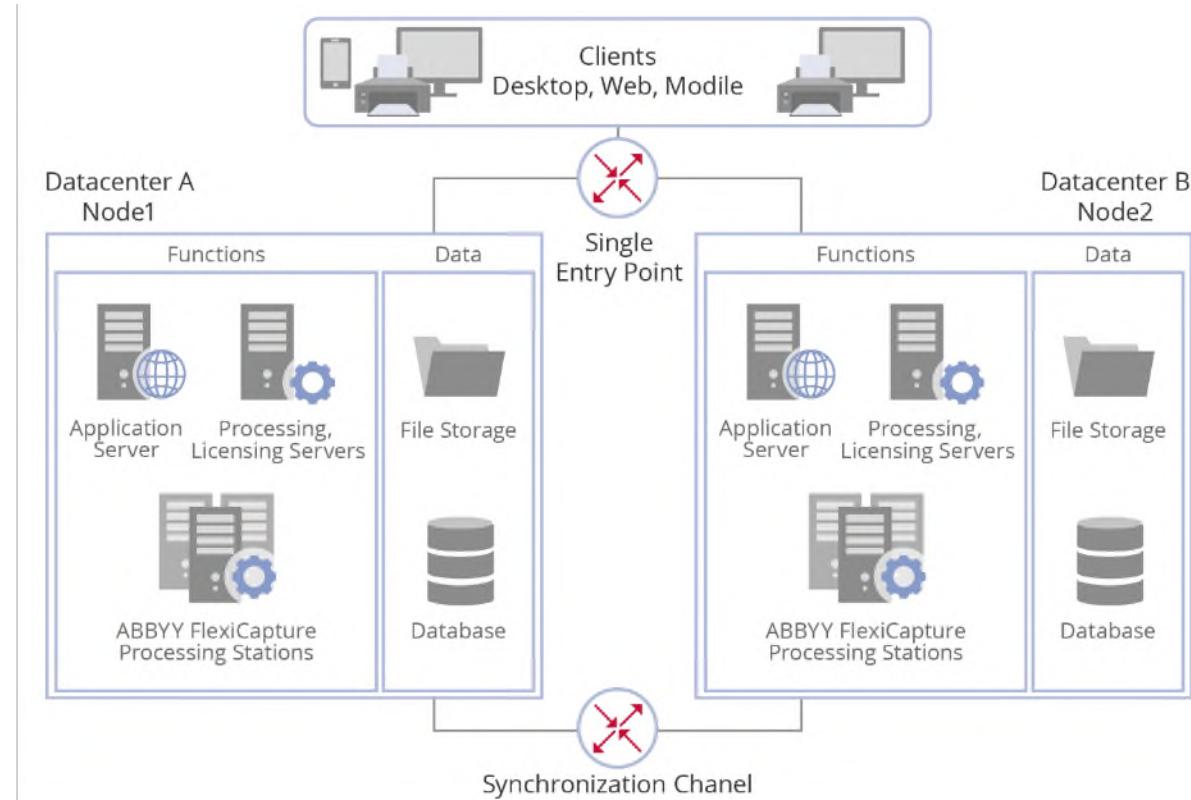
# Disaster Recovery Strategies

## Business Valuable Parameters

Recovery Time Objective (RTO) – How fast can return to operation

Recovery Point Objective (RPO) – How much data ready to lose

Recovery Capacity Objective (RCO) – Backup platform productivity



## Functions recovery

**Hot (Zero RTO, high RCO)**

The same infrastructure on the second node

**Warm (Zero RTO, low RCO)**

Only minimum unit to start processing

**Cold (Zero RCO, high RTO)**

Everything starts from scratch

## Data recovery

**Synchronous replication (Zero RTO, Zero RPO)**

**Asynchronous replication (Zero RTO, Non-zero RPO)**

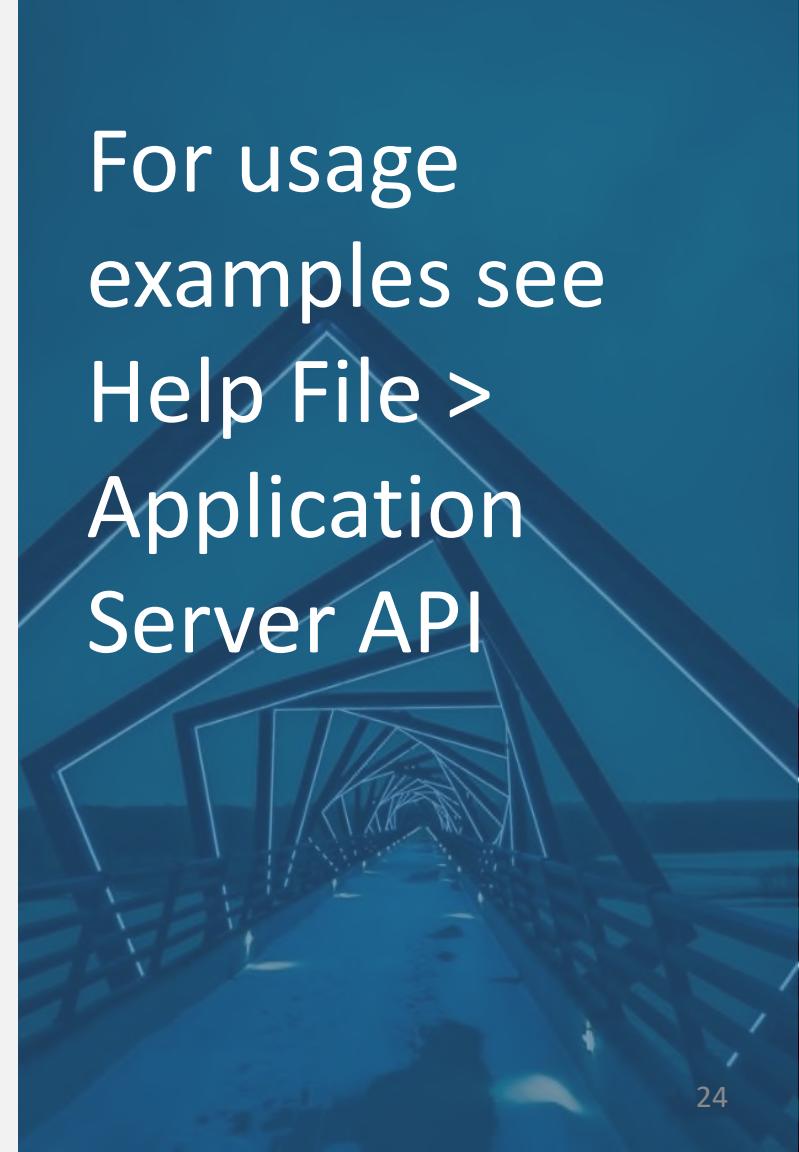
**Recurrent backup (Non-zero RTO, Non-zero RPO)**

The Application Server API allows an external system to:

- Create processing tasks, start processing and get the results
- Control a pre-configured processing

Create your own applications that use ABBYY FlexiCapture as a service

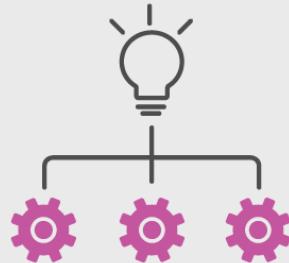
For usage examples see  
Help File >  
Application  
Server API

A dark blue abstract background graphic featuring a perspective view of a series of nested, illuminated hexagonal structures that resemble a tunnel or a complex architectural model.

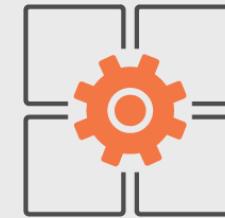
# Summary



ABBYY FlexiCapture provides automatic conversion of various forms and documents (of any structure and complexity)



Scalable architecture allows to select the necessary system configuration



Client\Server architecture ensures fast and effective document processing and data capture

# ABBYY®

## ABBYY FLEXICAPTURE

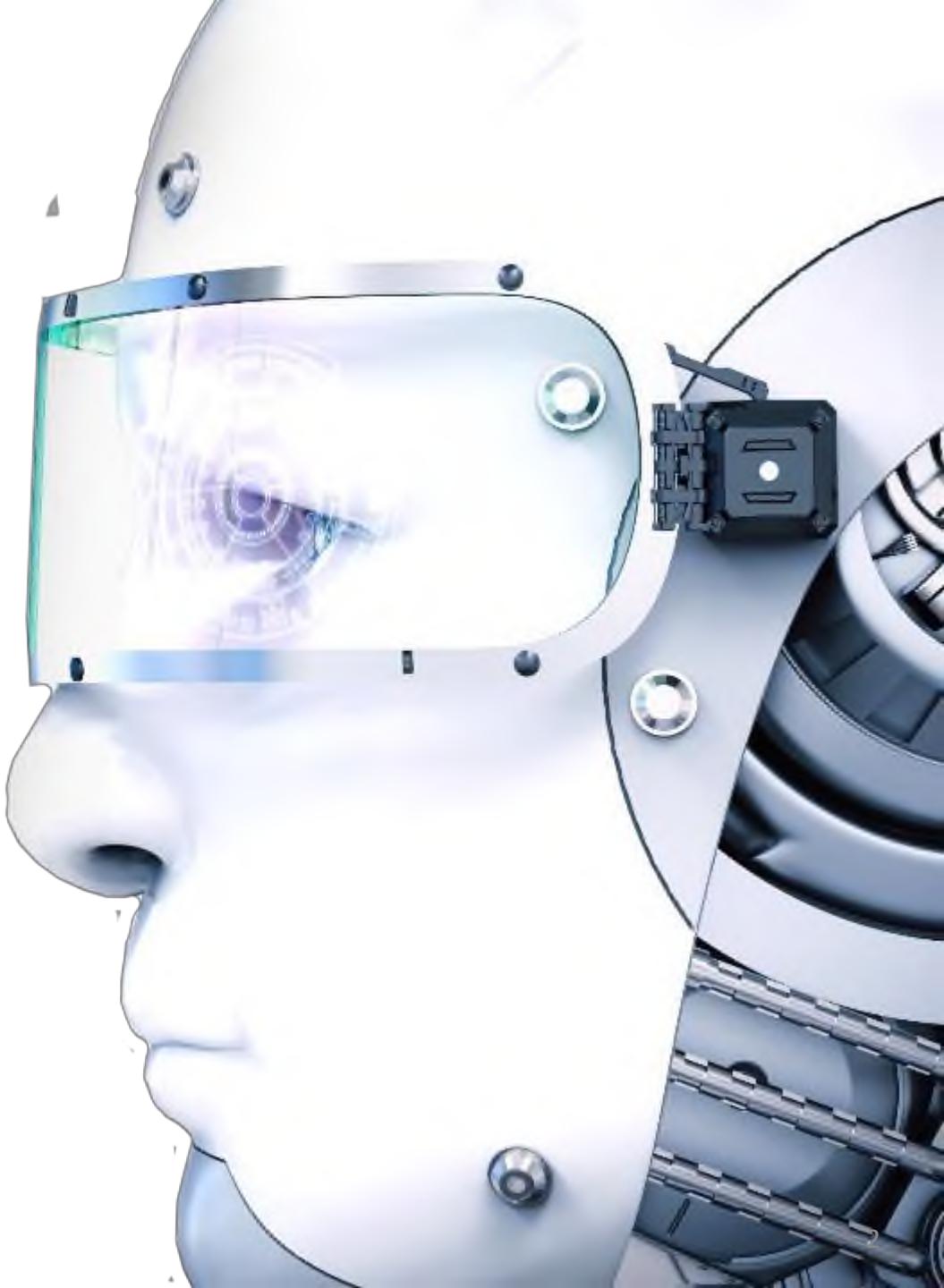
---

PROCESSING SERVER  
AND PROCESSING STATIONS



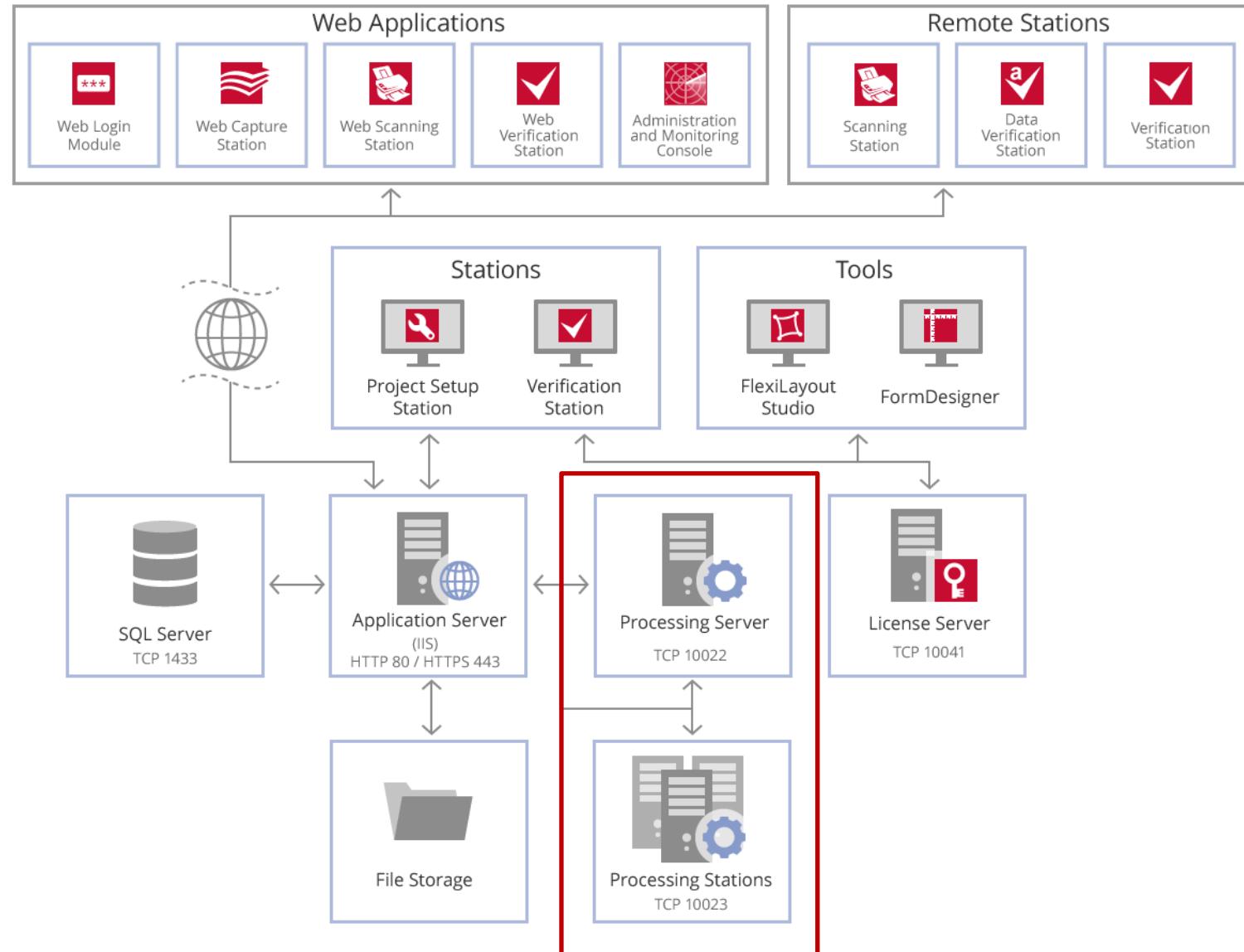
# Overview

- Processing Server and Processing Stations
- Processing Server Monitor
- License Distribution



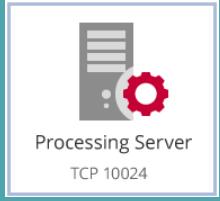
# FlexiCapture 12 Distributed. Architecture

ABBYY®



# Server and Processing Stations

ABBYY®



## PROCESSING SERVER

Distributes tasks among the Processing Stations



## PROCESSING STATIONS

Are used for:

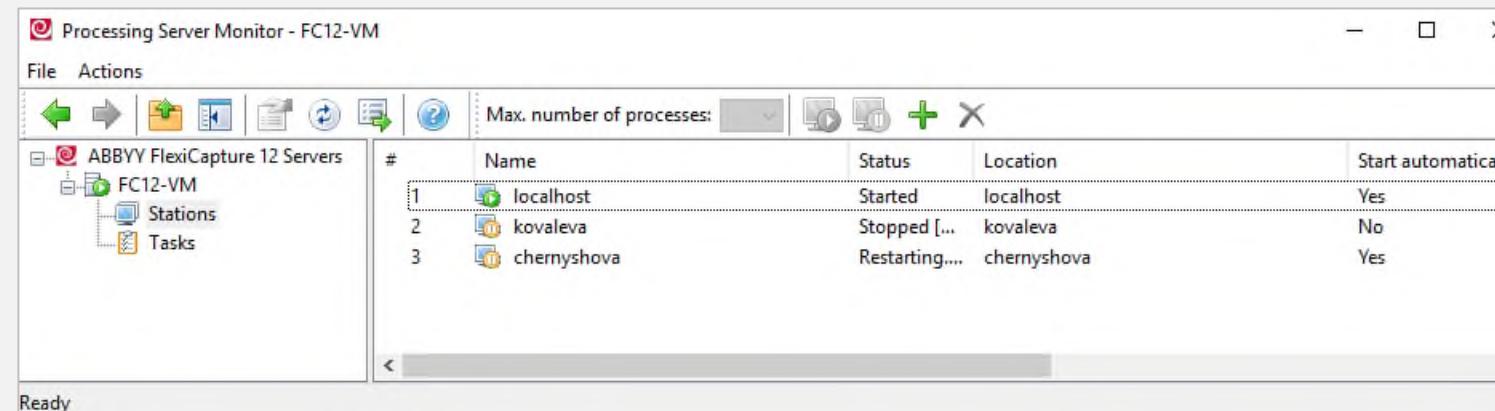
- Recognition
- Automatic importing images from Hot Folders
- Export

# Processing Server Monitor

ABBYY®

The Processing Server Monitor is used for monitoring the server and setting task priorities

- is automatically installed on a computer with the Processing Server



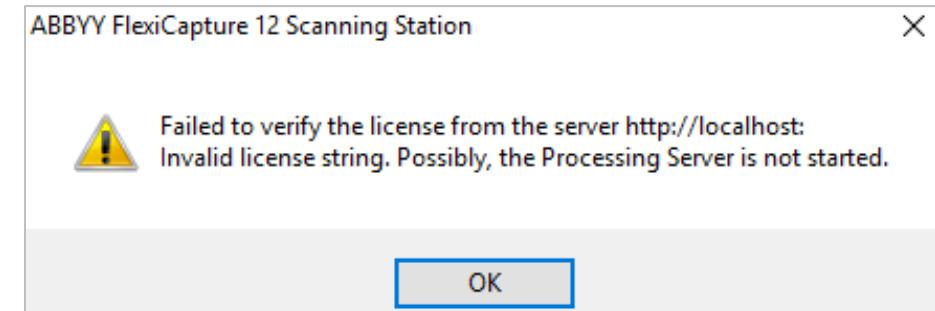
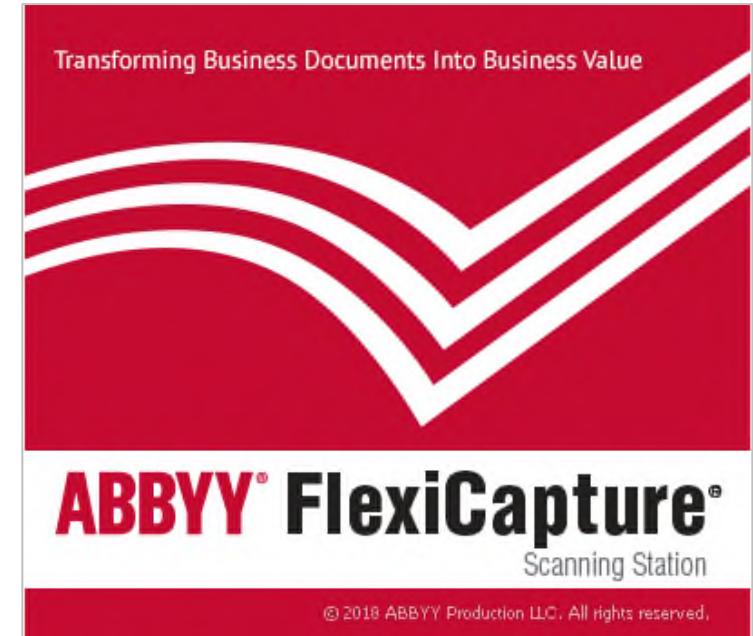
*Note!*

1. Only one Processing Server works in the system
2. Only one server can be added as a Processing Server to a Monitor

# License Distribution

ABBYY

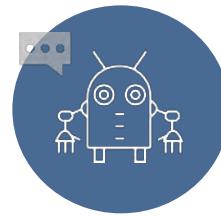
- The Processing Server is responsible for redirecting licenses to the Scanning Station and Data Verification Station
- Processing Server must be running to get a license!



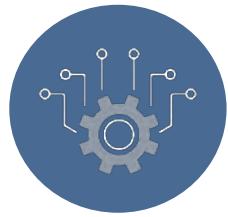
# Summary



The Processing Server  
is used to distribute tasks  
among Processing Stations



Tasks are completed  
automatically, with no operator's  
attention required



Processing Stations are used for:

- Recognition
- Automatic registration  
of images from Hot Folders
- Export



The Processing Server Monitor  
is an application for connection  
of the Processing Server  
with Processing Stations, and for  
task completion monitoring

# ABBYY®

## ABBYY FLEXICAPTURE

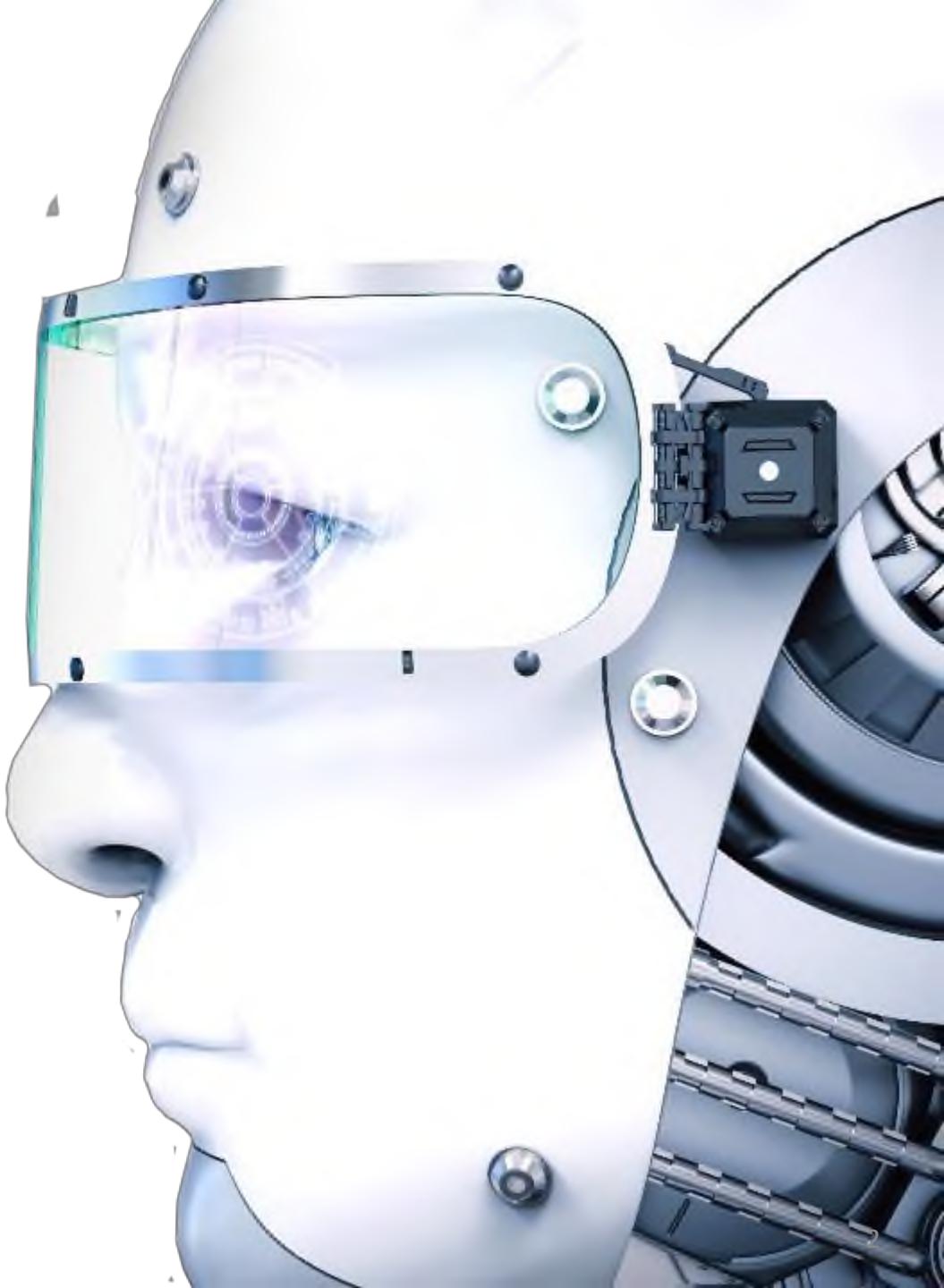
---

### LICENSING



# Overview

- Types of licensing
- License parameters
- Activating/deactivating licenses
- License Manager



FlexiCapture is protected  
with license

Serial number

e.g. **DVRP-1000-0000-1234-1234-1234**

System is not running  
without license

License can be contained in

- USB-dongle
- Software serial number  
(activation required)

## USB dongle

- Licensing key can be transferred to another computer
- Install the iKey driver or CodeMeter Key  
On 32-bit operating systems both drivers are installed automatically



**Serial number:** the license file must be activated on the ABBYY activation server

- License data is stored locally and is tied to the hardware of the machine
- Online license: all of its data is stored on ABBYY's cloud servers

	<b>Software protection</b>	<b>Hardware protection</b>
Use	<ul style="list-style-type: none"><li>For companies where USB dongles may not be used</li></ul>	<ul style="list-style-type: none"><li>Any versions</li></ul>
Activation	<ul style="list-style-type: none"><li>Via the ABBYY activation server (Internet)</li><li>By e-mail</li></ul>	<ul style="list-style-type: none"><li>Via the Internet</li><li>By e-mail</li></ul>
Deactivation	<ul style="list-style-type: none"><li>Via the Internet</li></ul>	<ul style="list-style-type: none"><li>Via the Internet</li></ul>

## Main parameters:

- Description
- License purpose
- Workstations, tools
- Productivity
  - number of pages and CPU cores
- Processing
- Custom components
- Import
- Export

## ABBYY FlexiCapture for Invoices

- custom component \*Invoice
- \*Invoice.AddOn.LineItems component for line items extraction

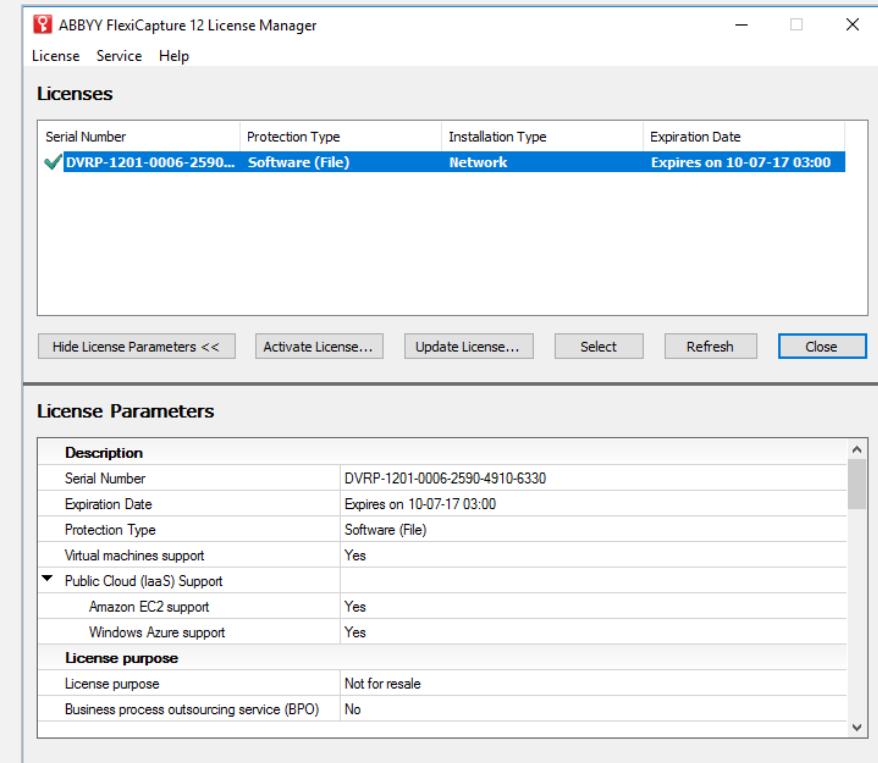
## Multitenant licensing

# License Manager

ABBYY®

License Manager enables you to activate and manage your licenses and view their parameters

Stores the licensing information for the entire system



ABBYY FlexiCapture uses two types of protection

- Serial number
- USB dongle

License restriction

- Number of pages, CPU, stations
- Time limit

License Manager

is used for managing licenses and viewing license parameters



## ABBYY FLEXICAPTURE

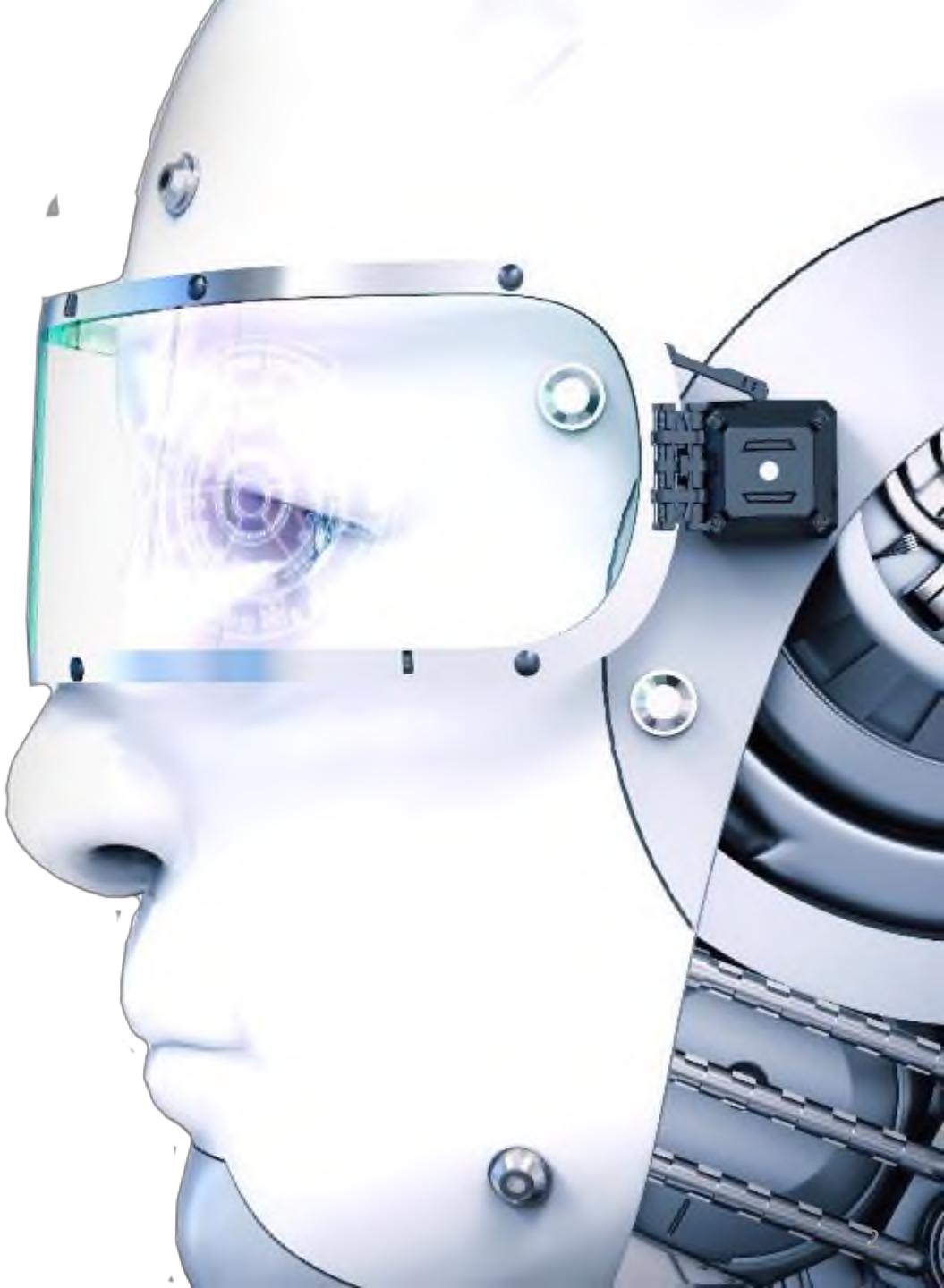
---

ADMINISTRATION  
AND MONITORING



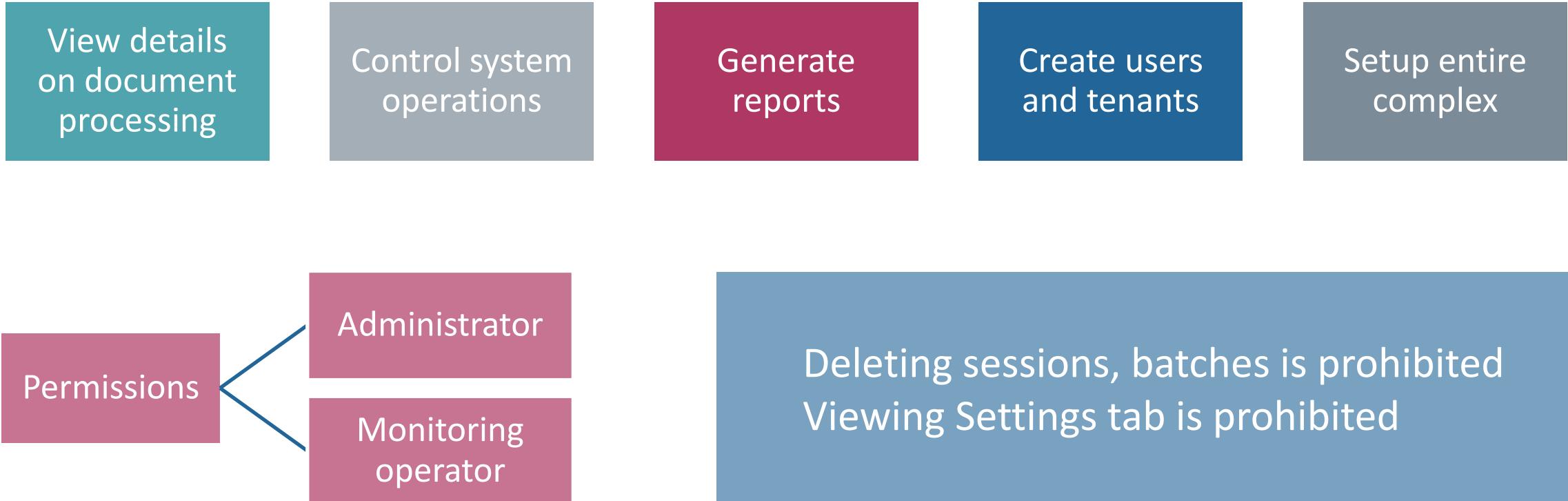
# Overview

- Monitoring
- Logging
- Reports
- System Cleanup



# Administration and Monitoring Console

ABBYY®



# Monitoring Options

ABBYY®

- Projects

- Productivity
- Batches
- Tasks
- Sessions
- Event log
- Errors

- Create filters

- Create tenants

- Create reports

- Custom reports

- Create e-mail notifications

# Automatic Clean Up

ABBYY®

The following objects can be cleaned up automatically

- Exported batches (Project Setup Station)
- Event Log records (Administration and Monitoring Console)
- Report data (Administration and Monitoring Console)

The screenshot shows the 'Settings' tab selected in the top navigation bar. On the left, there's a sidebar with options like Tenants, Projects, Users, Groups, Requests, Application Server, SMTP settings, Email notifications, Event logging mode, Cleanup (which is currently selected), and User agreement. The main content area is titled 'Automatic cleanup' with a green toggle switch indicating it is enabled. It states: 'Expired data is checked and deleted when the server is idle' and 'You can change automatic batch cleanup settings on the Project Setup Station by selecting Batch Type Properties > Export tab'. Below this, there are two sections: 'Event log' (Delete data older than 14 days automatically) and 'Report data' (Delete data older than 180 days automatically). Both sections have checkboxes next to them. A 'SAVE SETTINGS' button is located below these sections. At the bottom, there's a 'Manual cleanup' section with a 'DELETE' button.

Delete out-of-date data from the Server - reduce storage size necessary for keeping the FC projects

## Set the time limit for removal the Event log and Report data

Use *To be deleted* link to get information about number of objects to be deleted

**Manual cleanup**

Delete data older than  days **DELETE**

Processed batches: 4

Event log: 0

Note: Processed batches can be deleted only according to time limits specified in project/batch type properties

- If option is disabled even manual cleanup won't delete batches

# E-mail Notifications

Prevent service interruption, loss of documents or  
delays of processing

Customize messages for all email subscriptions

Subscription for groups of users

Available within a tenant

New user  
registration

Password  
reset

New tenant  
creation

Tenant  
renaming

New  
permission  
request

Exceeding  
license page  
limit

License  
expiring

Error and  
warnings

## Administration and Monitoring Console logs

General system errors that occur during document processing

## Task processing logs on the Processing Servers

Active tasks and their logs can be accessed via the Processing Server Monitor

## OS event logs on the servers

- errors and warnings in the local event log of the operating system
- can be accessed via the standard Event Viewer tool of the MMC console
- list of processing errors even if the Application Server is inaccessible at the time of error

# Downloads: Easy Installation of Scanning Station

ABBYY®

- Scanning Station can be easily installed and run just by clicking a link

Microsoft ClickOnce technology is used

- Installation and upgrade are very smooth

No administrator privileges are required  
are installed per-user, not per-machine

## Available Web-stations

[Web Data Verification Station](#)  
[Administration and Monitoring Station](#)  
[Web Scanning Station](#)  
[Web Scanning Station \(New\)](#)  
[Web Verification Station \(New\)](#)  
[Web Verification Station](#)  
[Web Capture Station](#)

## Requests +

Active    [Processed](#)  
There are no active requests

## Downloads

Scanning Station

[INSTALL](#)

[Hardware and software requirements](#)    [Installation instructions](#)

The Administration and Monitoring Console is used to



configure the Application Server



control the processing



set user permissions



generate the reports



## ABBYY FLEXICAPTURE

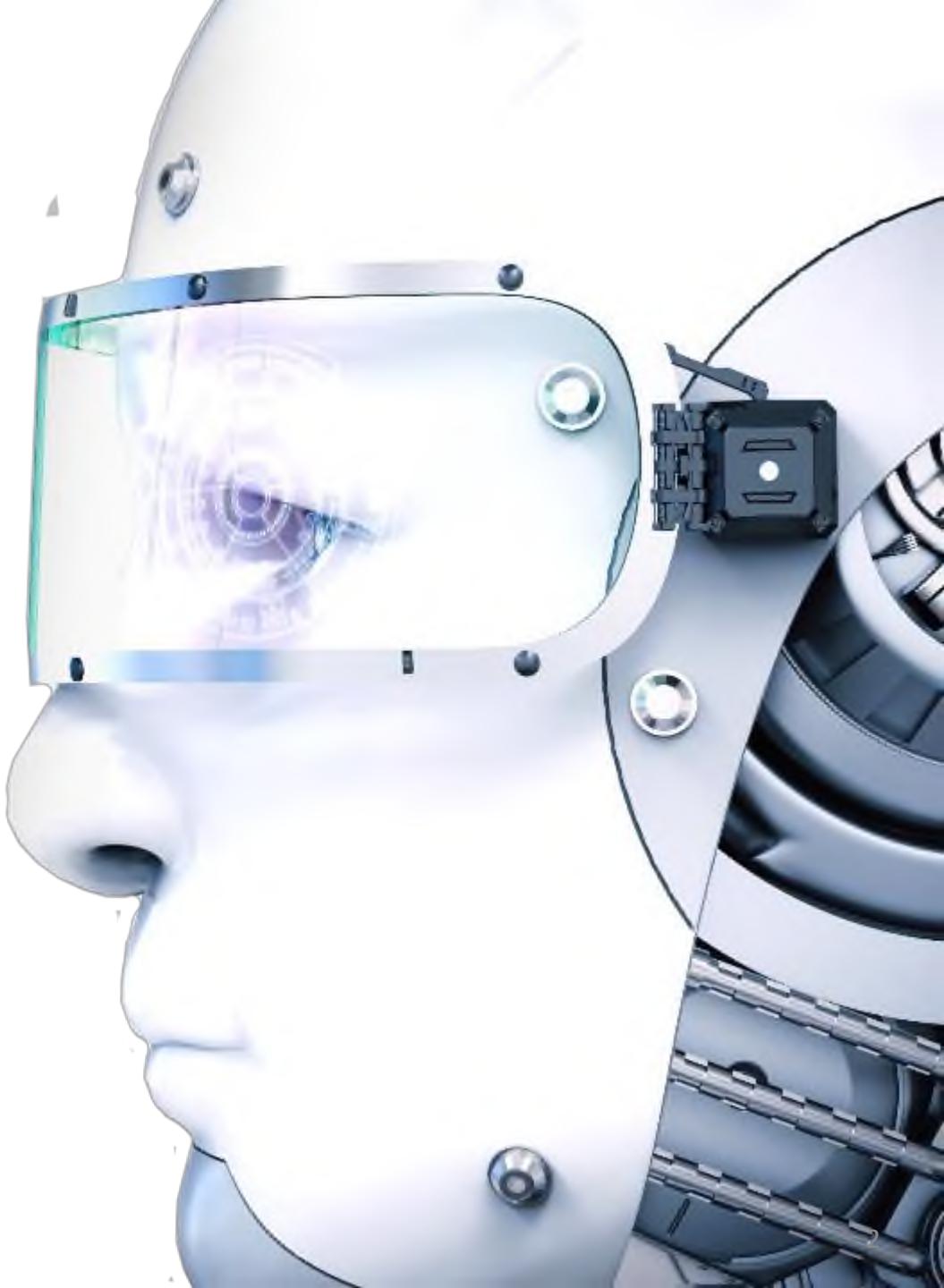
---

DOCUMENT PROCESSING  
STAGES



# Overview

- FlexiCapture Document Overview
- Major processing stages
  - Scan paper documents or load images
  - Recognize
  - Verify
  - Export results



# Document Structure

ABBYY

Banking\_eng



This is a scanned document page showing a bank statement. It includes a header with account information, a table of transactions with columns for date, description, debit, credit, and balance, and a footer with a barcode.

Page 1



This is a scanned document page showing a bank statement. It includes a header with account information, a table of transactions with columns for date, description, debit, credit, and balance, and a footer with a barcode.

Page 2



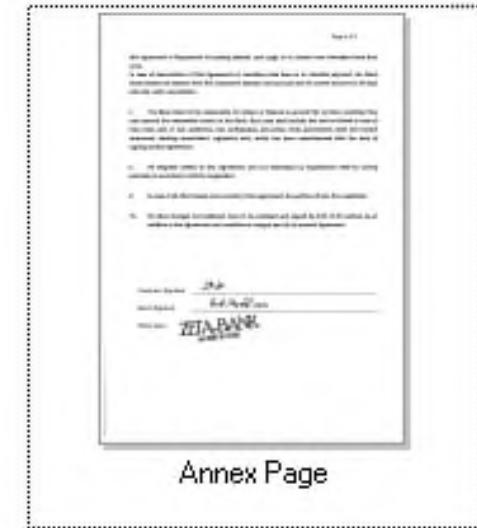
This is a scanned document page showing a bank statement. It includes a header with account information, a table of transactions with columns for date, description, debit, credit, and balance, and a footer with a barcode.

Page 3



This is a scanned document page showing a bank statement. It includes a header with account information, a table of transactions with columns for date, description, debit, credit, and balance, and a footer with a barcode.

Annex Page



This is a scanned document page showing a bank statement. It includes a header with account information, a table of transactions with columns for date, description, debit, credit, and balance, and a footer with a barcode.

Annex Page

# Document Structure

ABBYY®

Document Definition – E.g. Passport

Section 1 – the first page  
with photo

Page 1

Photo

Field 1  
Name

Field 2  
Date of Birth

Number of pages – 1

Identifiers

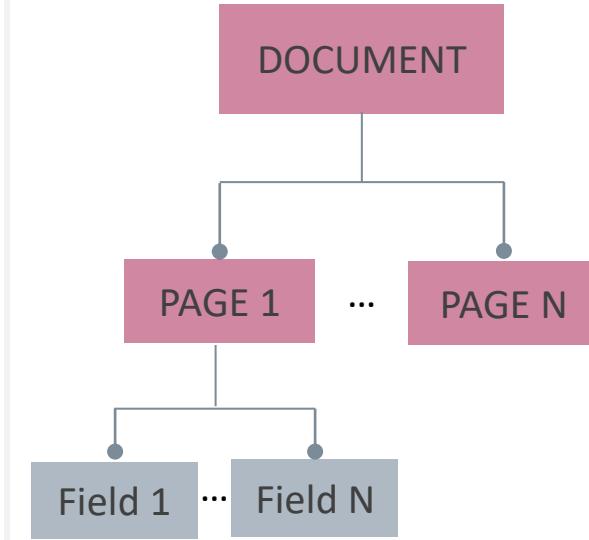
Validation rules

Export settings, etc.

Section 2 – Other pages

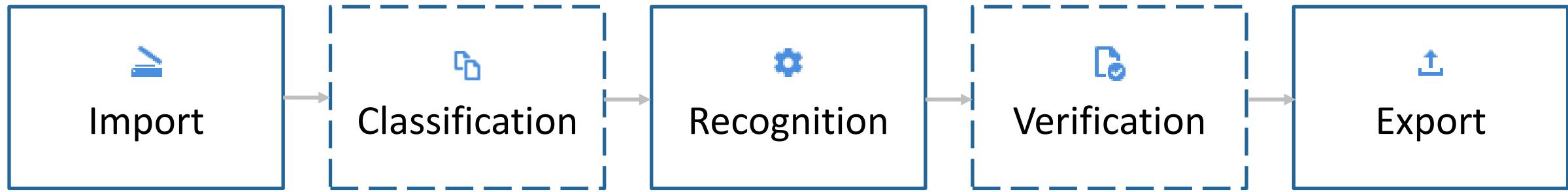
Page 2-N

Number of pages – 0..N

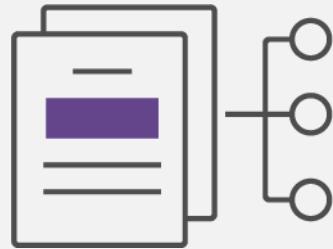


# Document Processing Stages

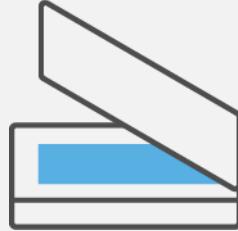
ABBYY®



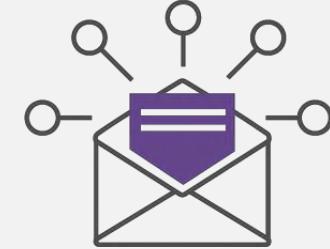
# Import



Create batches with  
documents to be processed



Scan paper documents  
or load image files  
manually on the Scanning  
Station

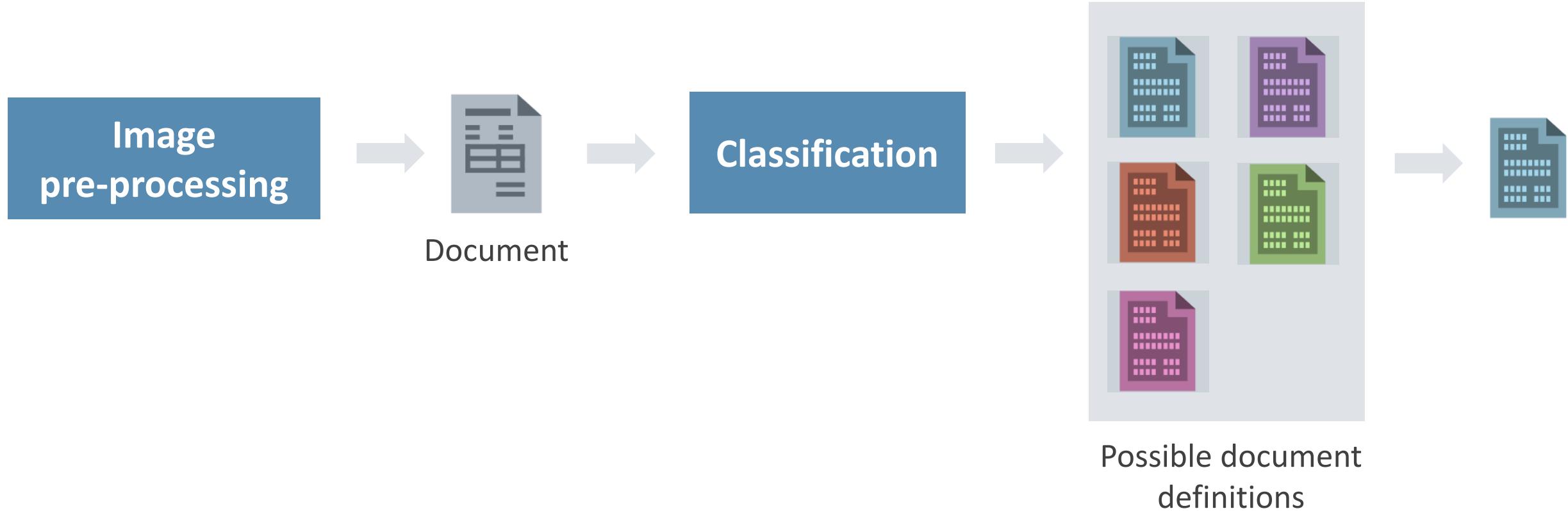


Automatically import  
images from

- pre-defined Hot Folder
- Mailbox
- FTP

# Classification

ABBYY®



1

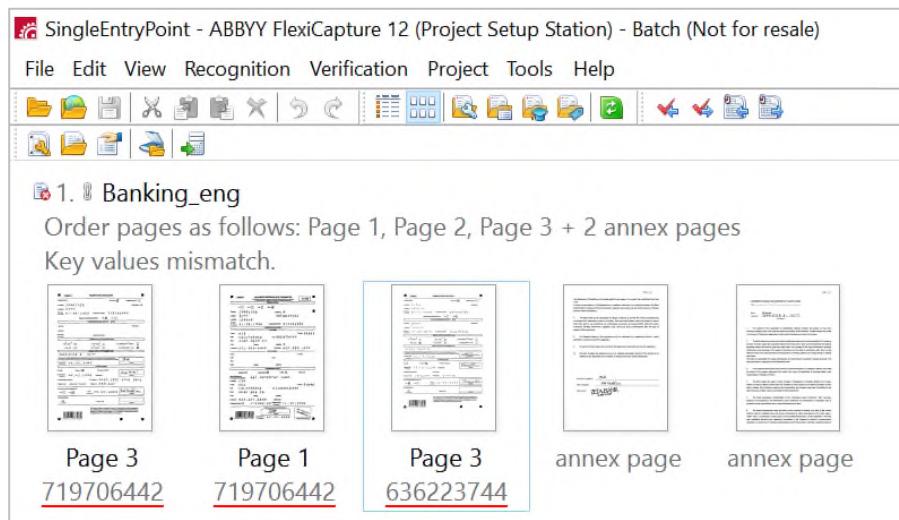
SCANNED IMAGES ARE  
ANALYZED

2

FIELDS ARE RECOGNIZED  
DATA ARE EXTRACTED

3

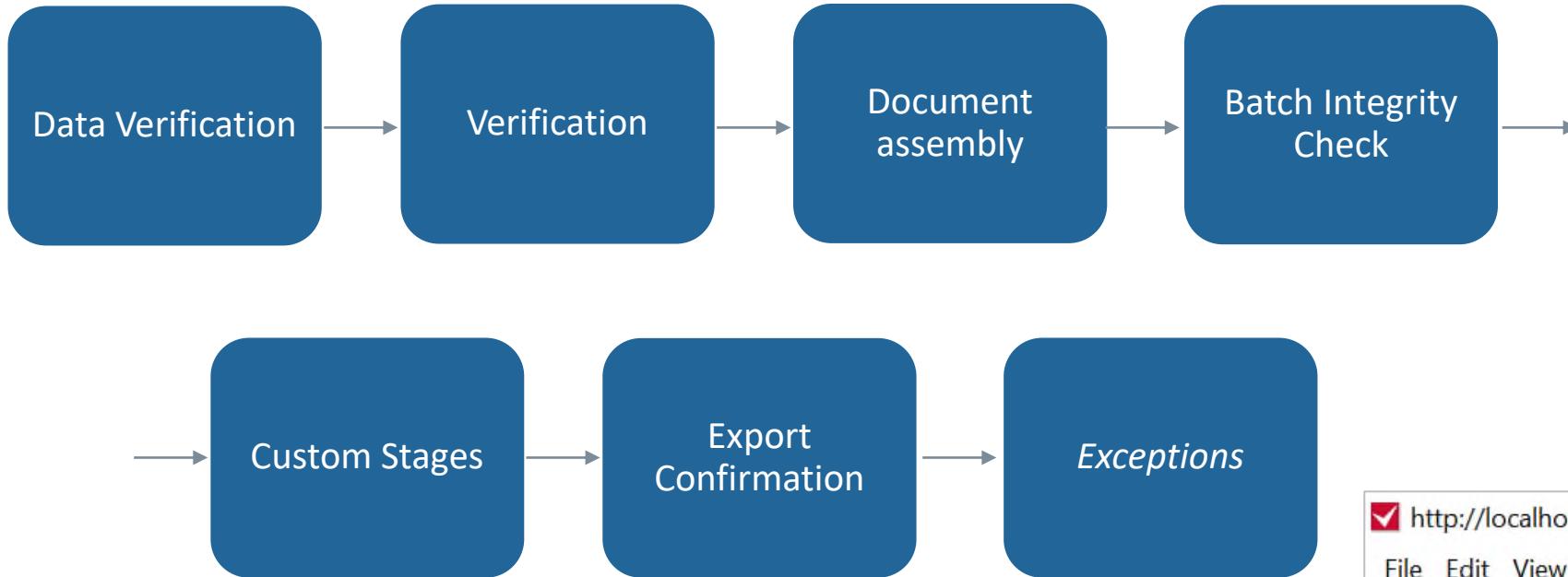
FORMAT CHECKS AND  
VALIDATION RULES ARE  
APPLIED



- Batches are recognized automatically on Processing Station
- No operator involvement required

# Document Verification Stages

ABBYY®



*Note: the order of verification stages can be rearranged!*

http://localhost/SingleEntryPoint - ABBYY FlexiCapture 12 (Verification)						
File Edit View Task Recognition Verification Tools Help						
Get Task Close Task						
Queue		Max Pri...	Personal...	Formed...	Processi...	
<Any>		Normal				
Exceptions		Normal				1
Data Verification		Normal				2
Verification		Normal				2

# Document Verification Roles



Data Verification Operator

Verification Operator

Senior Verification Operator

Custom Roles

# Data Verification & Verification

ABBYY

## Data Verification Station

- Verifying unreliable recognized characters

Field: "Name On Card"

NIHAL YALCIN  
NIHAL YALCIN

NIHAL YALCIN

The Customer/ the Holder agrees that any disputes w/ th the Bank arising from or in connection with this application shall be governed by the General Terms and Conditions, the Service Procedure for Private Customers and the Price Schedule of the Bank.

Document Date 07/10/2007

*[Signature]*

## Verification Station

- Checking the extracted data and document assembly
- Working with exceptions
- Manual routing and prioritization of documents if needed

Nationality BRITISH  
City of Birth LONDON Birth Date 01/06/1966

ADDRESS/CONTACT DETAILS

Address

Address type FIXED ZIP 97633  
Country USA State CALIFORNIA  
City BRENTWOOD Street 3700 MAIN ST

1 / 1

Incorrect page order and/or page set. Go to Document in thumbnail view and correct document page order.  
Addressing  
More than 1 checkmark(s) were selected in the group.

FORM 01  
CUSTOMER'S REFERENCE DATA CONFIRMATION  
Personal Details  
Mr Mrs Ms Mrs  
First Name JENNIFER Middle Initial A  
Last Name PITT Nationality BRITISH  
City of Birth LONDON  
Birth Date 01/06/1966 Social Security Number 719706442  
Address/Contact Details  
Country USA ZIP code 94633  
State CALIFORNIA City BRENTWOOD  
Street 3700 MAIN ST.

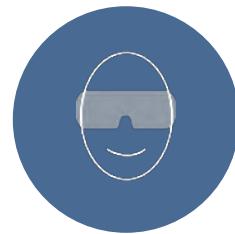
Automated export of batches on the Processing Station based on pre-defined export settings

Batch processing events are logged

- No operator involvement is required

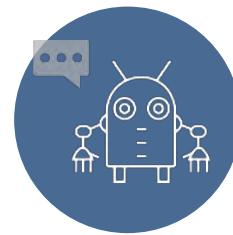
Task	Start-up time	Progress	Status	Priority
Preprocess document with source images	11.02.2015 16:28:09	100% (13 of 13)	Succeeded	Low
Importing images using profile 'HotFolder for ...	11.02.2015 16:12:50	100% (1 of 1)	Succeeded	Low
Importing images using profile 'Hot Folder for...	11.02.2015 16:12:45	100% (1 of 1)	Succeeded	Low
Recognizing pages	11.02.2015 16:28:15	100% (13 of 13)	Succeeded	Normal
Exporting documents	11.02.2015 16:28:26	100% (13 of 13)	Succeeded	High

## Document processing stages:



### Operators' jobs:

- Import (loading) images  
from scanner or from a folder)
- Verification  
correcting errors if required



### Jobs done automatically:

- Import from a Hot Folder
- Recognition
- Export (of data and images)

# ABBYY®

## ABBYY FLEXICAPTURE

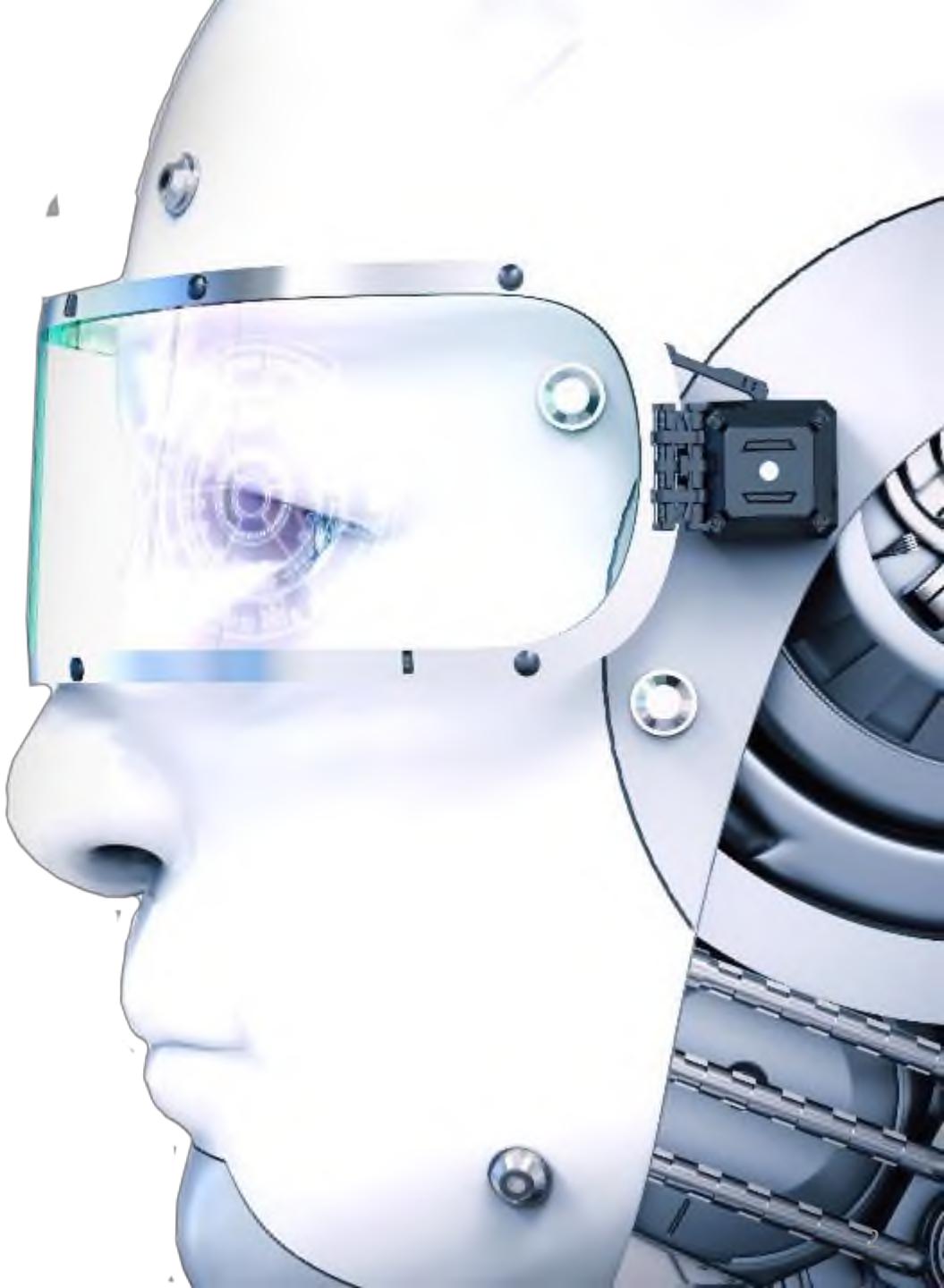
---

### PROJECT SETUP STATION



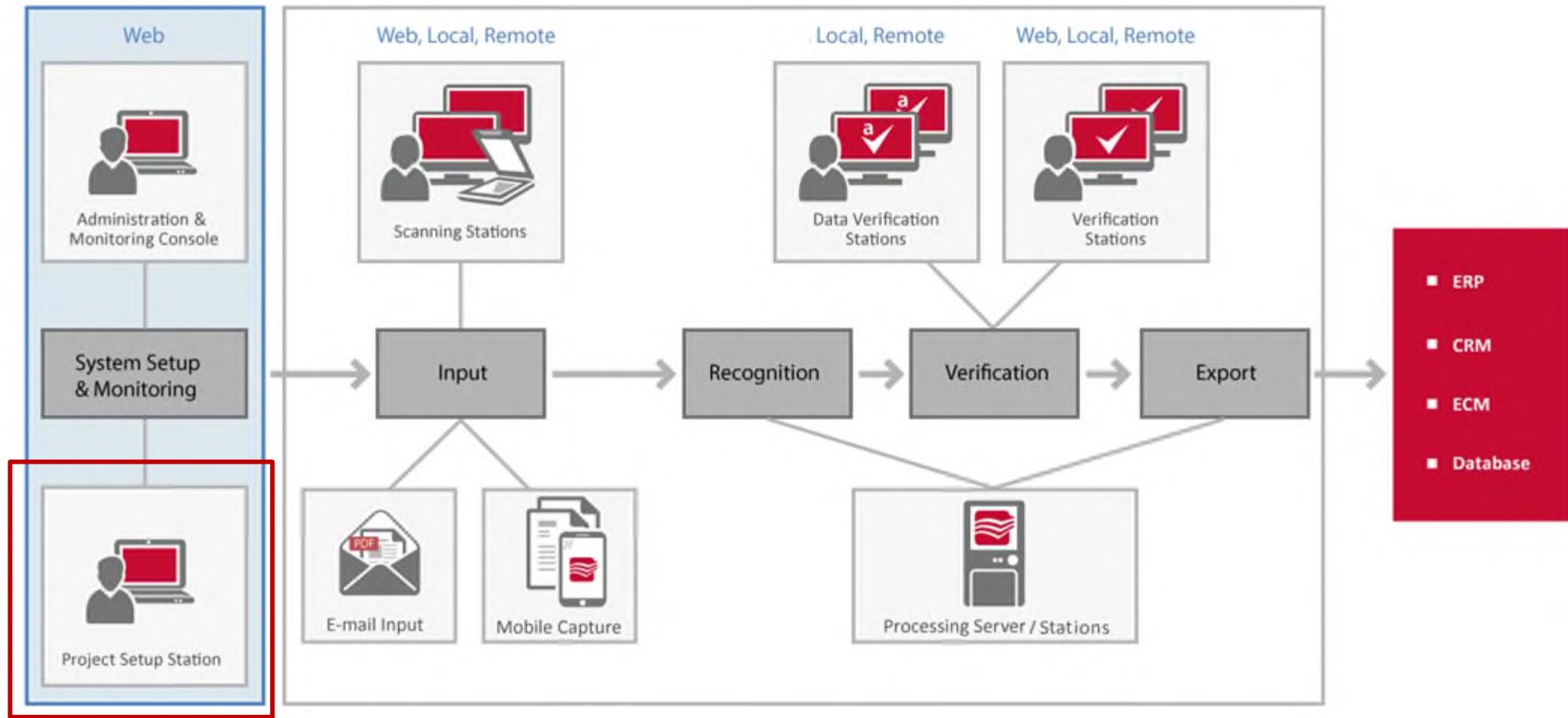
# Overview

- Project setup
- Project structure
  - Project properties
  - Document Definitions
  - Batch types
  - Import profiles
- Working with project
  - Project properties



# Project Setup

ABBYY®

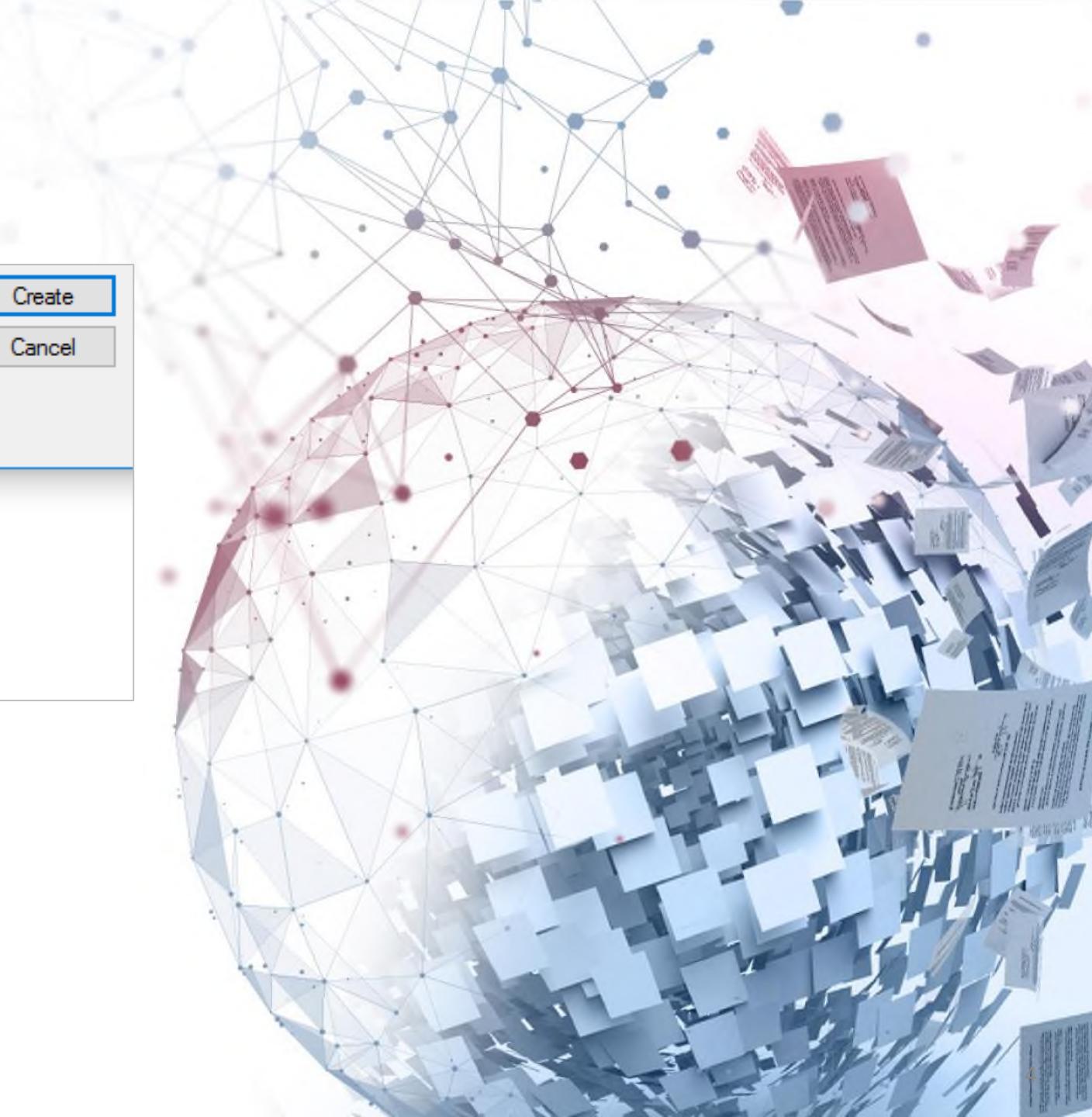


# Project types

File name:

Save as type:

Project type:



A project is a single environment uniting  
the document batches and settings  
required to process them

## Project includes:

- Document Definitions
- Batch Types
- Image Import Profiles
- Project settings
- Work Batches
- Classifier (if used)

# Creating/Editing the Project

## 1. Create a new project



Build version that had been used to create a particular project is saved to file

## 2. Setup project

## 3. Upload project to Server



To edit the existing distributed project open it from the server

## Permissions for project:

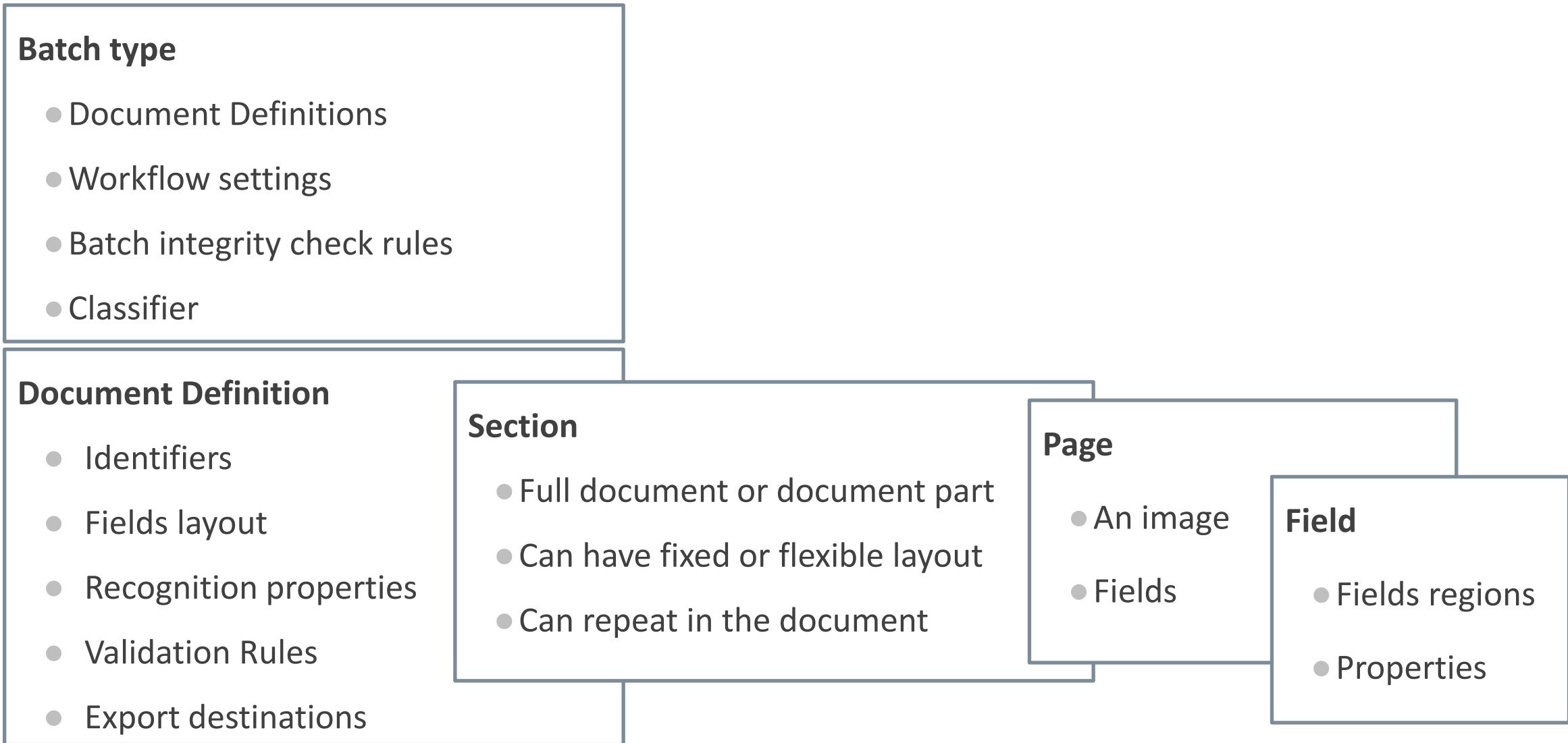
- Only Administrator can create a new project
- Project Settings Editor can adjust the settings of created projects only

Permission	
Roles	Projects
<input type="checkbox"/> Scanning Operator	<input checked="" type="checkbox"/> SingleEntryPoint The role given to all batches types
<input type="checkbox"/> Data Verification Operator	
<input type="checkbox"/> Verification Operator	
<input type="checkbox"/> Senior Verification Operator	
<input type="checkbox"/> Processing Server	
<input checked="" type="checkbox"/> Project Settings Editor	
<input type="checkbox"/> Monitoring Operator	
<input checked="" type="checkbox"/> Administrator	

# PROJECT STRUCTURE

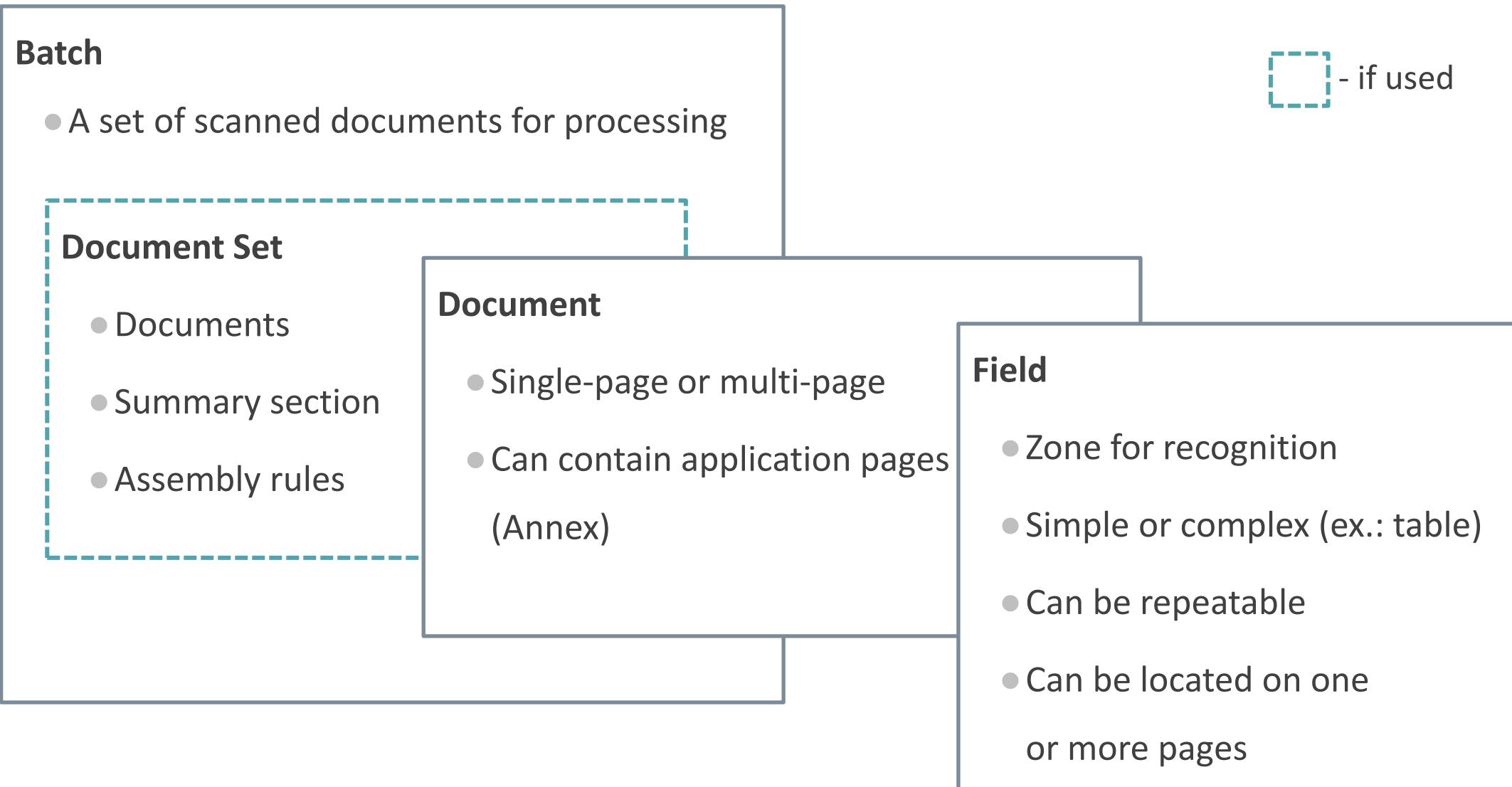
---





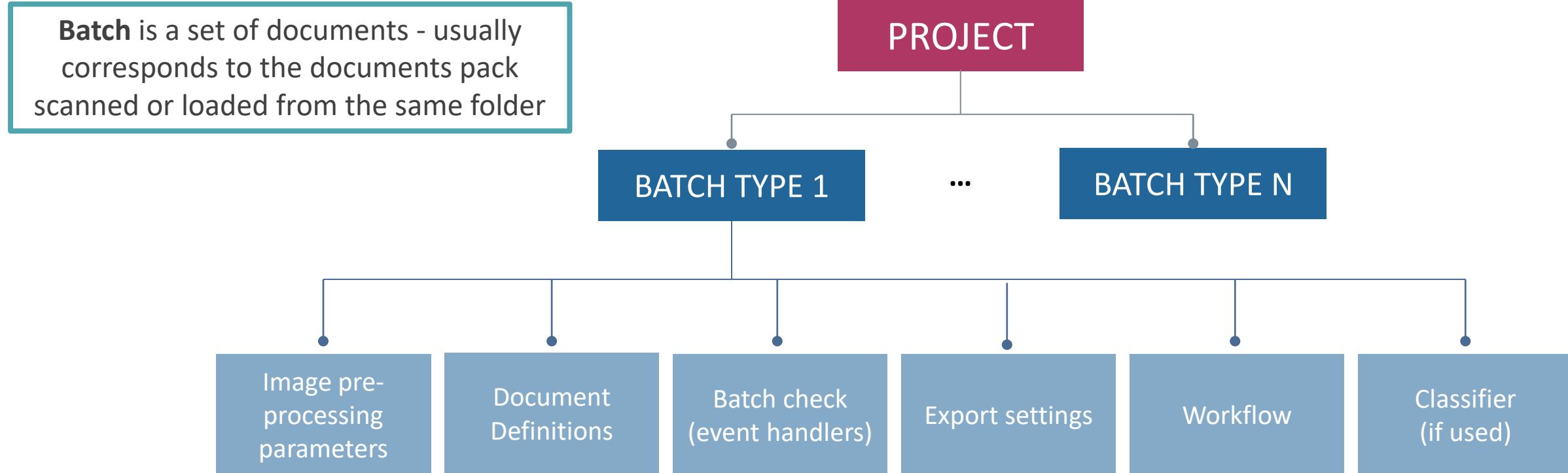
# Document Structure

ABBYY®



# Batch and Batch Type

ABBYY®



 *Batches of “Default” batch type are processed using the Project settings*

# WORKING WITH PROJECT

---





## WORKING BATCHES

Are used for document processing

- published versions of the Document Definitions are used

Working batches list is available at

- Project Setup Station
- Verification Station



## TEST BATCHES

Are used to test Document Definitions

- Published or unpublished versions

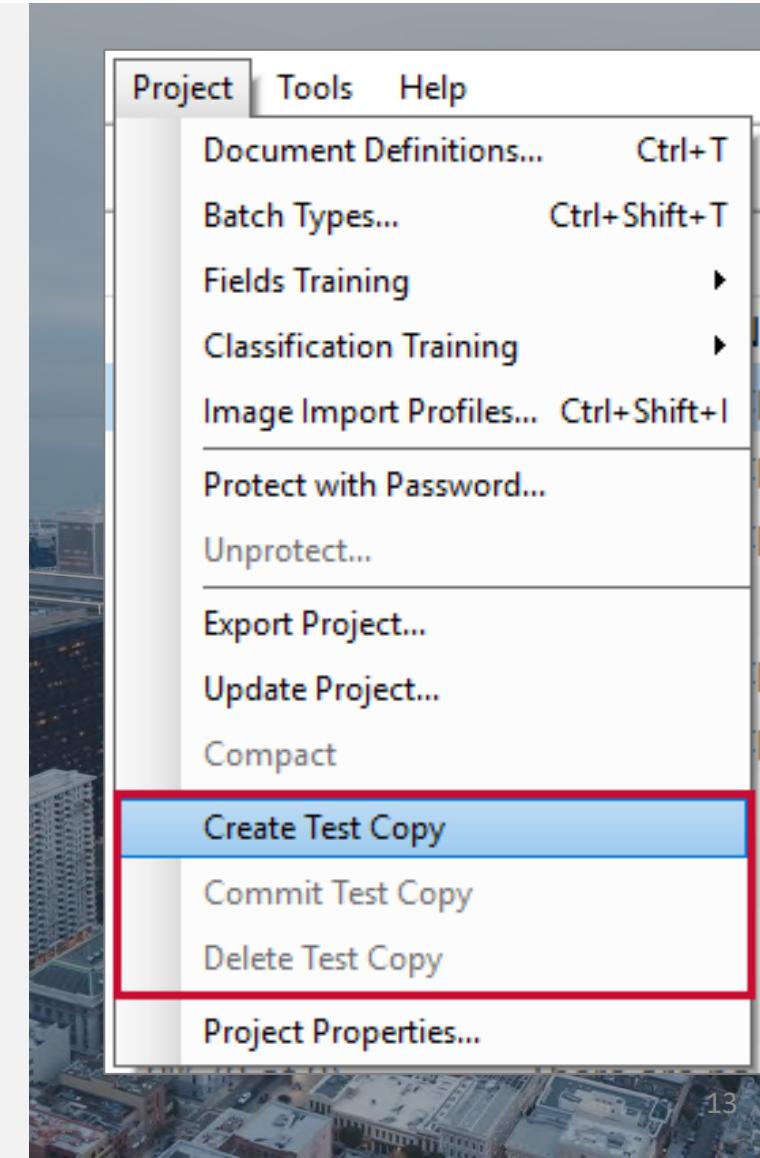
Test recognition quality and validation rules

*Working Batches and Test batches work for both local projects and projects uploaded to the server*

# Project Test Copy

- **Project Test Copy** can be used for testing the distributed processing
- Production Project and Test copy are available in the system simultaneously

Two separate environments with separate sets of batches



# Project Security, Export, Update and Delete

ABBYY®

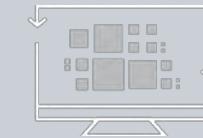
## Password protection

Restriction of access to project settings



## Project export: save all project files

E.g. send copy of the project to your colleague, Technical Support Service



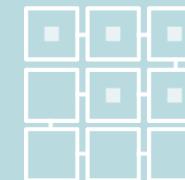
## Project update

Update the project with received changes



## Project deletion

Delete a project from the list of recently opened projects



A photograph showing the interior of a modern building with a large glass wall and a steel truss ceiling structure. Overlaid on the left side of the image is white text.

Project setup  
is carried out  
on the Project  
Setup Station

ABBYY FlexiCapture. Project Setup Station

To set up a project, you must define the following:

- Project properties
- Document Definitions
- Batch types
- Import profiles

Project settings can be tested using:

- Test batches – for local testing of document definitions settings
- Test Copy of project – for testing the distributed processing



## ABBYY FLEXICAPTURE

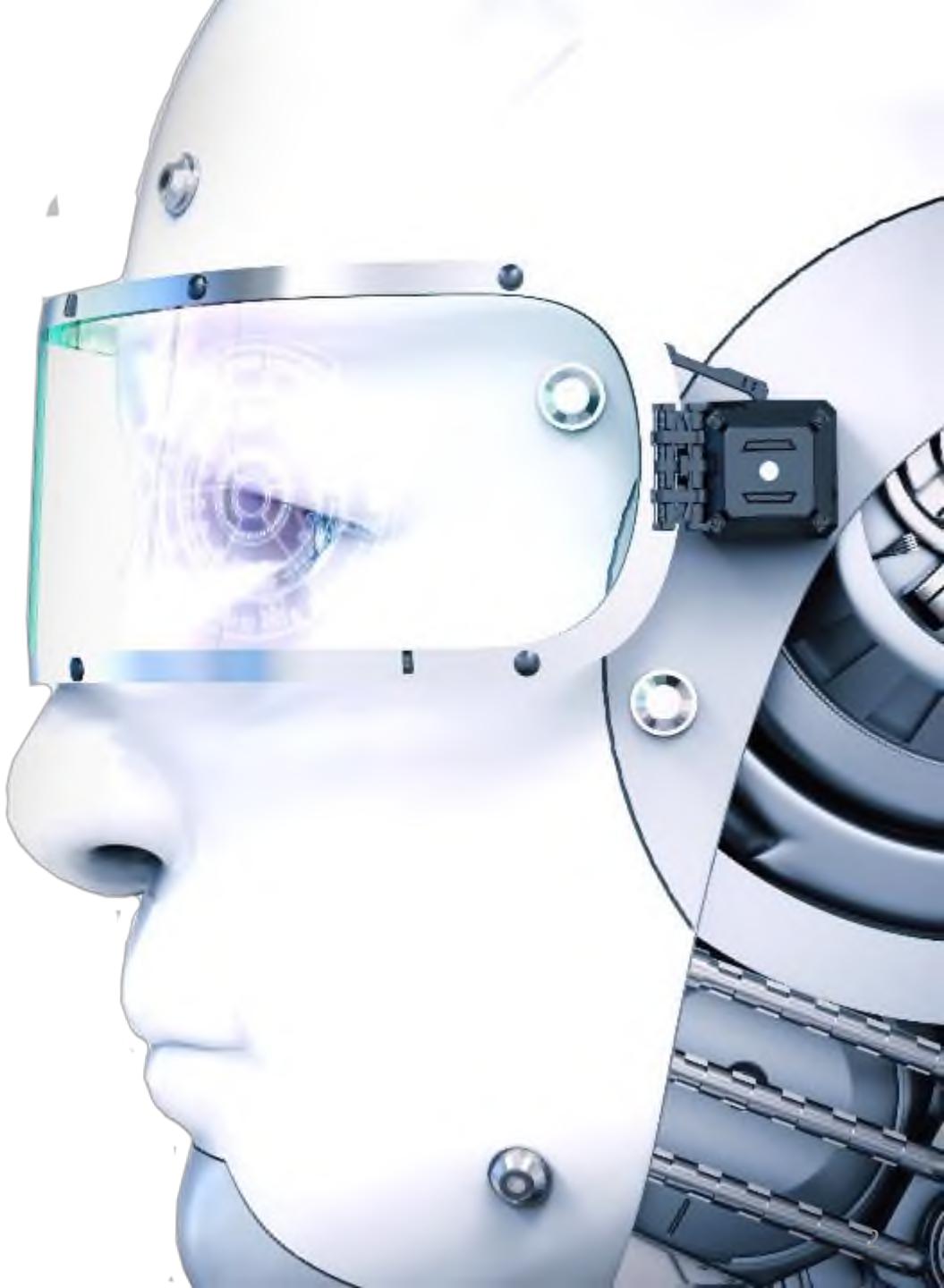
---

CREATING DOCUMENT  
DEFINITIONS



# Overview

- Document Definition
- Document Definition Structure and Objects
- Templates for semi-structured documents
- Document Definition Editor interface



## DOCUMENT DEFINITION

Defines the principles of identification and processing of a particular type of documents



# Ways to Mark Fields

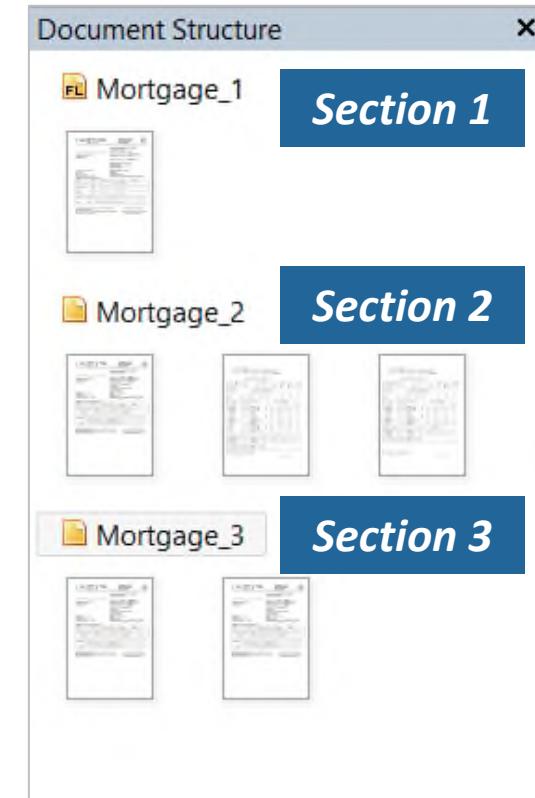
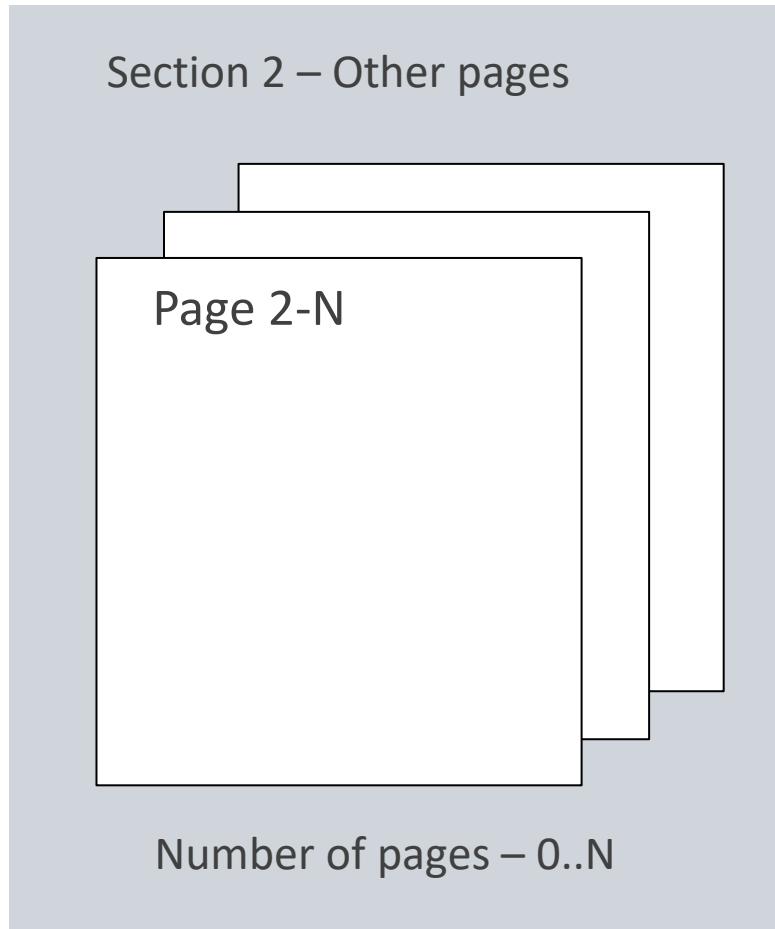
ABBYY®

DOCUMENT TYPE	EXAMPLE	MARKING TOOL
Semi-structured documents	Invoices Contracts	<ul style="list-style-type: none"><li>• ABBYY FlexiLayout Studio</li><li>• Field extraction training</li></ul>
Unstructured documents	Letters Tickets Business cards	Automatic search of regions by special description
Structured documents (hand-printed)	Applications Questionnaires Mortgage forms	<ul style="list-style-type: none"><li>• Document Definition Editor</li><li>• Manual fields' regions drawing</li><li>• ABBYY FormDesigner</li></ul>

# Document Definition Structure

ABBYY®

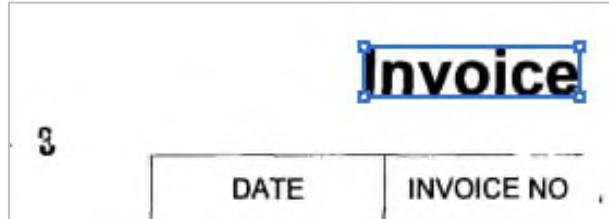
## Document 1 – Passport



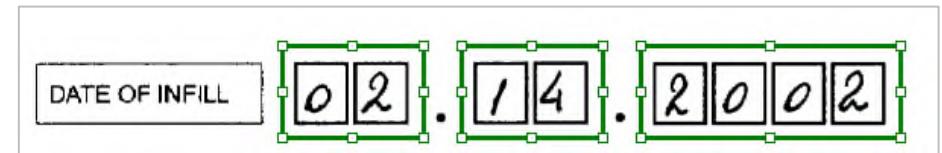
# Document Definition Objects

ABBYY®

## Identification elements



## Fields' regions (zones for recognition)



## Anchor elements (structured documents only)



# Document Identification

ABBYY®

To automatically select the Document Definition add **Identification elements**  
(unique objects for the document type)

Form Name

This form is a Subsidy for Children Application Form. It includes fields for date of birth (02.14.2002), parental information (YOUR NAME: MICHAEL), and a signature field (M). A note at the top states: "ATTENTION PLEASE! The form should be filled in with BLACK or DARK BLUE ink or ball pen with HANDPRINTED CAPITAL LETTERS according to the ABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890".

Company Name

Document Name

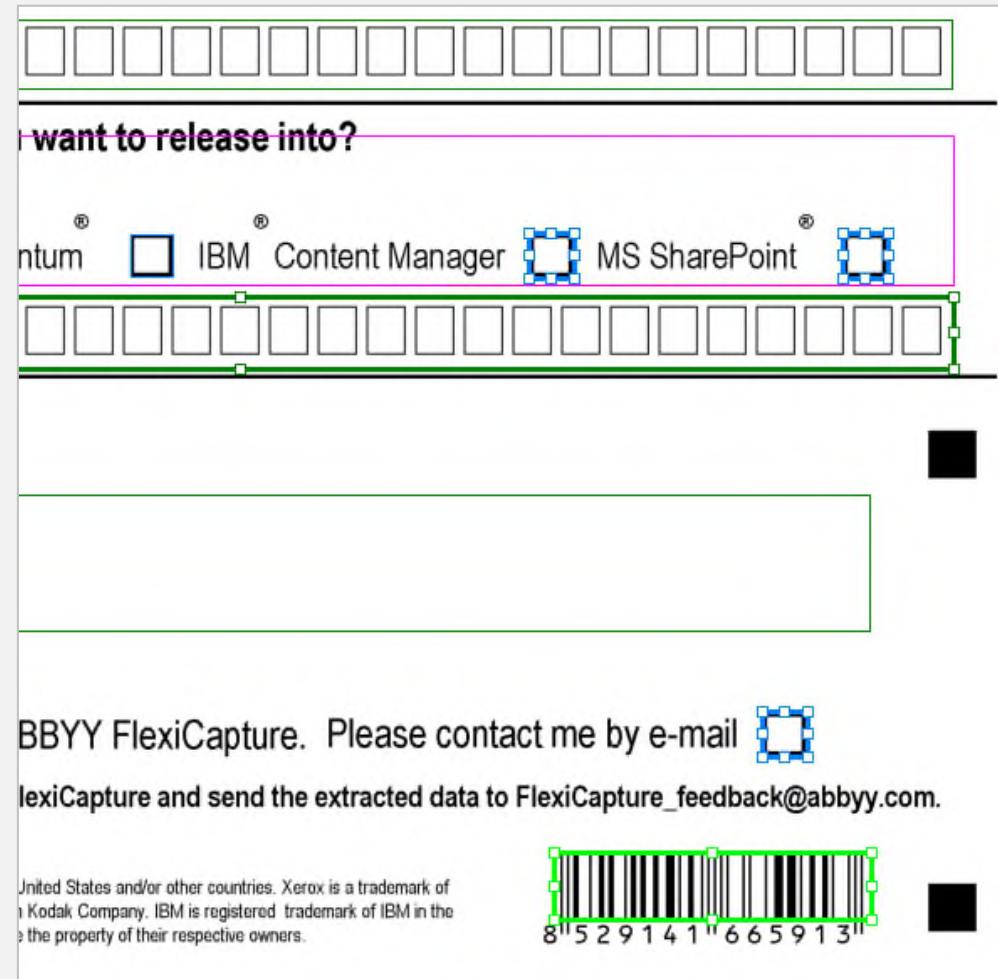
This invoice is from AA APPAREL, dated 6/09/04, with invoice number 99044. The recipient is US NORTHLAND, Doug Darrell, 18 Prince George Drive, Prince George, VA 23875. The document name is highlighted as "Invoice".

DATE	INVOICE NO
6/09/04	99044

## Fields' regions (zones for recognition)

- Text fields
- Checkmarks
- Barcodes
- etc.

## Fields have processing settings



# Main Steps of Document Definition Creation

ABBYY®

## 1 Wizard

Images for definition editor loading

Fields marking tool

## 2 Fields creation

## 3 Recognitions settings

## 4 Validation rules and data verification settings

## 5 Assembly and export settings

## 6 Definition testing

Document Definition contains:

- Identification elements
- Fields for recognition
- Processing settings

Several methods are available to mark fields:

- FlexiLayout
- Document Definition Editor – autodetection or manual marking

Document Definition Editor: tree of fields, image view, toolbar, data form

# ABBYY®

## ABBYY FLEXICAPTURE

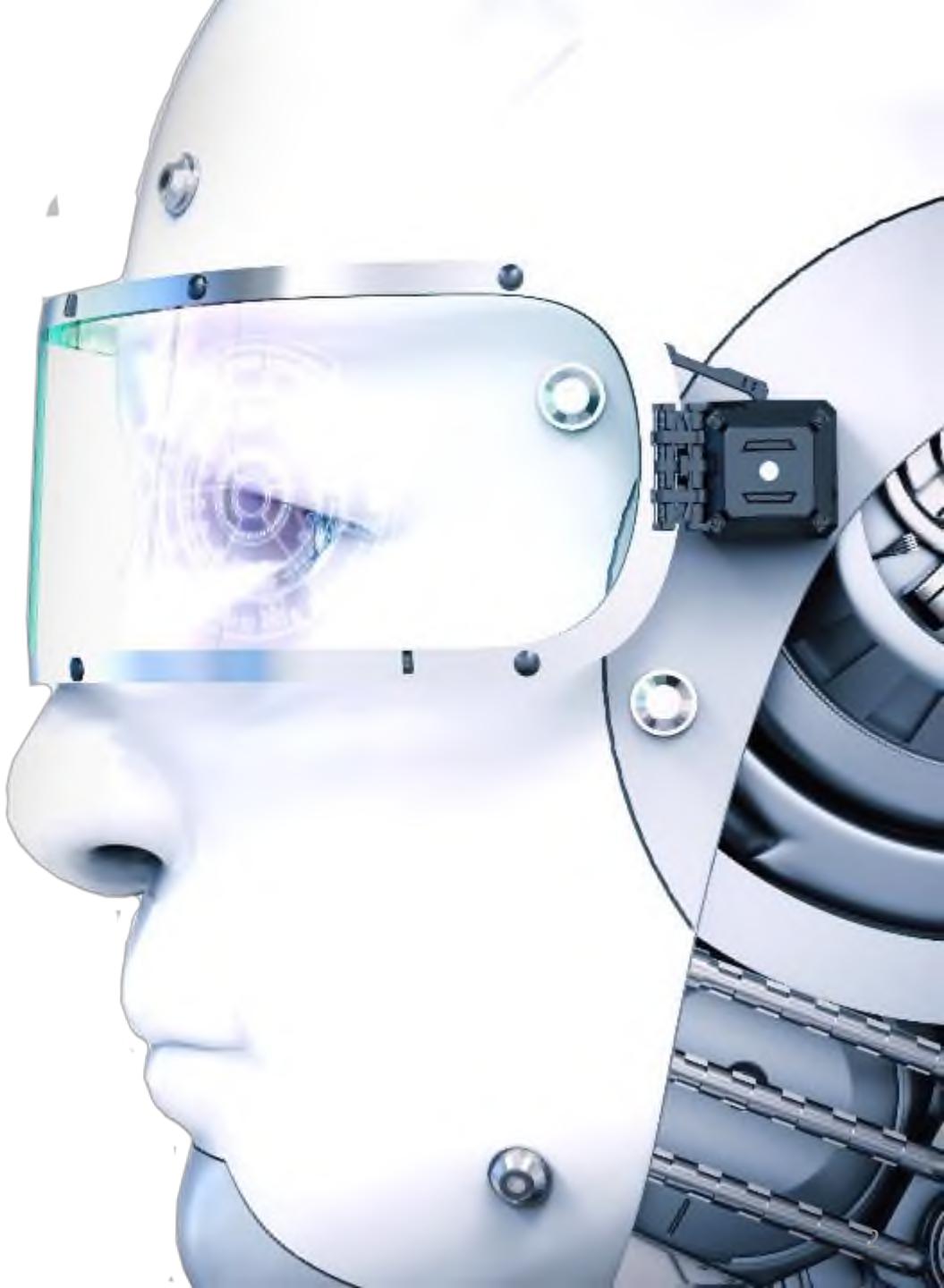
---

FIELDS CREATION



# Overview

- Types of Fields
- Basic field recognition properties
- Data Form view



Recognition fields

Service fields

Fields without regions

Groups of fields

Non-recognized region

Dependence on field types and properties:

- recognition results
- field treating by operators
- values for export

## For a semi-structured document

All fields are inherited from the FlexiLayout

## For a structured document

In the Document Definition Editor

- Manually
- Automatically

First Name	[Placeholder: First Name]
Last Name	[Placeholder: Last Name]
City of Birth	[Placeholder: City of Birth]
Birth Date (mm.dd.yyyy)	[Placeholder: Birth Date]

DATE OF BIRTH	06.23.1968	FAMILY STATUS
ADDRESS AND CONTACT PHONE:		
Street	HARPER VALLEY	ab

# Recognition Fields

ABBYY®

Text field

Main Credit Card Number

5447 5847 4707 4747

Checkmark

by e-mail

Checkmark Group



Barcode



Picture



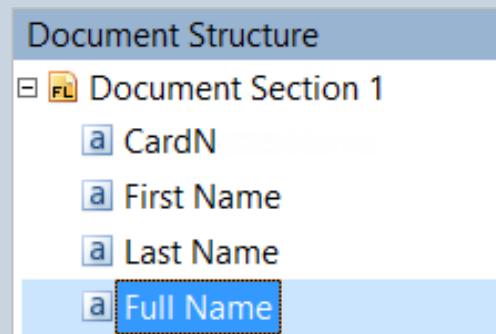
Table

Reference	Description	Transcription	Line	Unit	Qty	Unit Price	%	Total Unit	Notes
INPUT.PRESENTATION	Display A (PA142)	2020 material direct		Line	0	761.00		0.00	
INPUT.PRESENTATION	Display B (PA142)	2020 material direct		Line	1	460.00		460.00	
INPUT.PRESENTATION	Display C (PA142)	2020 material direct		Line	2	120.00		240.00	
INPUT.PRESENTATION	Display D (PA142)	2020 material direct		Line	3	260.00		780.00	
INPUT.PRESENTATION	Display E (PA142)	2020 material direct		Line	4	110.00		440.00	
INPUT.PRESENTATION	Display F (PA142)	2020 material direct		Line	5	30.00		150.00	
INPUT.PRESENTATION	Counter Display (PA142)	2020 material direct		Line	6	100.00		600.00	
INPUT.PRESENTATION	Counter Display (PA142)	2020 material direct		Line	7	100.00		700.00	
INPUT.MATERIALS	Mineral Oil (M1)	2020 material direct		Line	8	250.00		2000.00	

# Fields Without Regions

- Calculated field (no region => no recognition)
- Fields without regions may be used to store additional information  
E.g. result of merging data by a rule

Fields in the document structure



Recognized document

<u>CardN</u>	4567 2312 4455 0912
<u>First Name</u>	James
<u>Last Name</u>	Mills
<u>Full Name</u>	James Mills

# Service Fields

- Contains additional information on batch and document processing
- Values are specified automatically when processing document

The screenshot shows two windows related to service fields in ABBYY FlexiCapture.

**Create Field** window:

Type	Hotkey
Text	Alt+Shift+T
Checkmark	Alt+Shift+C
Checkmark Group	Alt+Shift+M
Barcode	Alt+Shift+B
Picture	Alt+Shift+P
Table	Alt+Shift+L
Group	Alt+Shift+G
Service Field	

**Service Field Properties** window:

Source: Flexible Section Variant ID

Parameter Name: (dropdown menu showing options)

Data type: (dropdown menu showing options)

Maximum text length: (text input field)

Available options in the dropdown menus:

- Source: Flexible Section Variant ID, Batch documents count, Batch name, Batch pages count
- Parameter Name: (dropdown menu showing options)
- Data type: (dropdown menu showing options)

Fields can be merged into *Groups* to

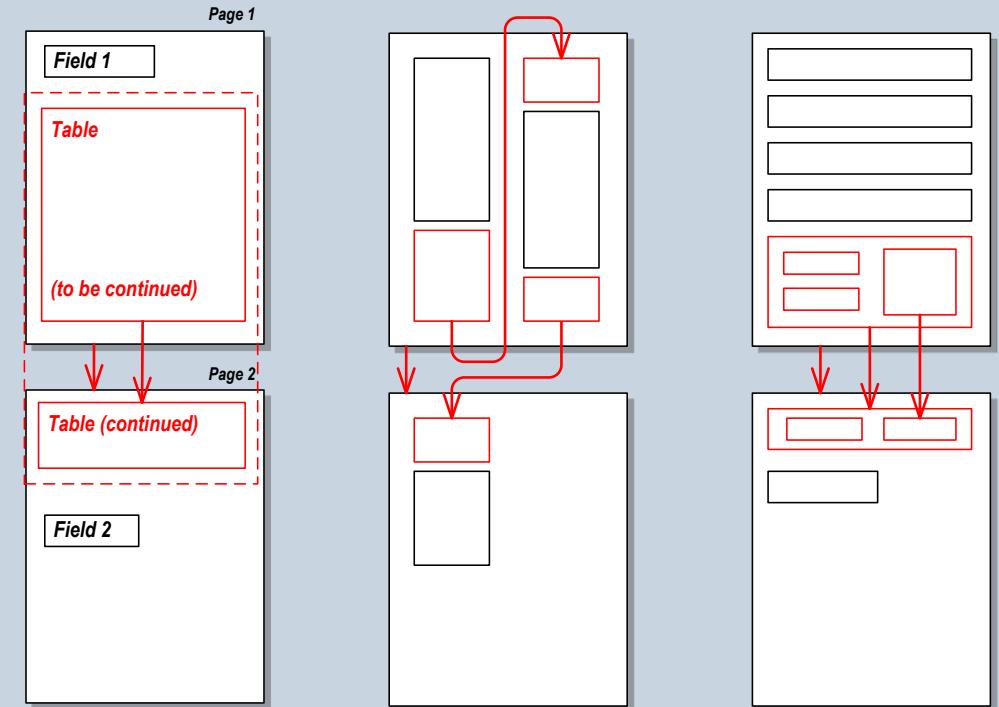
- simplify the structure of the Document Definition
- be able to perform group operations on multiple fields

- Name Group
  - [a] Name from 1099 form
  - [a] Name from Paycheck Stub
  - [a] Name from Tax Return
  - [a] Name from US Passport
  - [a] Name from W-2 form
  - [a] Name from Credit Card Statement
- SSN Group
  - [a] SSN from Paycheck Stub
  - [a] SSN from Tax Return
  - [a] SSN from W-2 form

# Fields with Several Regions

Fields with several regions:

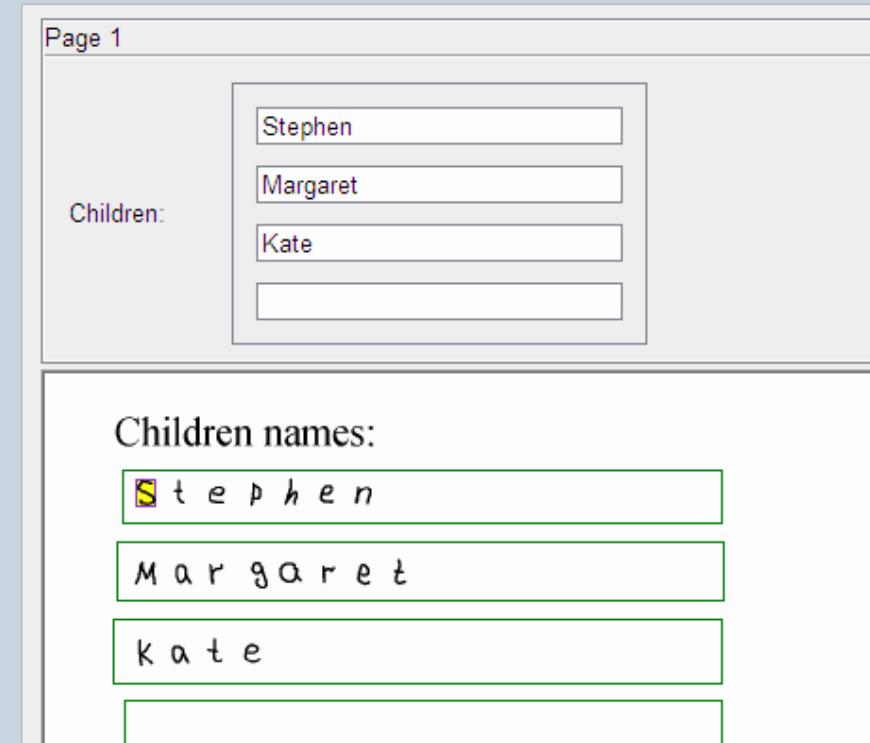
- comprise several regions on the image
- values from all regions are combined and exported together as one field
- *Continue Region...* command



# Fields with Several Instances

Fields or field groups that occur multiple times and contain similar information

- names of children
  - account numbers
  - etc.
- 
- *New Instance...* command



# Text Field: Marking Type

F l e x i C a p t u r e	Simple
<u>F l e x i C a p t u r e</u>	Underlined
<u>F.l.e.x.i.C.a.p.t.u.r.e</u>	Simple comb
<div style="border: 1px solid black; padding: 2px;">F l e x i C a p t u r e</div>	Text in frame
<div style="border: 1px solid black; padding: 2px;">F.l.e.x.i.C.a.p.t.u.r.e</div>	Comb in frame
<div style="border: 1px solid black; border-collapse: collapse; padding: 2px;">F l e x i C a p t u r e</div>	Partitioned frame
<div style="border: 1px solid black; border-collapse: collapse; padding: 2px;">F l e x i C a p t u r e</div>	Char box series
<div style="border: 1px solid black; border-collapse: collapse; padding: 2px;">F l e x i C a p t u r e</div>	Gray boxes

## Important:

Incorrect settings for “Marking type” and “cells” may adversely affect the quality of recognition!

# FIELD RECOGNITION SCRIPT

---



# Field Recognition: Script

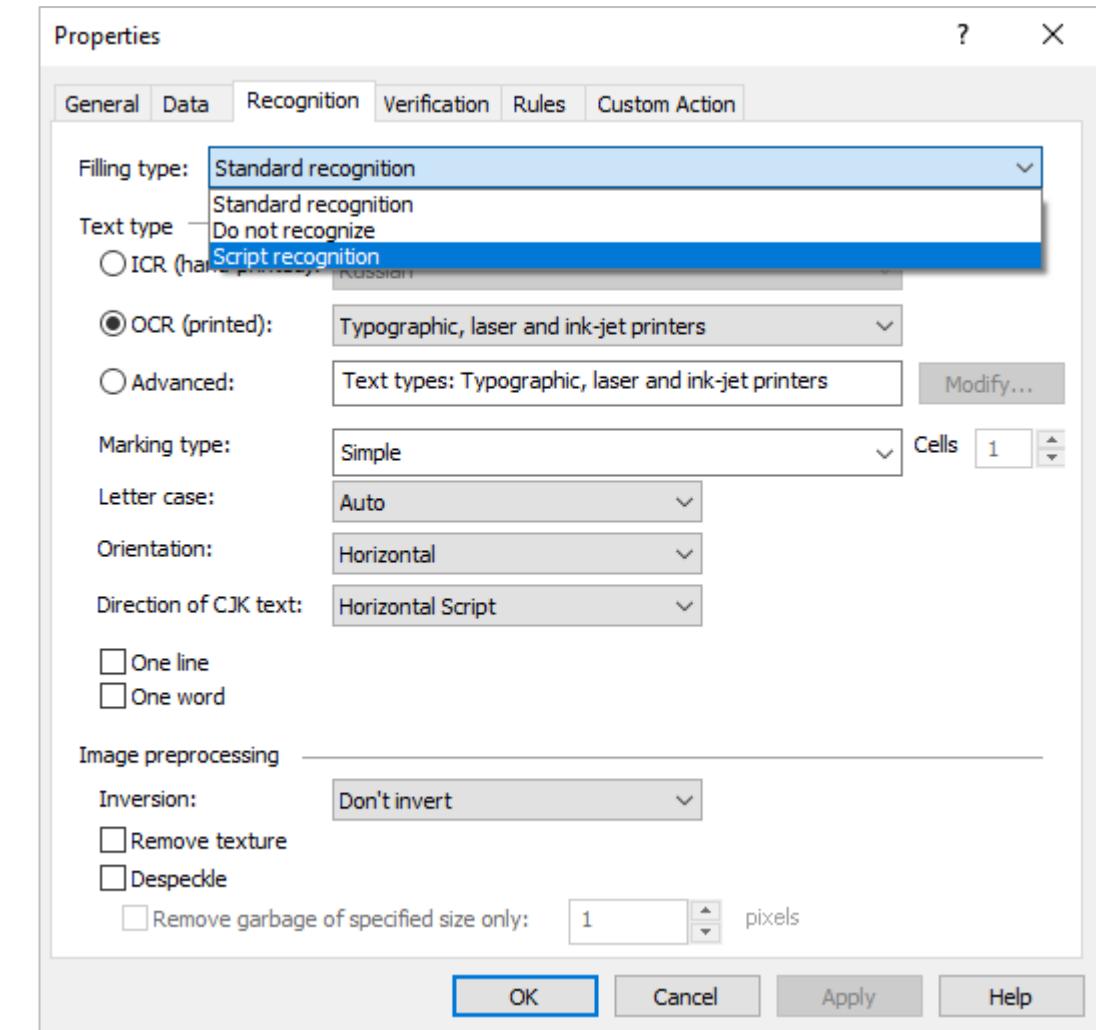
## Via custom recognition script

- Access to field regions
- Access to document registration parameters
- No access to the document fields

Field region can be recognized using  
the external engine

Image area being recognized is passed  
to the script as an input parameter

- *FieldRegion* parameter



Document Definition contains:

- Fields for recognition (with regions)
- Additional fields (calculated or service)

For semi-structured documents, fields are inherited from a FlexiLayout

Fields can be multi-page:

- With regions continuing on several pages
- With repeating instances



## ABBYY FLEXICAPTURE

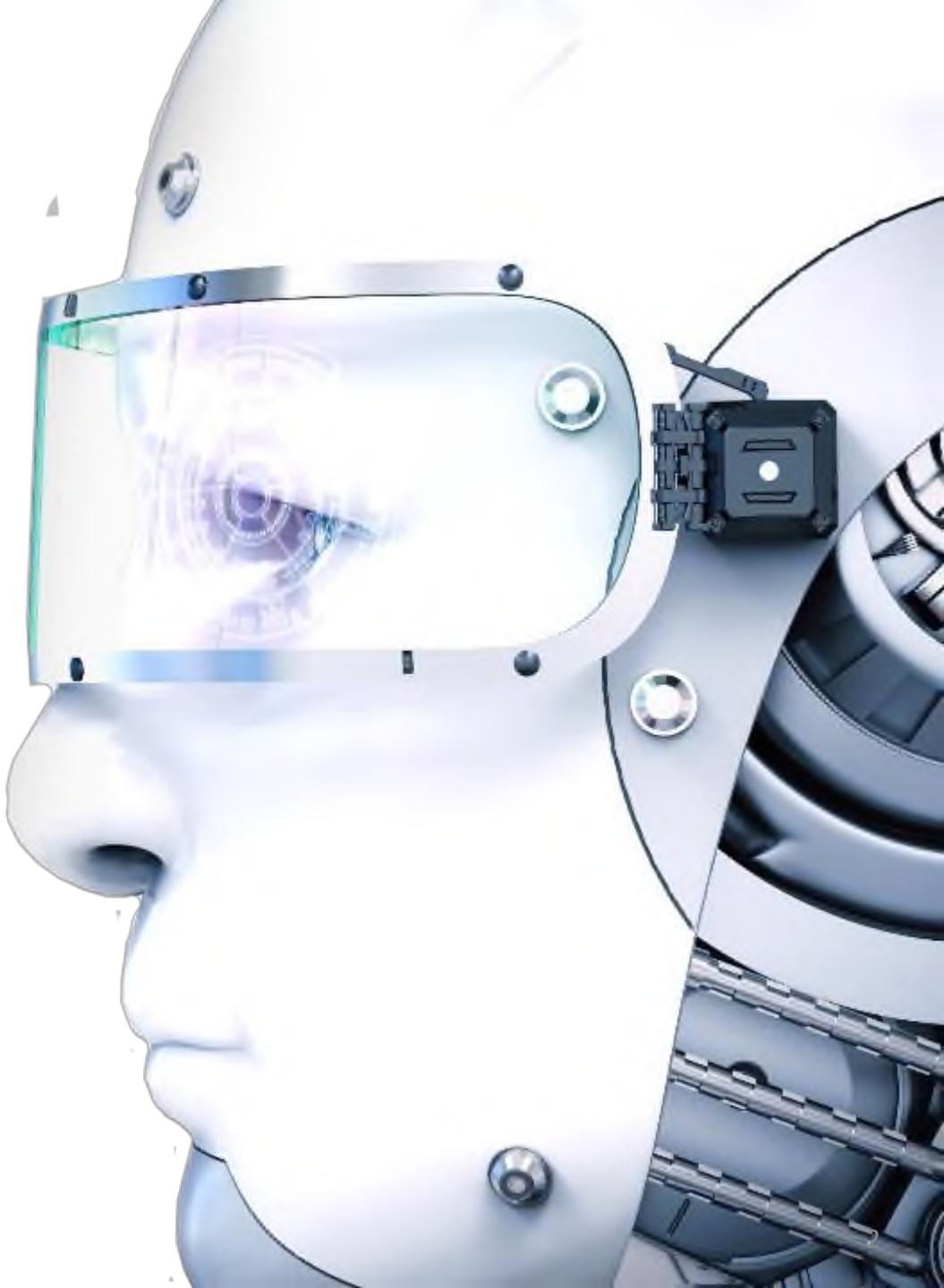
---

### DATA TYPES



# Overview

- Why improve recognition quality?
- Data types
  - Languages and alphabet
  - Format checks
- Overview of the standard data types
- Creating custom data types
  - Alphabet
  - Dictionaries
  - Regular expressions
- Recognition options



The Administrator's task is to reduce the amount of data to be checked:

- Data type

Describing alphabets and dictionaries

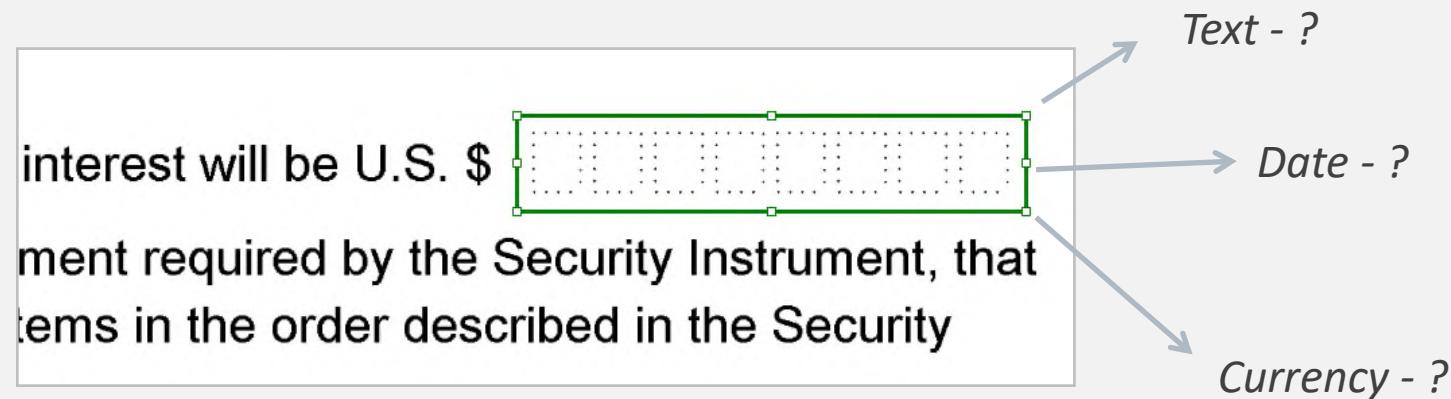
Checking the format of the data and imposing constraints

- Recognition options

Data type is a set of possible values in a field and automatic data normalization and validation

- Has certain additional constraints on the format of data

can be used to check if the data is correct

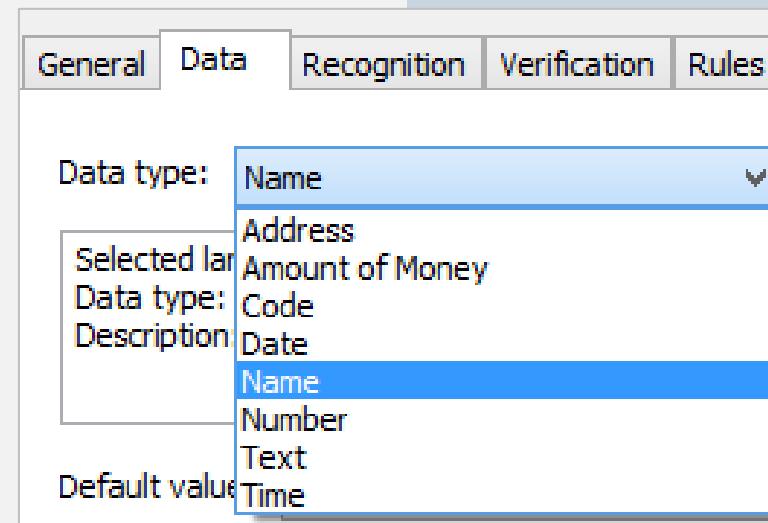


Predefined dictionaries and alphabets  
for common types of data

- Names
- Numbers
- Address details
- Etc

User-created types of content based on

- Dictionary
- Regular expression
- Alphabet



# Data Type Parameters

ABBYY®

## Content

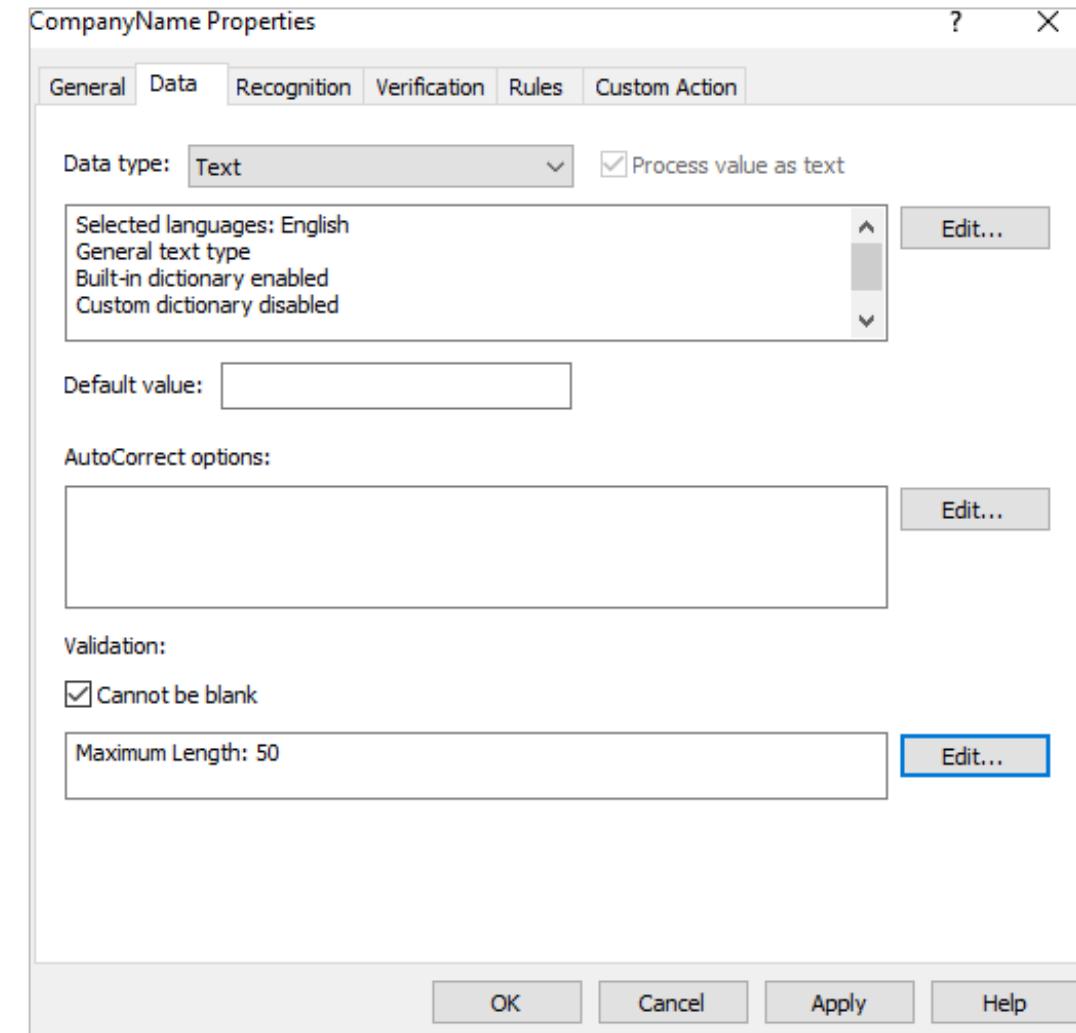
- Increases recognition quality by applying alphabet and dictionaries
- Checks the field format

## AutoCorrect

- Value conversion
- Normalization

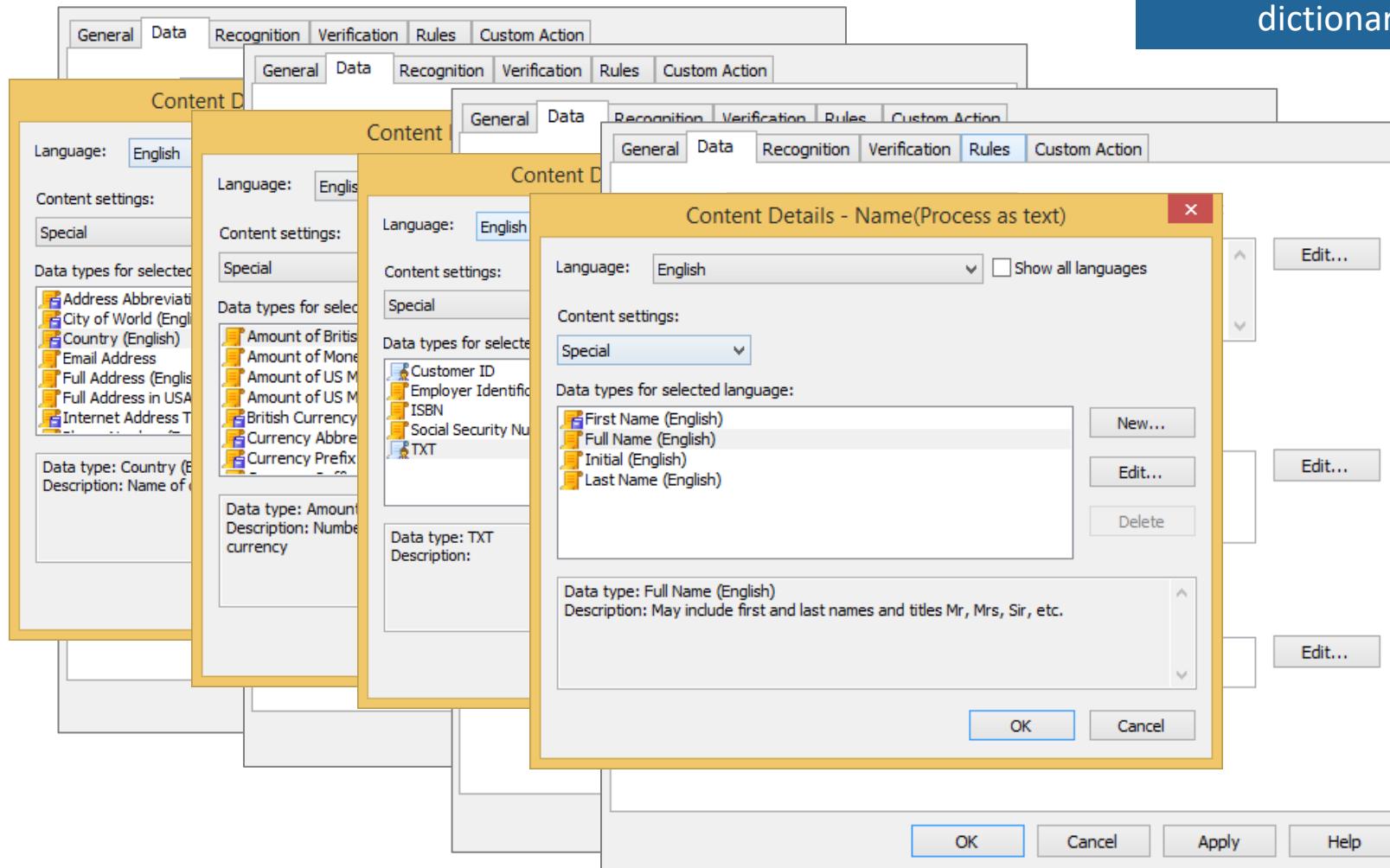
## Validation

- List of allowed values
- Number constraints
- Date interval

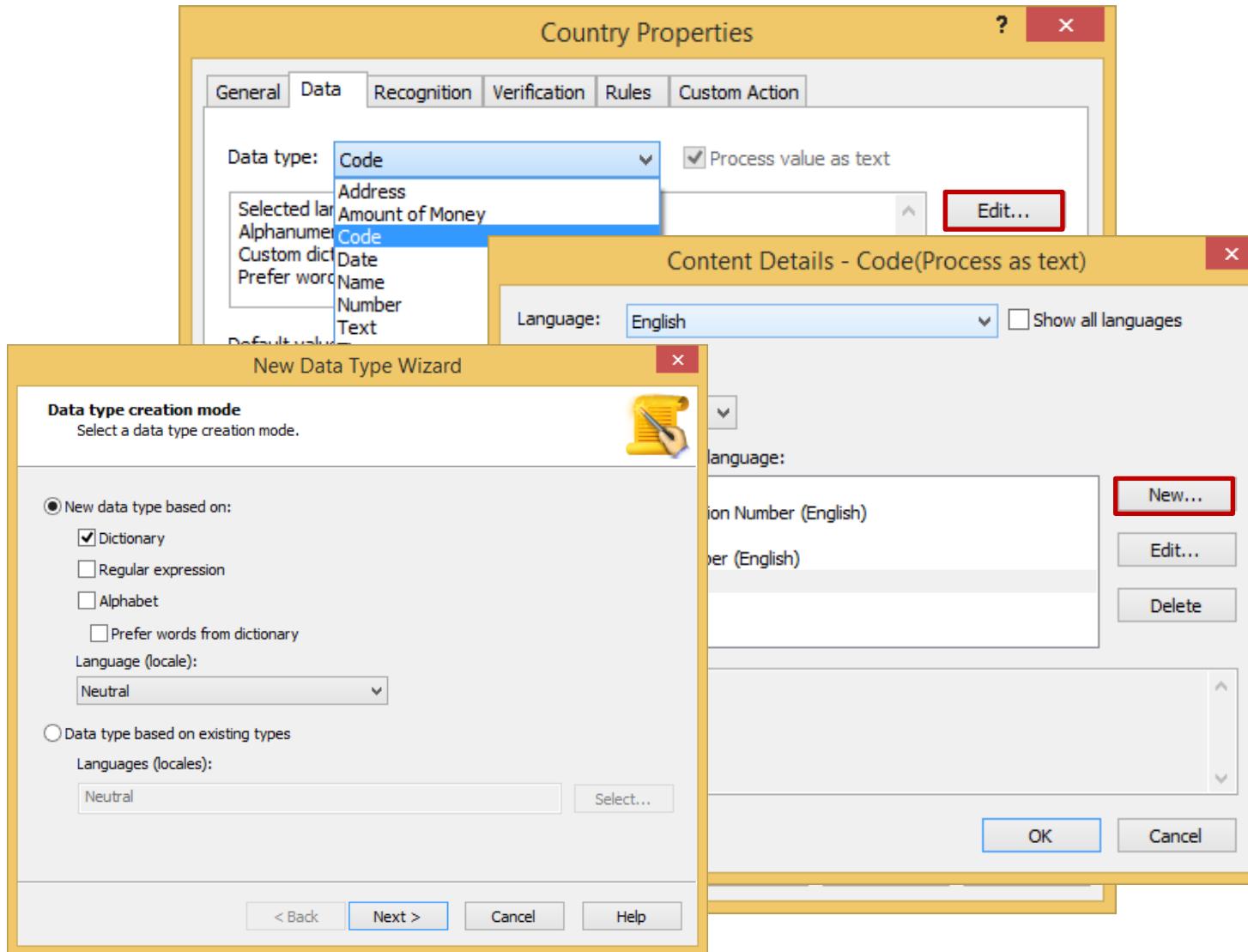


# Standard Data Types

Set of built-in  
dictionaries



# Custom Content Types



If the standard data types do not completely describe the data in the fields, you can create your own **custom data type** based on:

**a dictionary**

(a word list or \*.txt, \*.dic files)

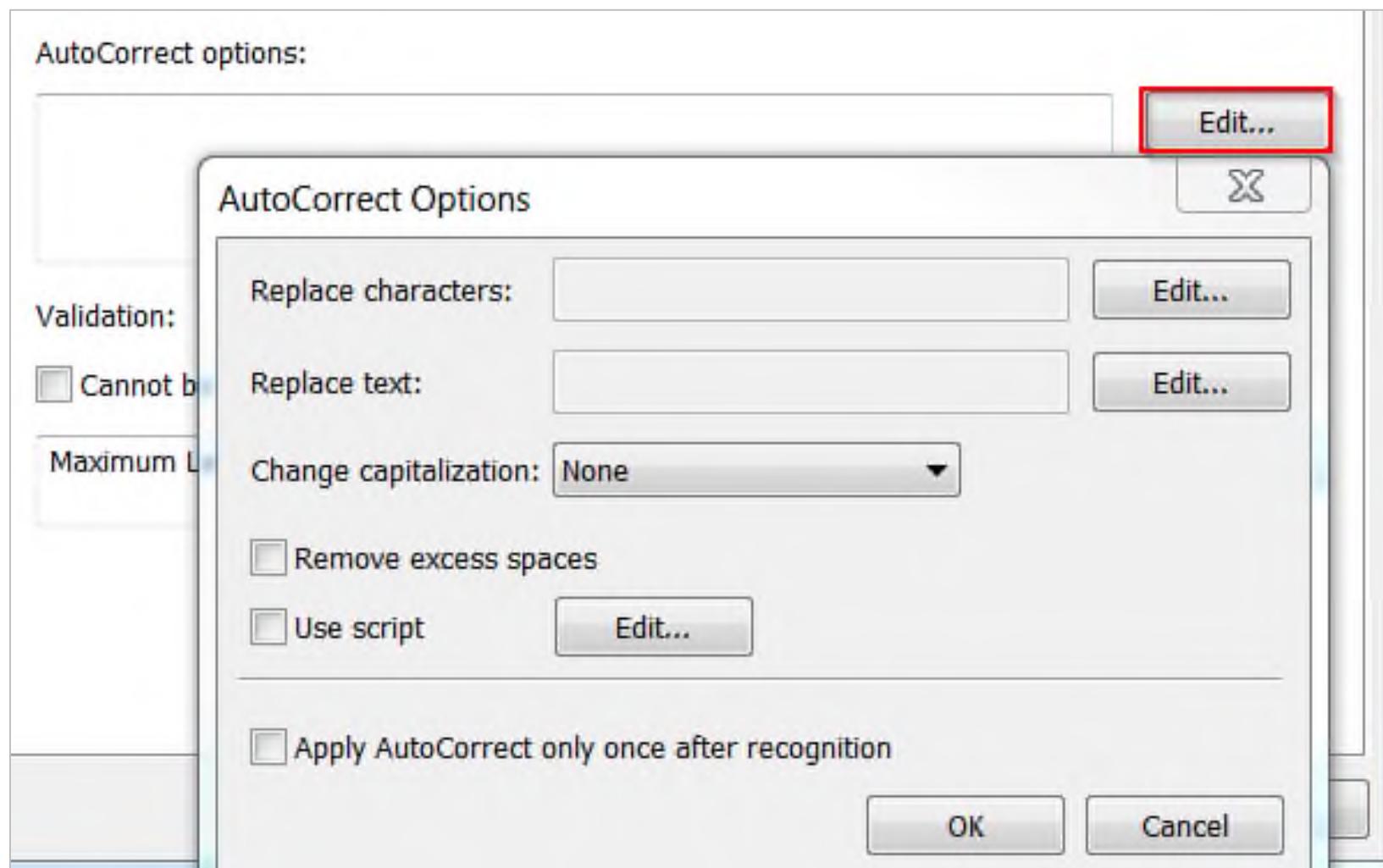
**a regular expression**

describes the format of the data limits the set of words possible in this field (works when a field is recognized)

**an alphabet**

the set of allowed characters (+ limiting subsets)

# AutoCorrect Options



# Date Types: AutoCorrect Options

ABBYY®

After recognition fields' values can be automatically modified:

- Characters can be replaced

“Ms Merilyn Monro”  
=> “Merilyn Monro”

- Text may be changed

EliCe -> ELICE

- Letter case may be changed

“Peter Harms” =>  
“Peter Harms”

- Excess spaces may be removed

- A script may be applied

# Autocorrecting by Script. Sample

ABBYY®

This rule checks the number of uncertain characters in the field and clears the field if the number of uncertain characters is greater than or equals three.

[VBScript]

```
Dim totQty, suspQty
Dim i
totQty = me.Symbols.Count
suspQty = 0
for i = 0 to totQty - 1
  if me.Symbols.Item(i).IsSuspicious = true
    then suspQty = suspQty + 1
  end if
next
if suspQty >= 3 then
  me.Text = ""
end if
```

---

Recognition quality can be improved by specifying the data types for the fields

---

Wide range of pre-defined data types (dictionaries and data formats) are available

---

Custom data types can be created

---

Data types allow to check and adjust the format of recognized data



## ABBYY FLEXICAPTURE

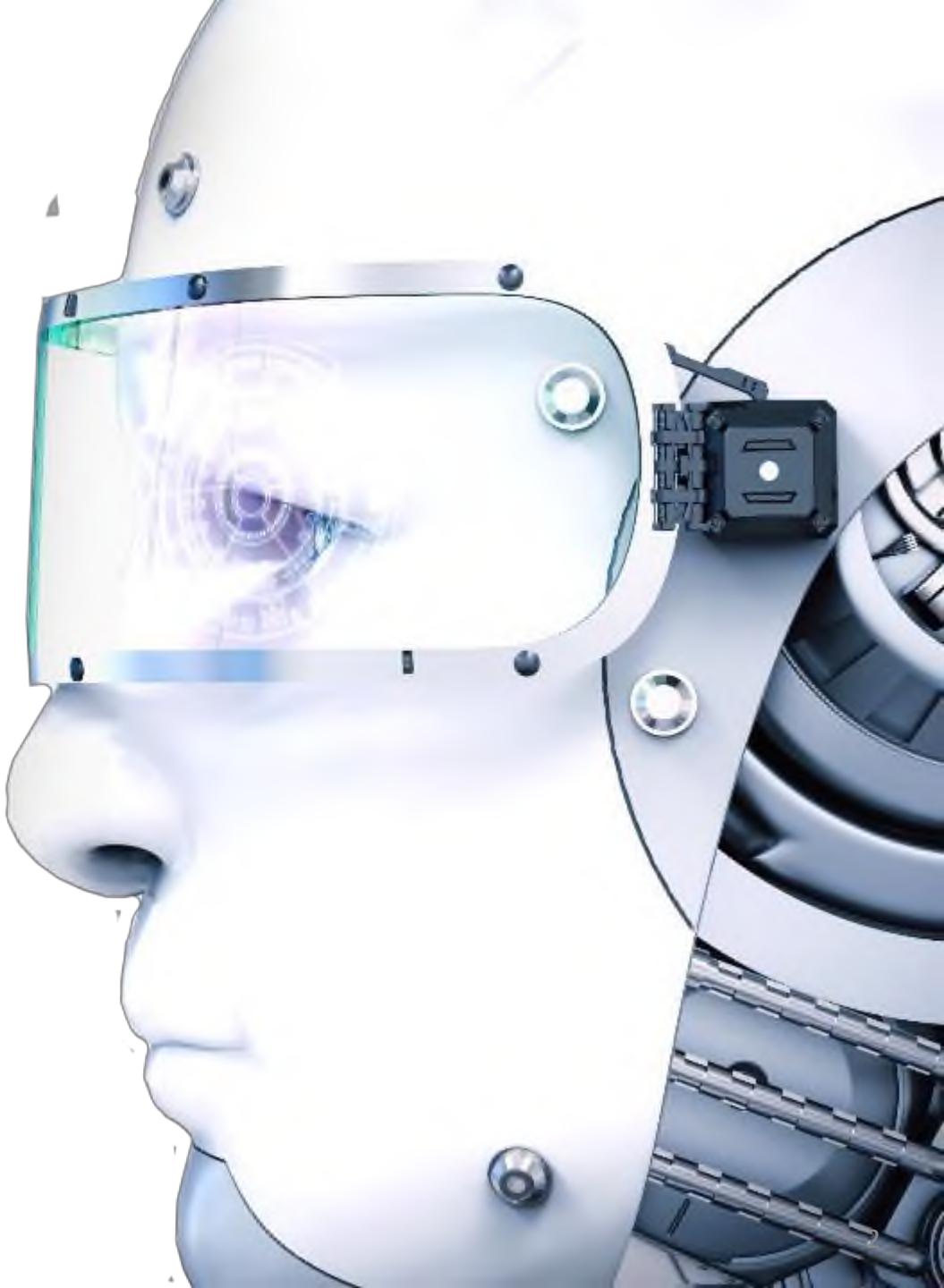
---

### VALIDATION RULES



# Overview

- ○ Why Implement Rules
- ○ Rules usage scenarios
- ○ How rules work
  - Rule creation
- ○ Rule strictness
- ○ Types of rules
- ○ Rules and their properties



# Why Validate Data

ABBYY®



## Check document structure

Make sure that no pages  
are missing



## Check data format

Make sure that no data  
is missing



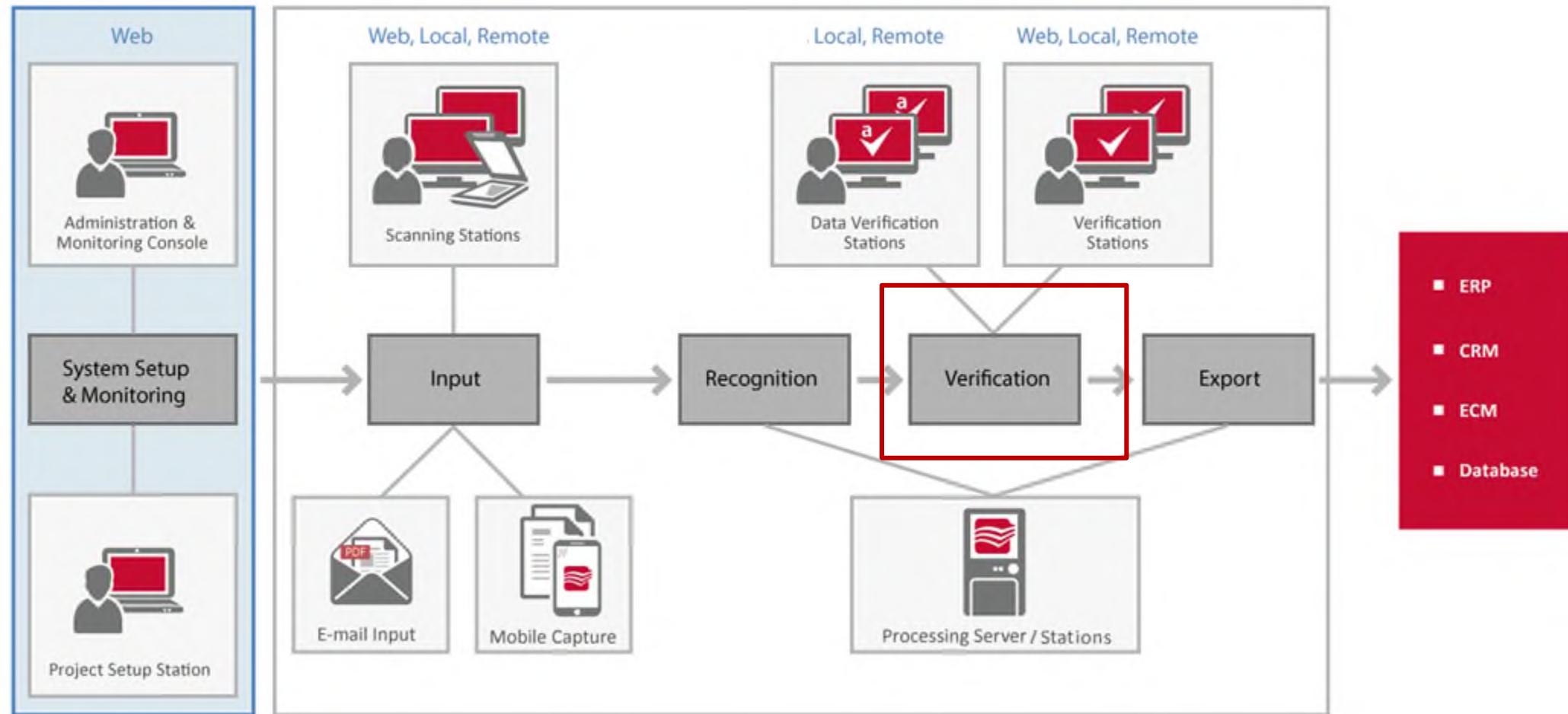
## Check the connections between data in different fields

Make sure that data integrity  
is maintained

Data can be modified and validated by using rules

# Document Processing: Data Validation Stage

ABBYY®



# Why Implement Rules

ABBYY®

Automatic check of the recognition results and data modification conditions:

- Data integrity check

Data completeness

Fields values relations

- Field values modification

Merging fields

Deleting excess data

Data normalization and formatting

# Rules Usage Scenarios (1)

## CASE

Check the presence of recognized value in the database

e.g. CompanyName

ID	CompanyName
1	INATON, Inc
2	AXA APPAREL
3	D. Brawn Manufacture
4	Techhouse
5	J. Goldsmith Manufacture
6	FLY Systems, Inc



## CASE

Merge several values recognized in separate fields in one

E.g. name details can be recognized in separate fields, but only the Full Name field with merged values must be exported

Name

First Name	<input type="text" value="Anna"/>	Initial	<input type="text" value="M."/>	Last Name	<input type="text" value="Ephron"/>
Full Name	<input type="text" value="Anna M. Ephron"/>				

Red arrows point from the individual First Name, Initial, and Last Name fields to the combined Full Name field below them.

# Rules Usage Scenarios (2)

## CASE

Calculate sum in the table cells and compare it with the total sum field  
E.g. any financial document

- 1 -

■ FORM 01 CUSTOMER'S REFERENCE DATA CONFIRMATION  
Basic information for entry of information on the Customer's Data in accordance with the General Terms and Conditions and the procedures established by applicable laws

PERSONAL DETAILS

Mr  Ms  Mrs  Misses

First Name: JENNIFER Middle Initial: A  
Last Name: PITT Nationality: BRITISH

City of Birth: LONDON Birth Date:

## CASE

Compare several fields  
E.g. names from different fields of the document

- 2 -

■ FORM 02 ACCOUNT APPLICATION

First Name: JENNIFER Middle Initial: A  
Last Name: PITT  
Birth Date: 01/06/1966 Social Security Number: 19706442

KINDLY REQUEST TO OPEN (CHECK ONLY ONE)

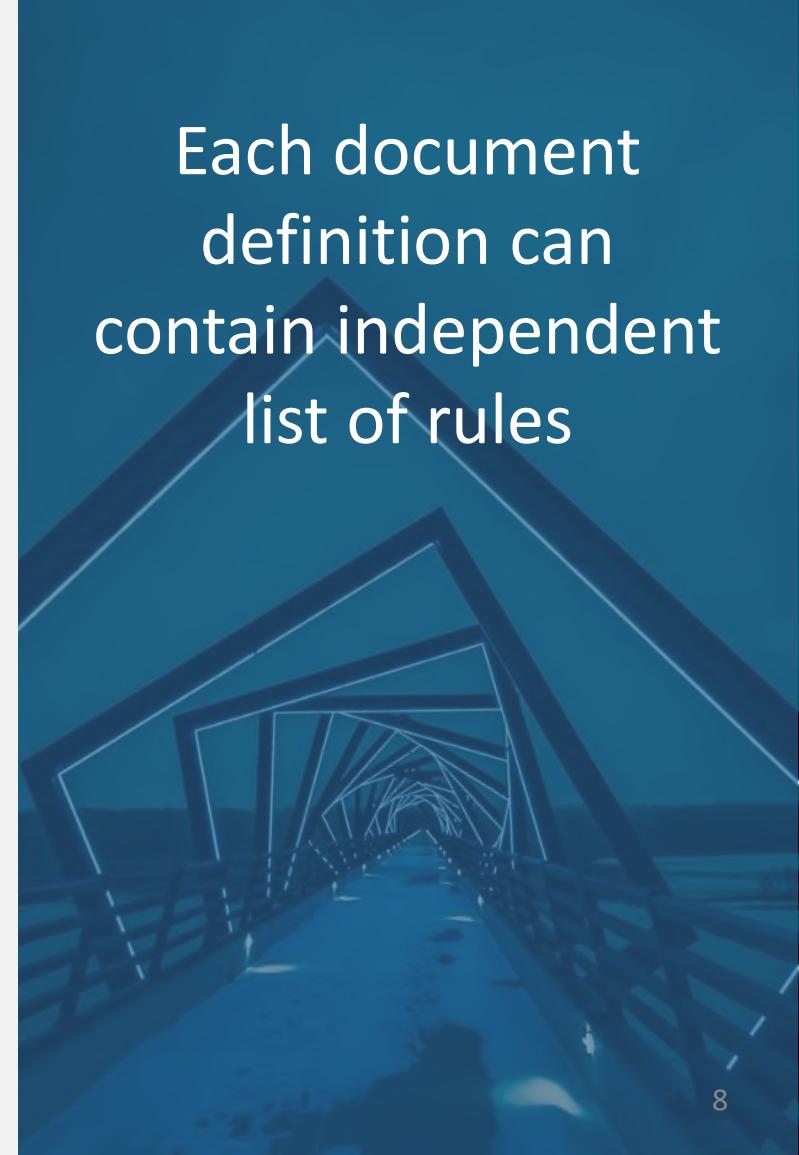
Settlement Account  Current Deposit Account  
 Card Account  "ZETA Bank" Account  
In the Currency (Enter Currency Code): 840

Rule can be

- Standard (adjusted by the UI)
- Custom (written in JScript, VBScript, .NET)

Rule can be applied to

- Simple fields (e.g. checkmark, text field)
- Elements of the complex fields (e.g. table cell)

A large, abstract graphic element on the right side of the slide. It features a perspective view of several nested, translucent blue hexagonal outlines that recede into the distance, creating a tunnel-like effect. The overall color palette is a monochromatic blue.

Each document definition can contain independent list of rules

# Types of Rules

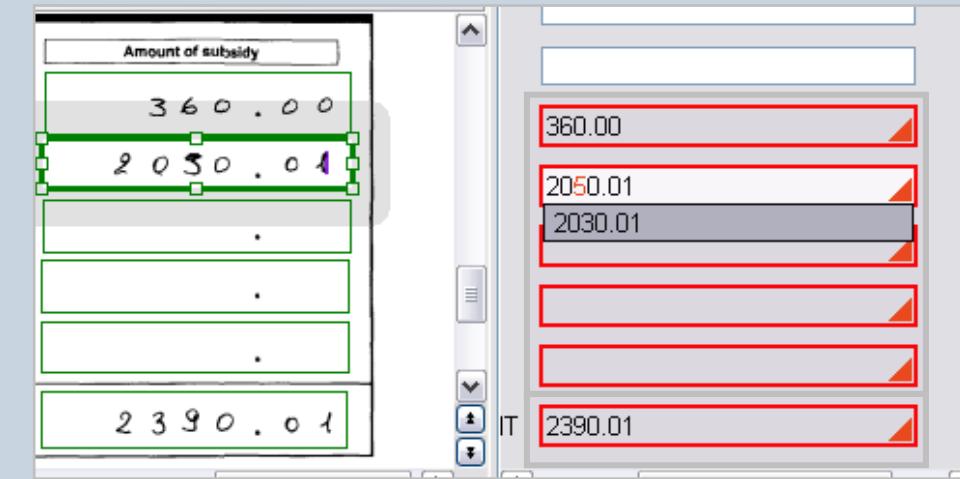
## Calculate Sum Rule

- Data modification rule
- The sum of fields indicated for summing up is automatically calculated and saved into the result field

Amount of subsidy	360.00
\$	360.00
\$	360.00
\$	.
\$	.
\$	.
Amount1	360.00
Amount2	360.00
Amount3	
Amount4	
Amount5	
TOTAL AMOUNT	720.00

## Check Sum Rule

Sums up the values of several fields and compares the result with the number specified in the rule or in another field



# Types of Rules (2)

## Compare Fields Rule

Compares the values of several fields

- fuzzy comparison is possible

▲ SSN Group

SSN from Paycheck Stub	701-41-742; 701-41-742:
SSN from Tax Return	701-41-7424 701   41   7424
SSN from W-2 form	701-41-7424 701-41-7424

## Database Check Rule

Compares the field value with the values from the database

- fuzzy comparison is possible

Database Table

A	B	C
ID	Code	PositionName
1 0001		Director
2 0002		President
3 0003		Chief Executive Manager
4 0004		Manager
5 0005		Accountant
6 0006		Assistant
7 0007		

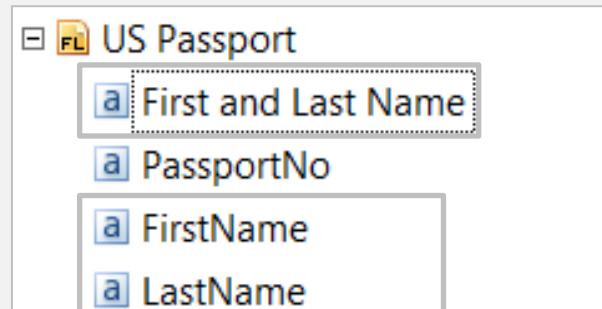
Position \* **MANAGER**

Position	Manager
----------	---------

# Types of Rules (3)

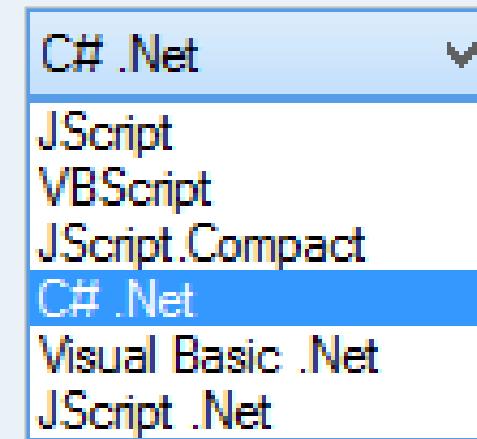
## ***Merge Fields*** Rule

- Merges the values of several fields into one
- Fields without region are used to store data
  - e.g. when merging data by applying a rule



## ***Script*** Rule

Scripts to perform data checks and modifications can be used



# Rules and Object Hierarchy

ABBYY®

A rule can be created in the properties of:

- Document Definition
- section
- field

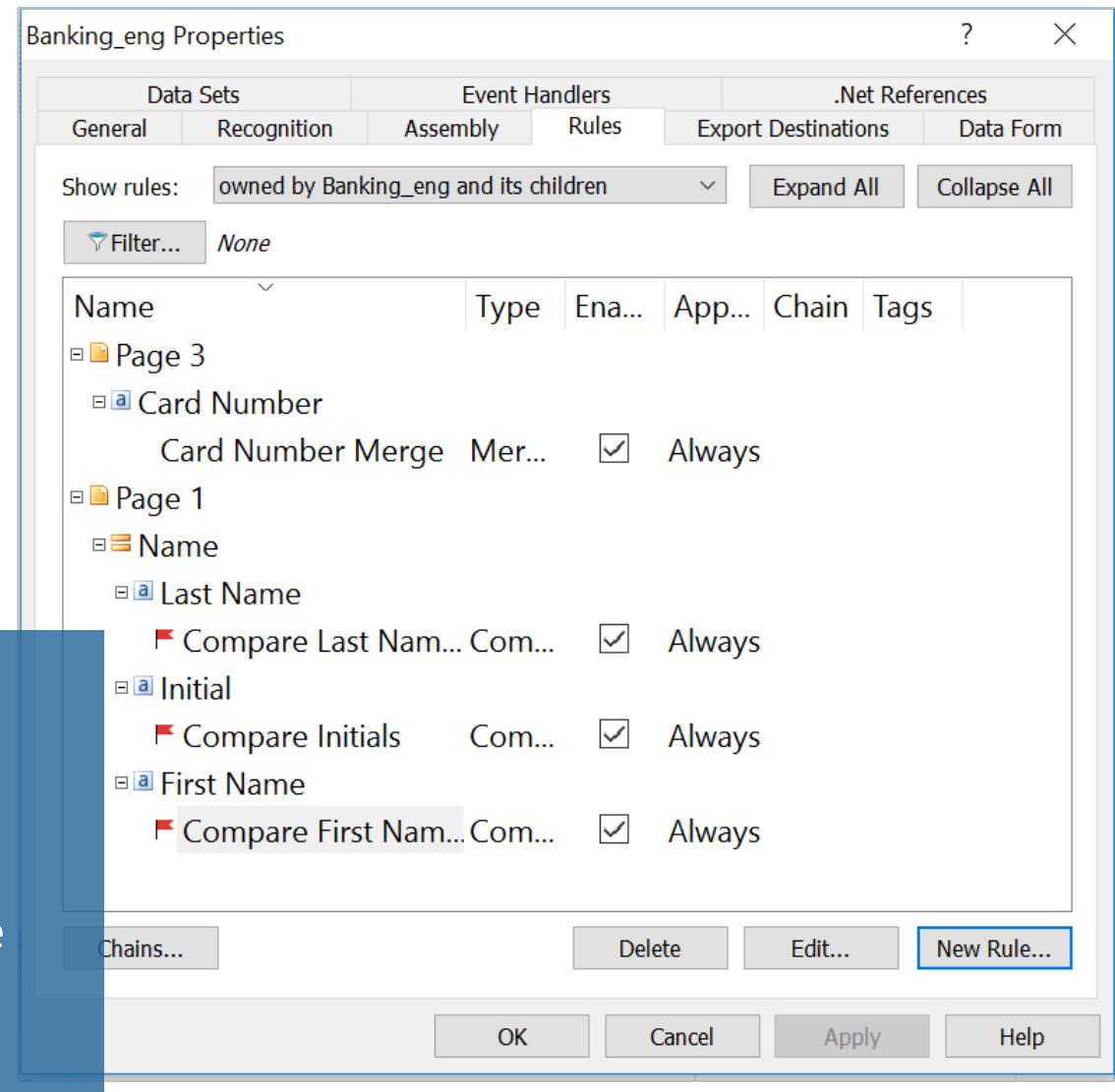


## Note:

In most cases, it makes no difference at which level to create a rule.

## Exception:

tables and repeating fields (fields with multiple instances) – rules must be created in the properties of these elements only.



---

Rules allow to

- avoid mistakes in the captured data in case of wrong filling in the documents
- check additional relations between the document fields

---

Rules are created for each documents type separately

---

Data errors exposed by rules are checked and corrected by the operators at the document verification stage



## ABBYY FLEXICAPTURE

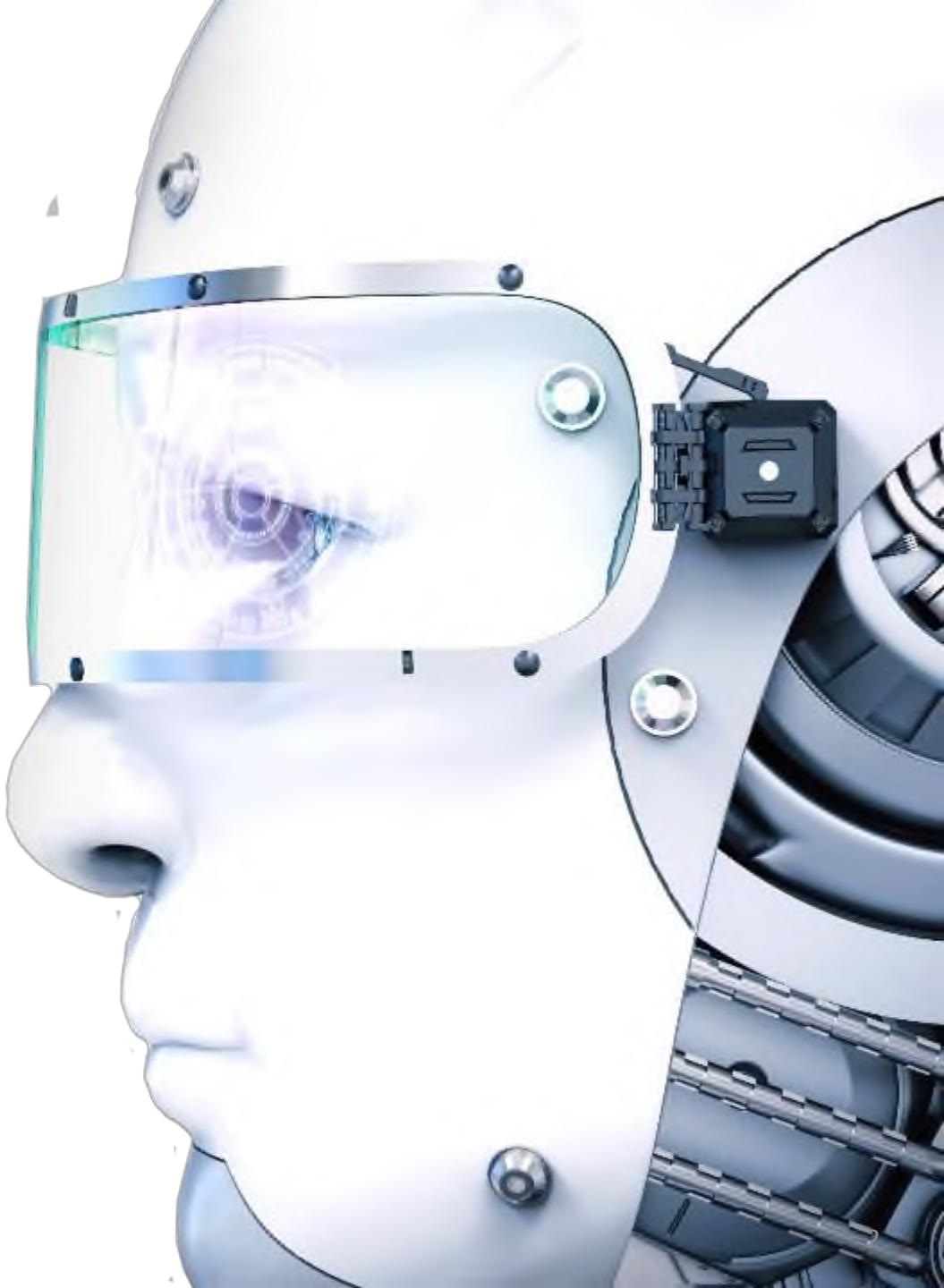
---

### ADVANCED VALIDATION RULES



# Overview

- DB check usage scenarios
- Script rules
  - Settings
  - Script editor
  - Samples
- Custom Action script



# DB CHECK USAGE SCENARIOS

---

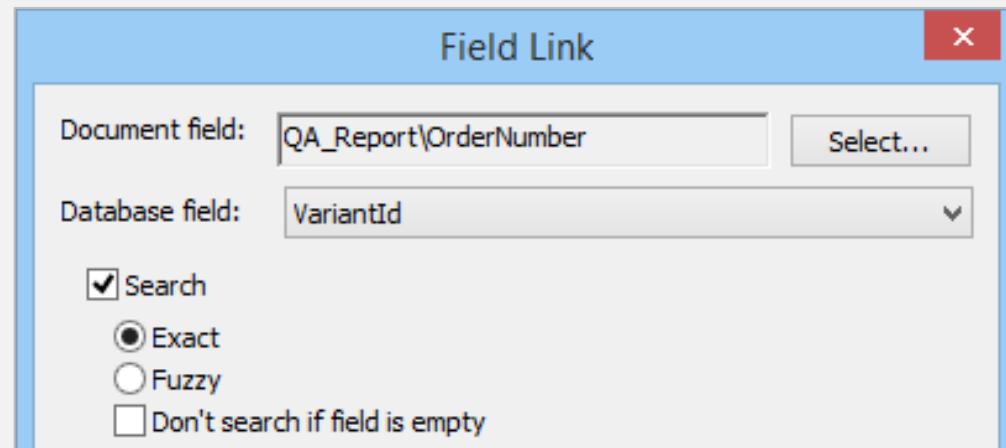


# Simple Database Lookup

1. Check if the value is present in the database
2. Comparison types:

Exact - the recognized value and the value from the database must be the same

Fuzzy - search for the most appropriate word in the database

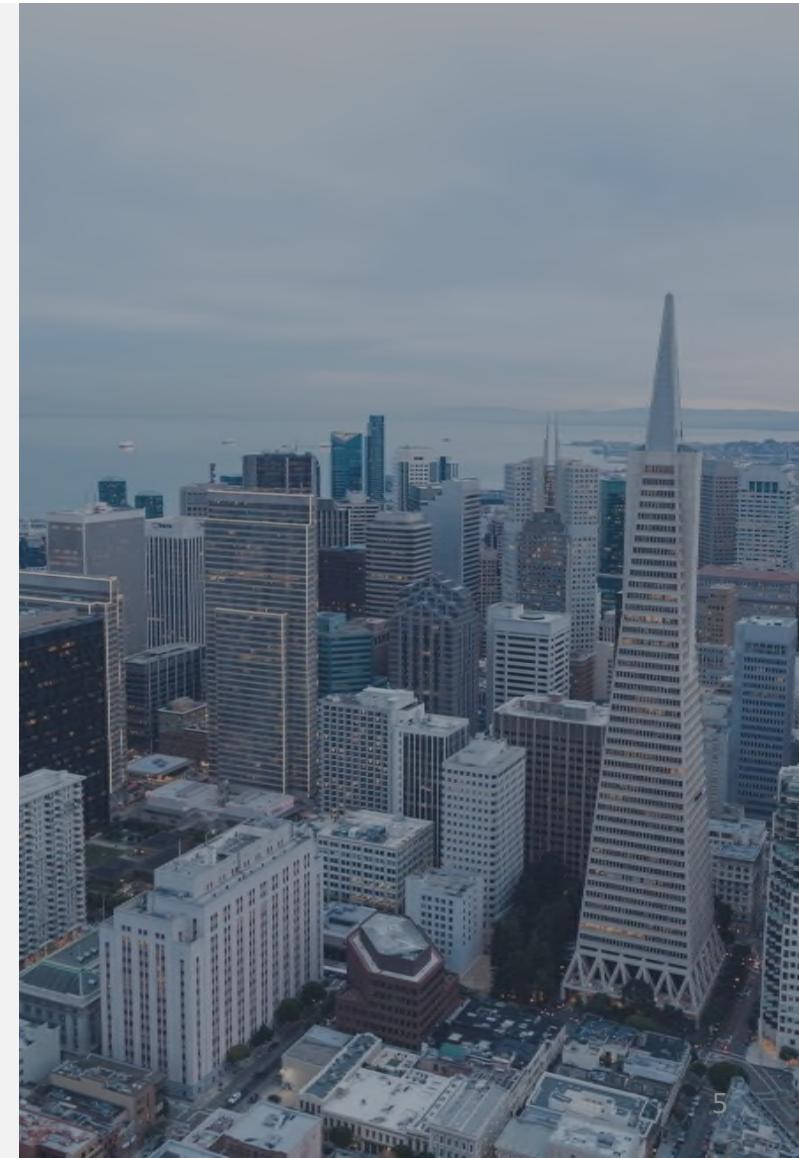
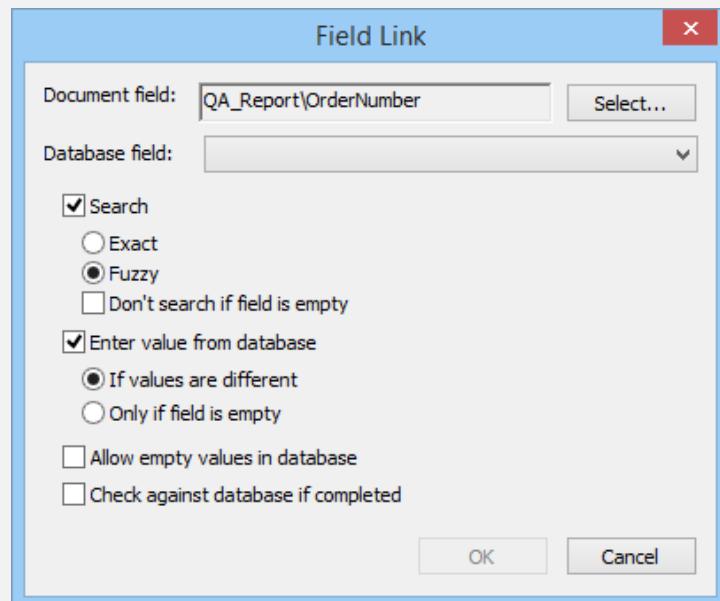


Certain number  
of errors is allowed  
depending  
on the value length

# Database Lookup with Substitution

ABBYY®

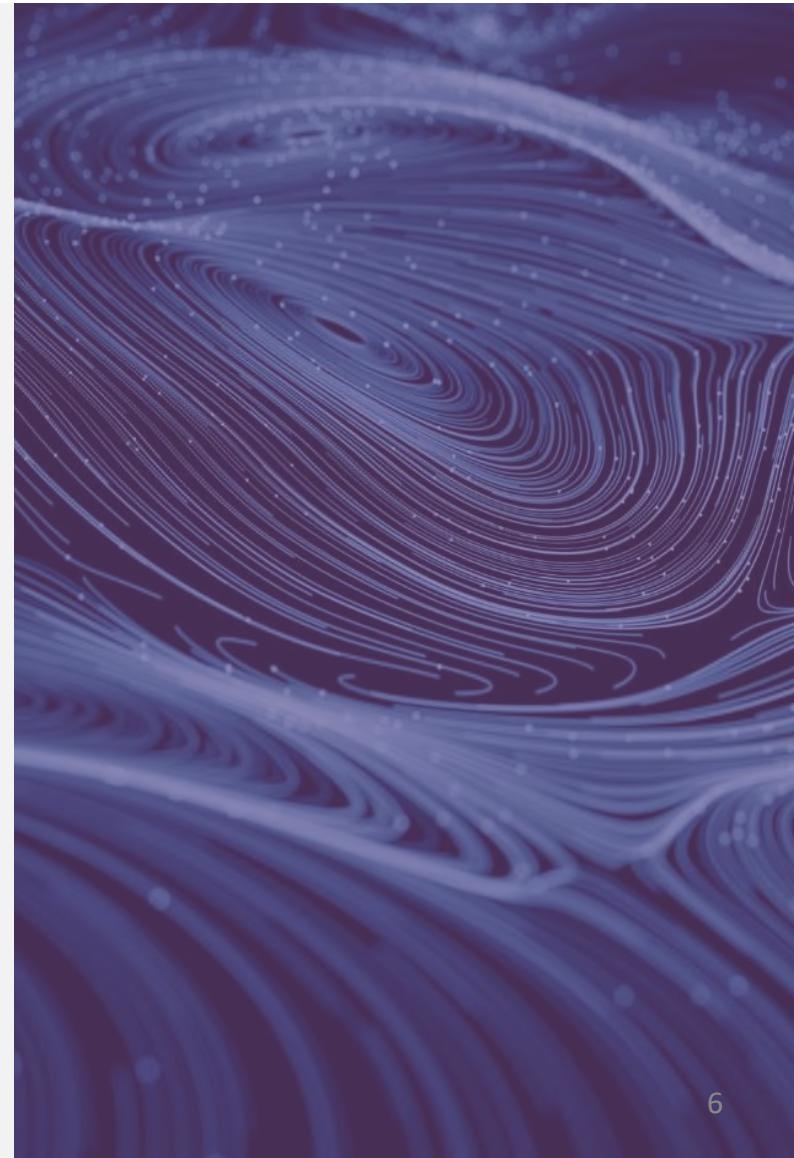
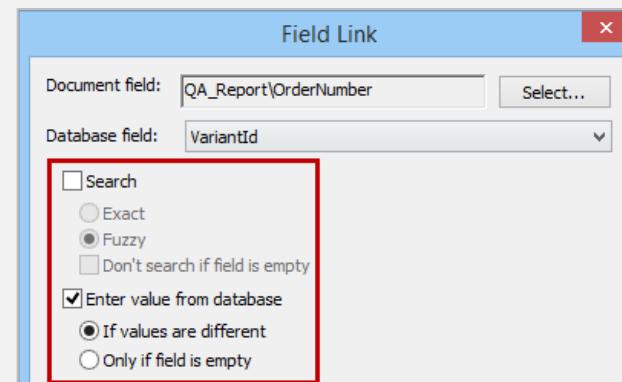
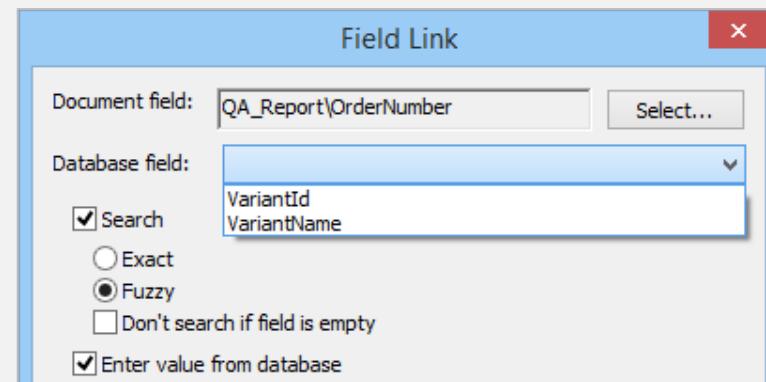
1. Check if the value is present in the database by a fuzzy search
2. If value is detected it is *substituted* by a database value
  - value normalization



# Database Lookup with Values' Retrieval

ABBYY®

1. Check if the value is present in the database
2. If value is detected, additional values are retrieved from the DB table
3. At least two links are required:
  - *Exact* or *fuzzy* comparison - to detect the DB string
  - Uncheck *Search* and check *Enter value from database* – to retrieve the value



# SCRIPT RULES

---



Document validation and fields normalization based on the values of one or more fields

## In the event of error:

- Set an error flag
- Add error description
- Specify focus field
- Propose replacement variants

# Script Rule Settings

ABBYY®

## List of Available fields

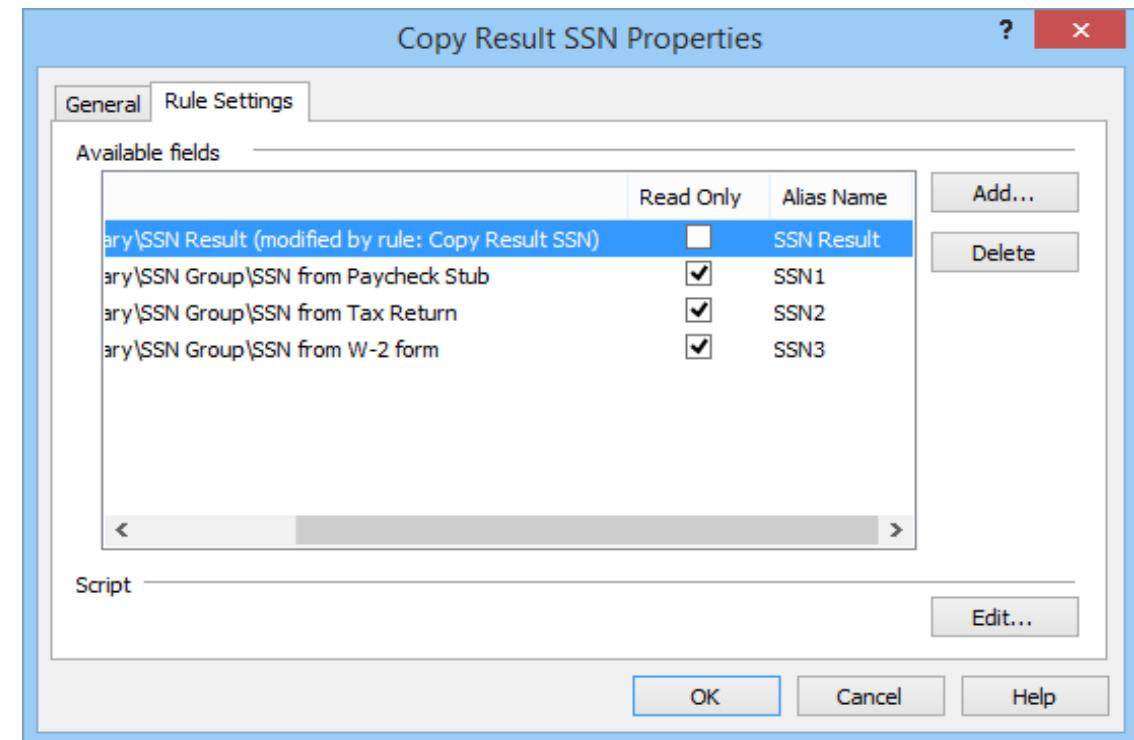
- fields linked by the rule must be added

## Read only option

- must be unchecked if the field value is modified by the rule

## Alias Names

- can be used to shorten the fields' names in script code

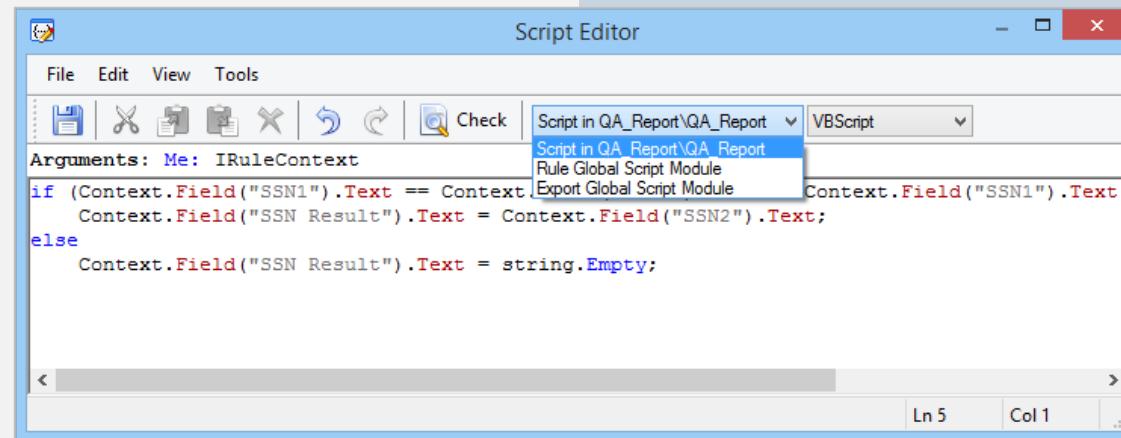


You can view (and copy) the code of any script rule specified in the Document Definition

- Speeds up the creation of script rules

Global script modules: ability to define procedures used by several scripts in one global module

- Global export module
- Global rule module



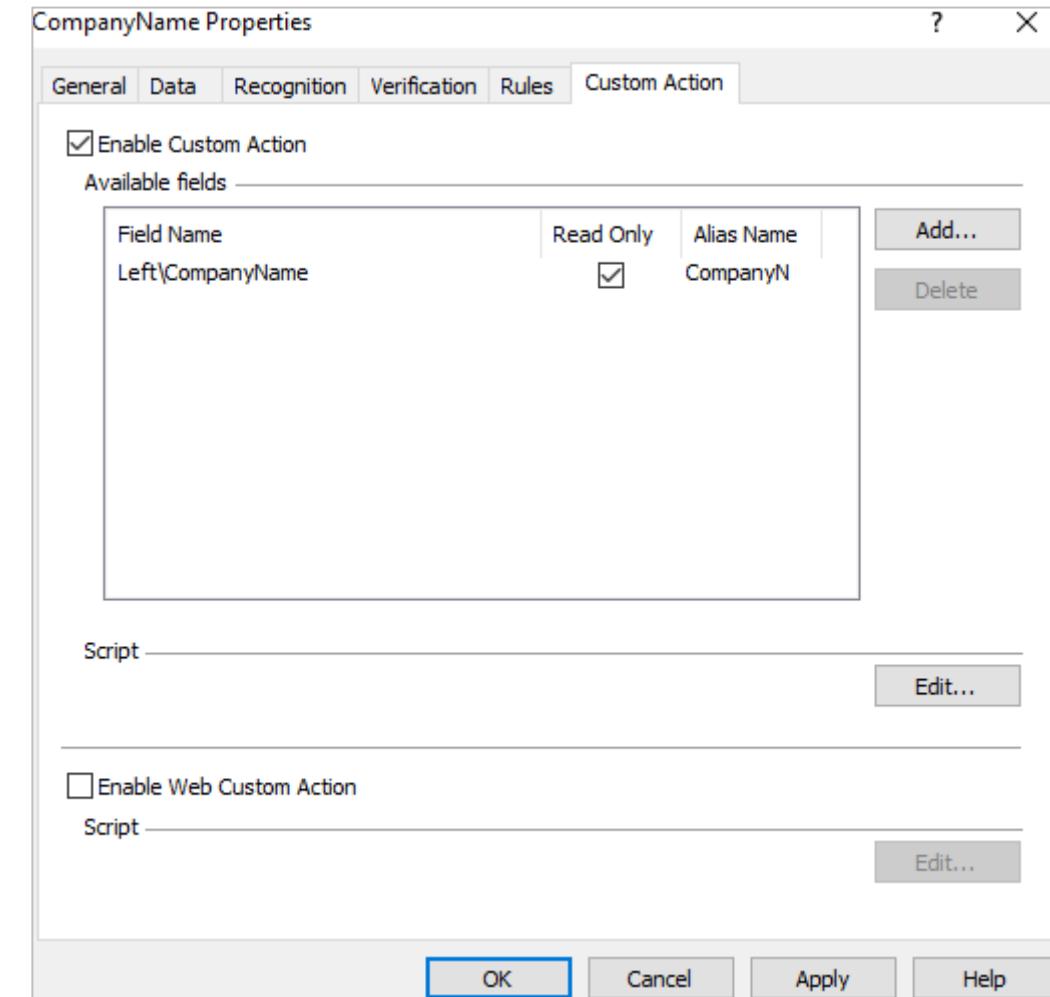
# Custom Action Script

ABBYY®

Additional script intended for the field modification

Applied to the field value only on the Operator demand

Web Custom Action is available



---

Database check rule offers wide possibility to check the value against the database, substitute values and retrieve the additional fields

---

Script rules allow to automatically run document checking and data modification of any complexity

---

Custom action script allows running field modification script on user demand



## ABBYY FLEXICAPTURE

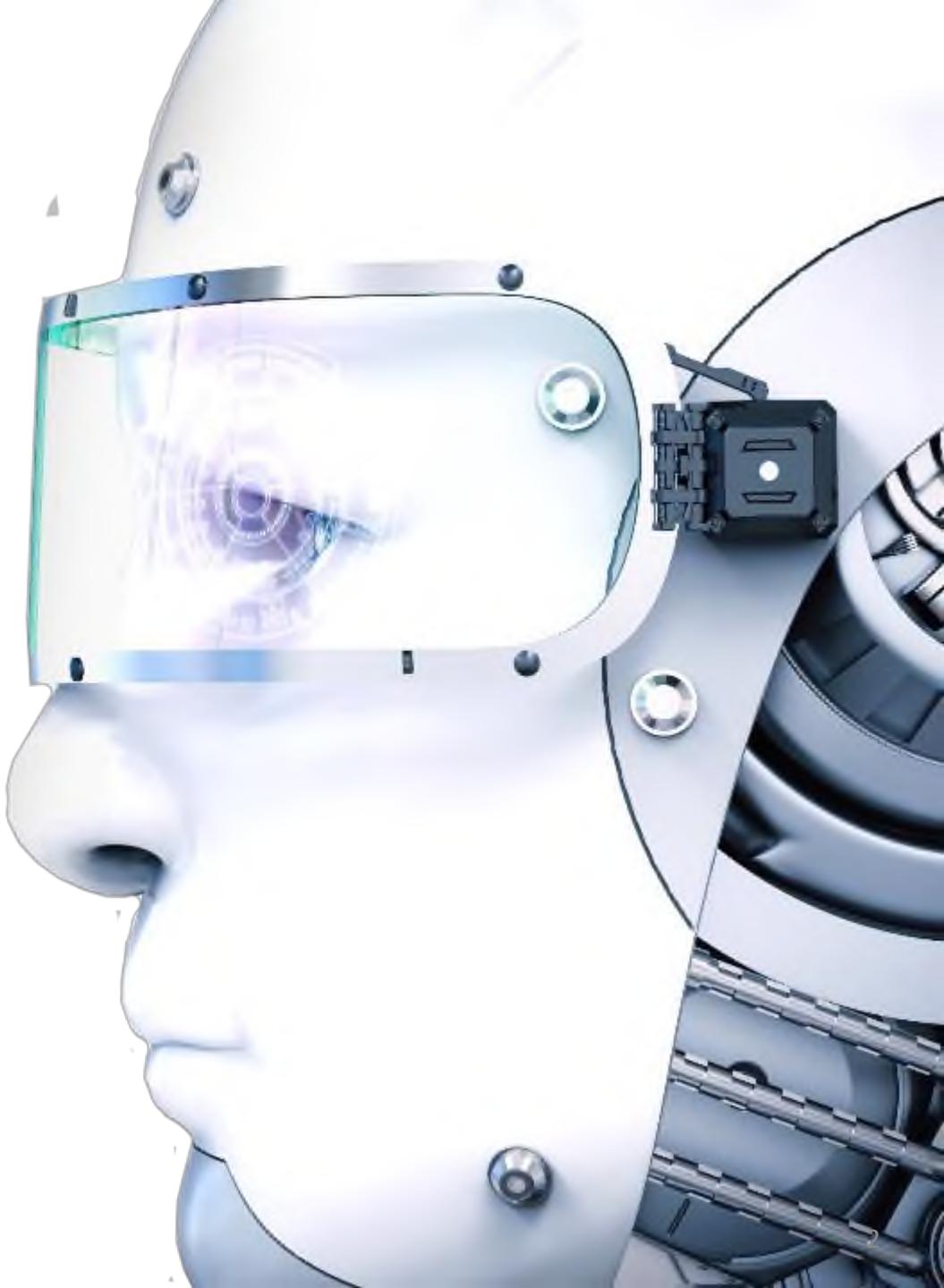
---

### DOCUMENT STRUCTURE

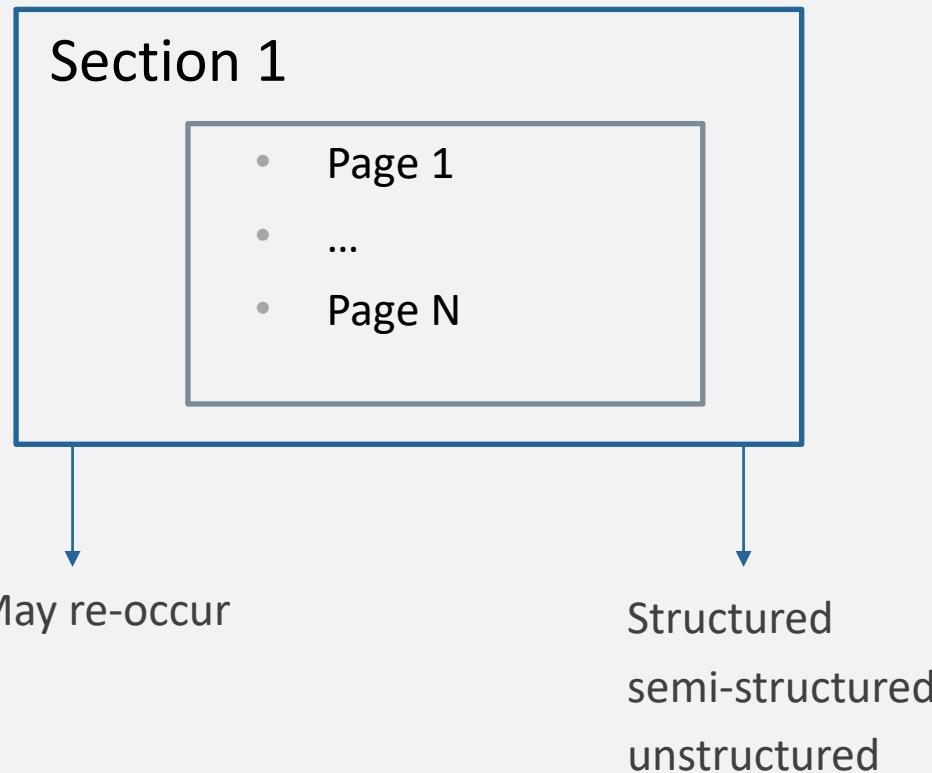


# Overview

- Processing multi-page documents
- Document Definition
  - Section
  - Page
- Assembly options
  - Order of sections
  - Number of repetitions
  - Annex pages
  - Checking key fields



## Hierarchy of Document Definition:



Section structure allows to process documents

- With repeating pages or sets of pages
- With optional pages
- With fields whose regions continue over several pages
- With annex pages

# Using Sections

A one-page document consists of one-page section

*Section 1 – page A*

The form contains fields for First Name, Last Name, Middle Name, Social Security Number, Phone Area/Building/Address Line 1, and a note to 'PLEASE ISSUE BANKING CARD FOR: IF NO'. It also includes sections for Address, Banking Card Type, and Cardholder Information, along with a barcode at the bottom.

Several pages of one document arranged in a particular order without any repetitions can be placed into the same section

*E.g.: Section 1 – page A, page B, page C*

*Assembly : ABC*

The three pages show the document structure as it would appear if it were split into three separate pages. Each page contains identical sets of fields for personal information, card details, and cardholder information, reflecting the assembly concept where multiple pages of the same document are combined into a single section.

## Example 1

- *Section 1 – page A (occurs once)*
- *Section 2 – page B (occurs twice)*
- *Section 3 – page C (occurs once)*
- *Assembly: A-BB-C*

## Example 2

- *Section 1 – page A (occurs once or twice)*
- *Section 2 – page B, page C (occurs twice)*
- *Assembly : A-BC-BC or AA-BC-BC*

A section may  
be a repeating  
element

# Multi-Page Section Example

ABBYY®

Fields with regions that continue onto the next page

E.g. table in a bill of lading

Part of the document (set of pages) may re-occur

E.g. form may re-occur from 1 to N times depending on the filing company

# Multi-Page Section Example

ABBYY®

## “Subsidy” document

Subsidy for Children Application Form

Subsidy Amount

Child Personal Details (the page re-occurs from 1 to 5 times depending on the number of children in the family stated on page 2)

Subsidy for Children Application Form

ATTENTION PLEASE! This form should be filled in all BLACK or BLUE BLOCKS or in red pen with HANDWRITTEN CAPITAL LETTERS according to the following document: ABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890 Example:

DATE OF INFILL:  .  .

Parental information  
YOUR NAME:  
First Name  Middle Initial   
Last Name

DATE OF BIRTH:  .  .  FAMILY STATUS:  Married  Single

ADDRESS AND CONTACT PHONE:  
Street   
Apartment #  City/Town   
State  ZIP Code   
Phone #  (  )

YOUR SOCIAL SECURITY NUMBER (SSN)

If you have 'Married' family status please complete the below section on your spouse:  
First Name  Middle Initial   
Last Name

In this section please indicate if you have children:  
1.   
2.   
3.

Section 1:  
“Application”

Subsidy calculation

#	Name of the child	Amount of subsidy
01		\$ . . .
02		\$ . . .
03		\$ . . .
04		\$ . . .
05		\$ . . .

Section 2:  
“Subsidy Amount”

Child information

SSN  (as it was indicated on the page 2) SSN   
YOUR NAME:  
First Name  Middle Initial   
Last Name

DATE OF BIRTH:  .  .

SUBSIDY CATEGORY:  General  Extended  Special  Nominal

AMOUNT OF SUBSIDY: \$

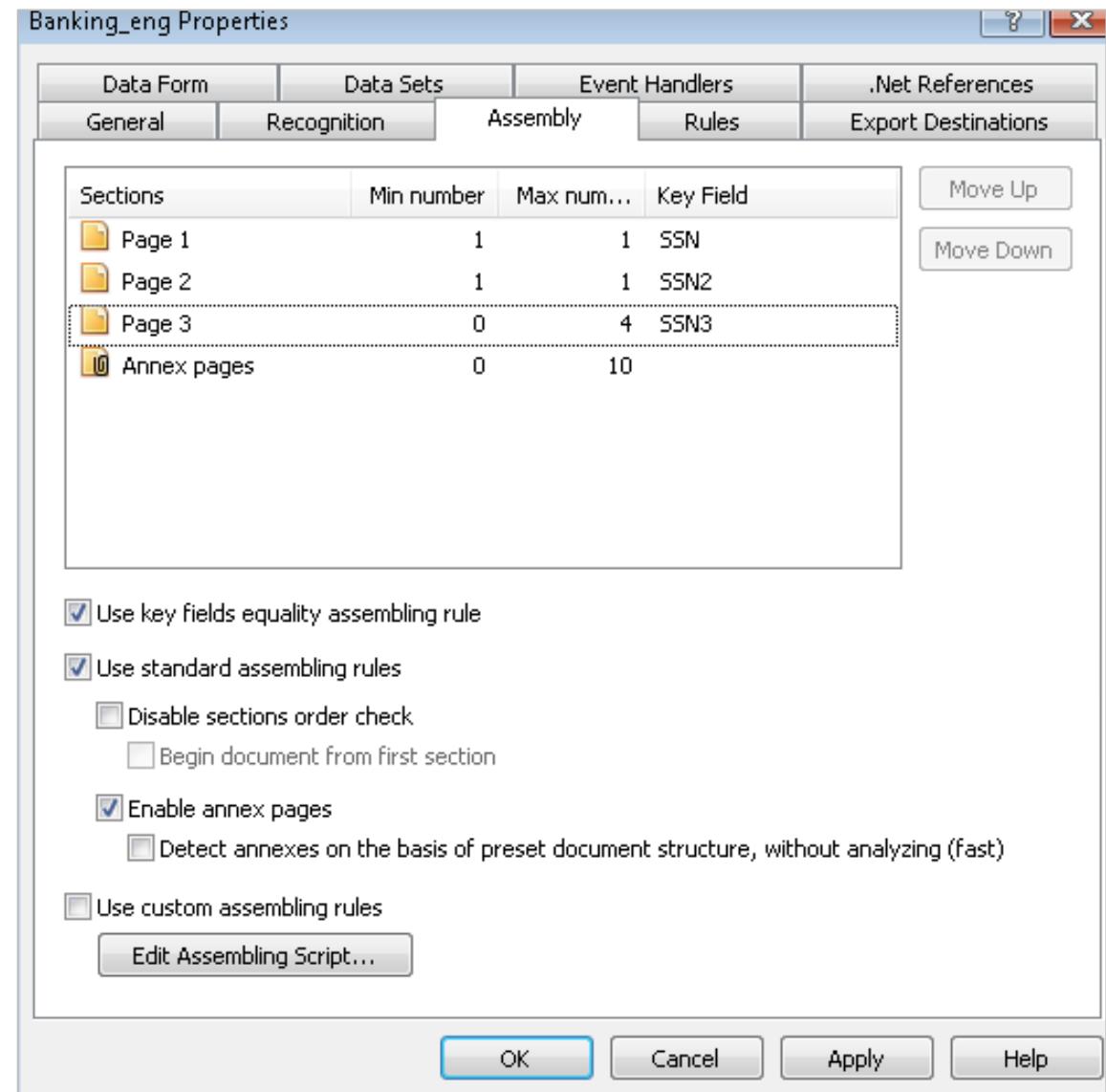
Section 3:  
“Child Personal Details”

# Multi-Page Section

ABBYY

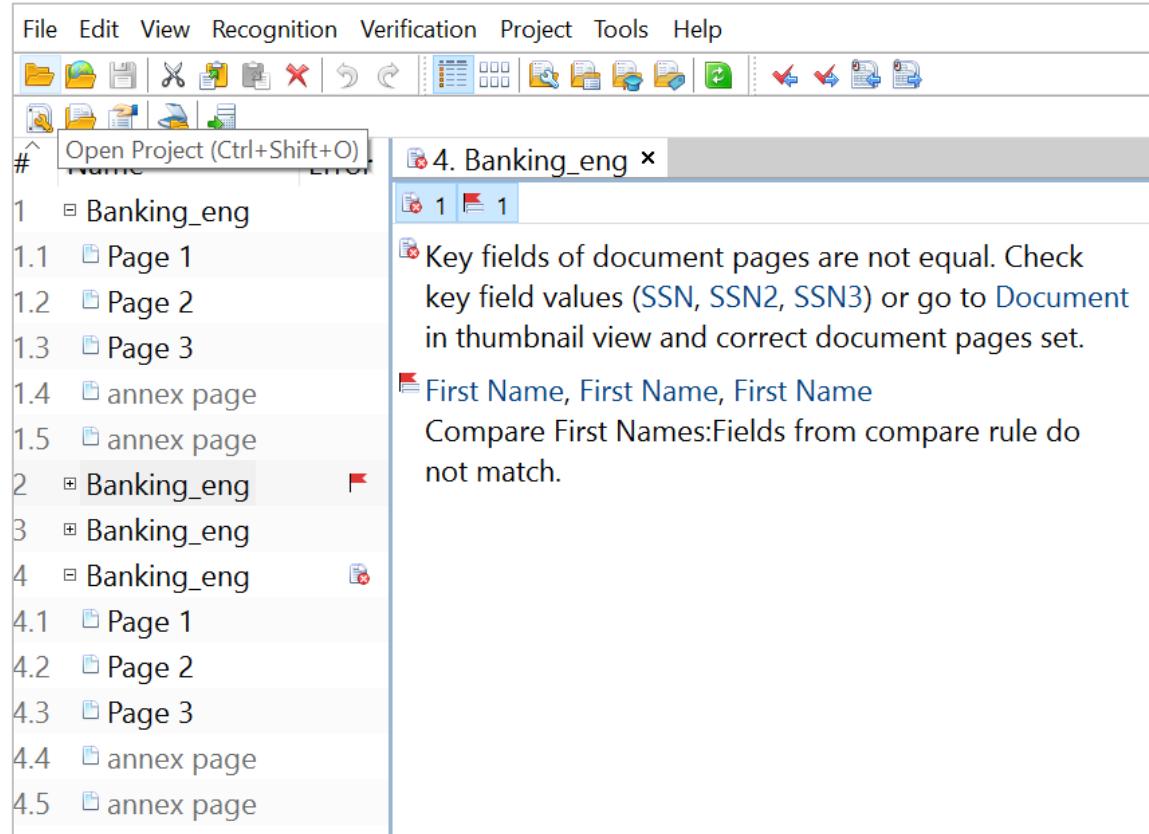
You need a multi-page section if:

- Part of the document contains fields with regions that continue onto the next page  
E.g. table in a bill of lading
- Part of the document (set of pages) may re-occur  
E.g. form may re-occur from 1 to N times depending on the filing company



# Document Assembly

ABBYY®



Pages are automatically classified and assembled into documents during recognition in accordance with the structure defined in the Document Definition

---

Hierarchy: Document Definition >  
Section > Page

---

A section is a logically selected part  
of a document

- Consists of at least one page
- May be one- or multi-page and  
described either in a FlexiLayout or a  
Document Definition

---

Document may include Annex  
(non-recognized) pages

---

Documents are assembled according  
to the document structure described  
in the definition

# ABBYY®

## ABBYY FLEXICAPTURE

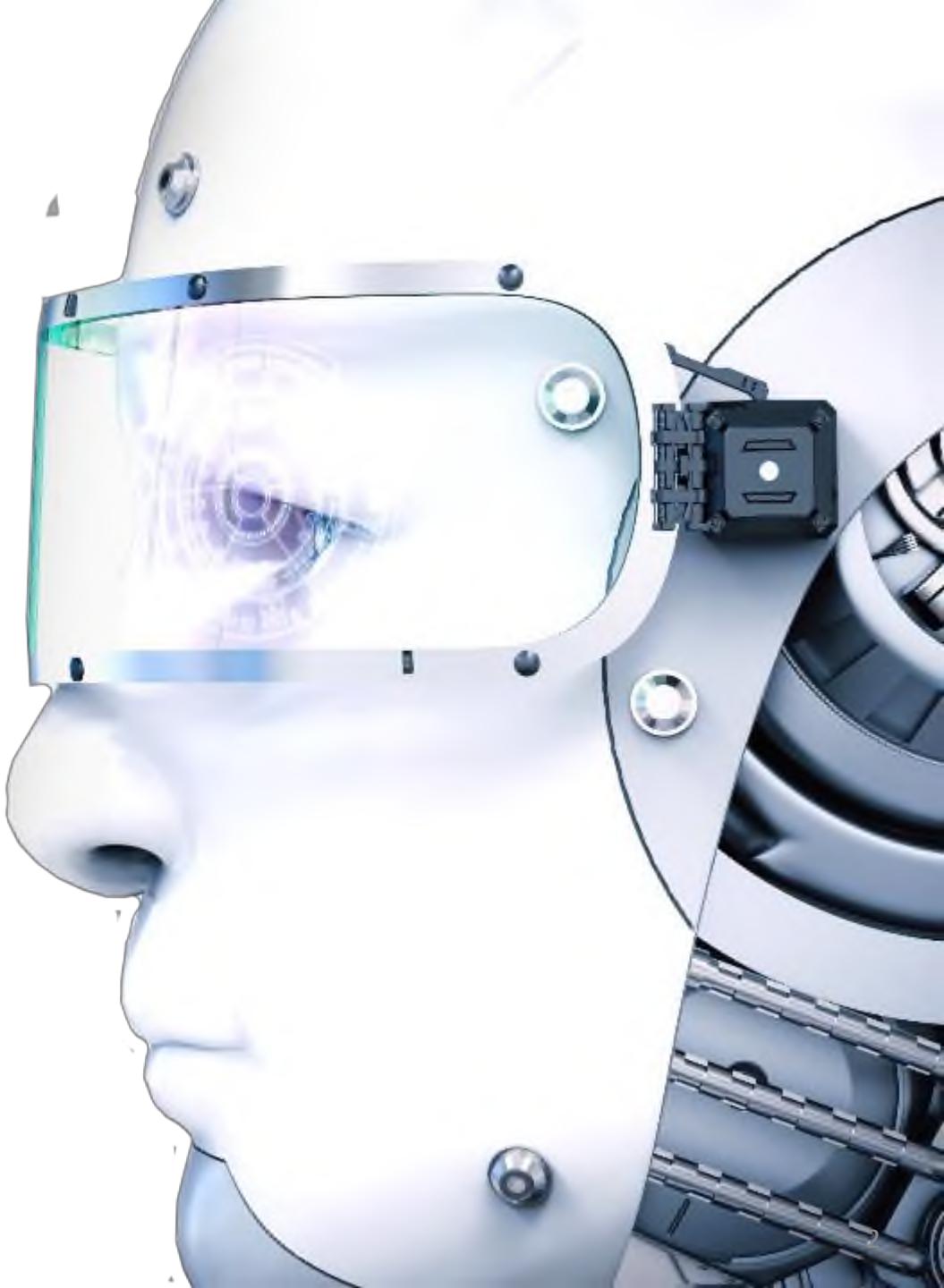
---

### SCANNING STATION



# Overview

- Batch types creation
- Batch types properties
- Custom Scripts
- Web Scanning Station



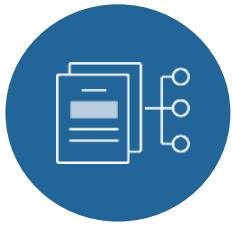
# BATCH TYPES CREATION

---

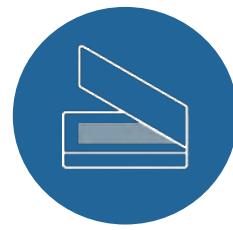


# Scanning Station. Main Workflow

ABBYY®



Create a new empty  
batch



Scan or add images  
to the batch



Manually pre-process  
images



Send batch

# Roles at Scanning Station

ABBYY®



## ADMINISTRATOR

*Creates Batch Types*

Exports Batch Types to Server  
(or saves locally)



## OPERATOR

Imports Batch Types or receives them automatically

Optionally can edit Batch Types

Scans batches with documents using pre-defined Batch Type settings

# BATCH TYPES PROPERTIES

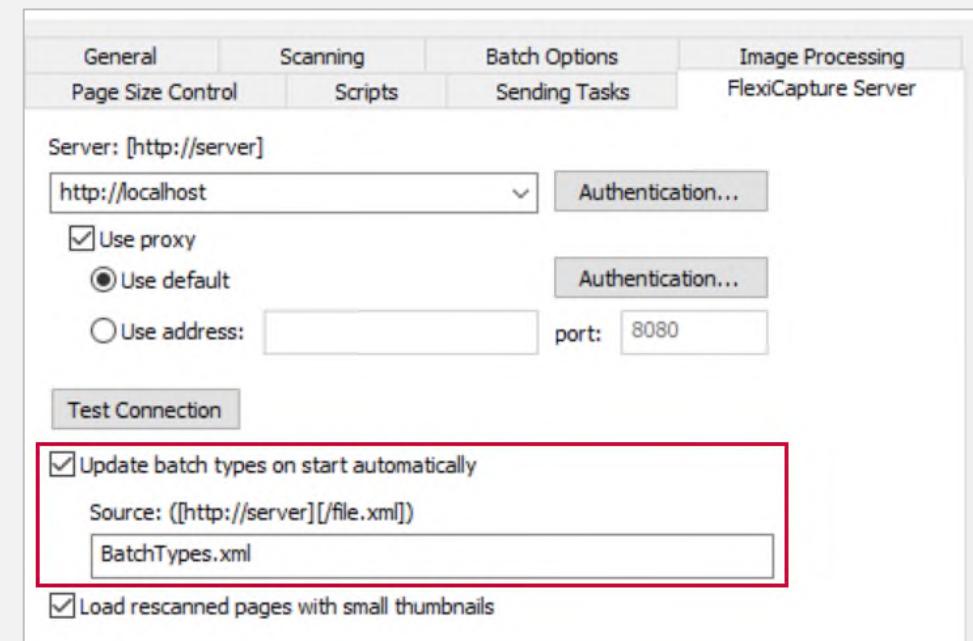
---



# Export and Import of Batch Types Settings

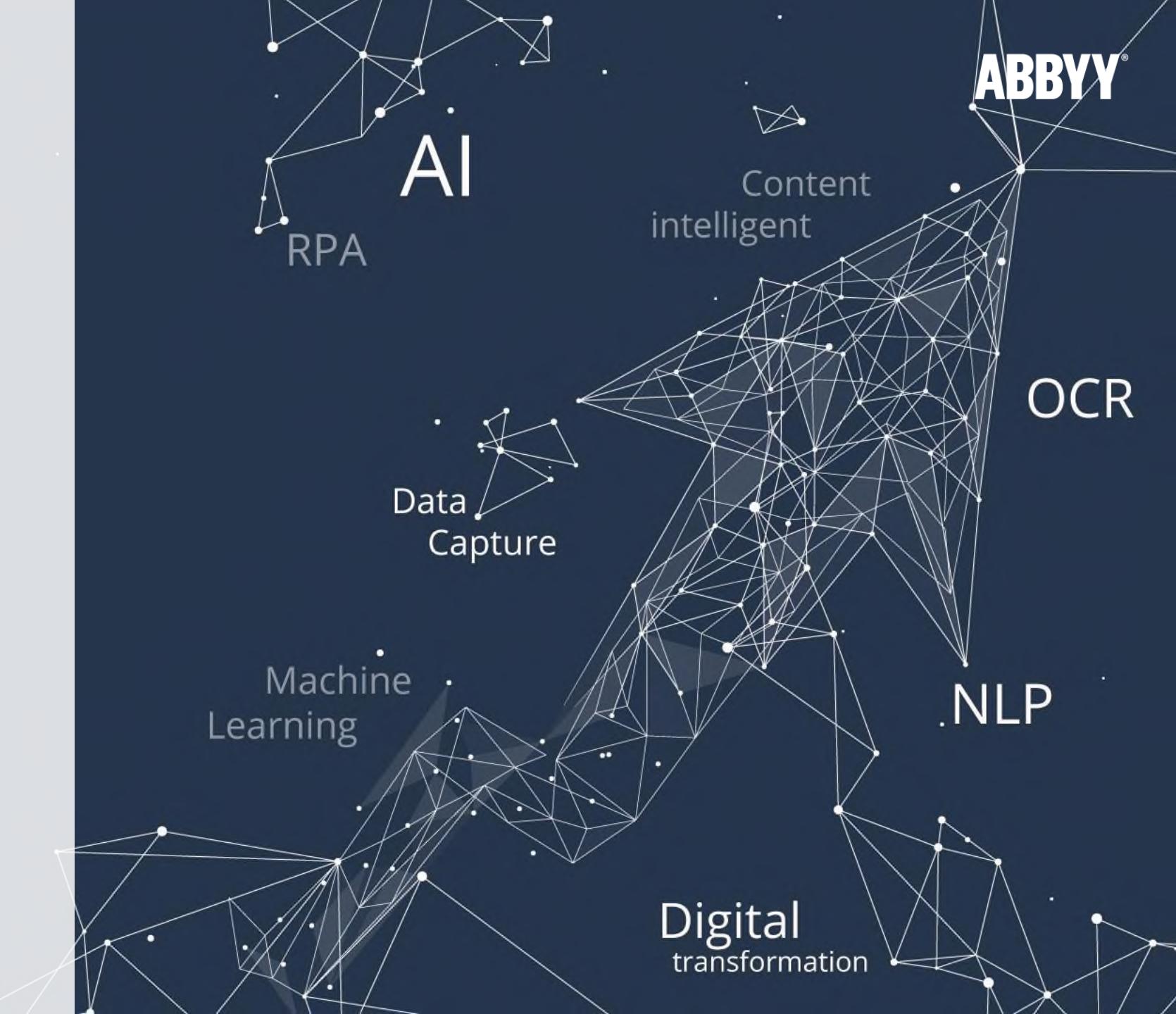
ABBYY®

- Batch types can be saved locally or on the Server as an XML file
- Settings can be imported:
  - manually
  - automatically if specified in station's options



# RESCANNING

---



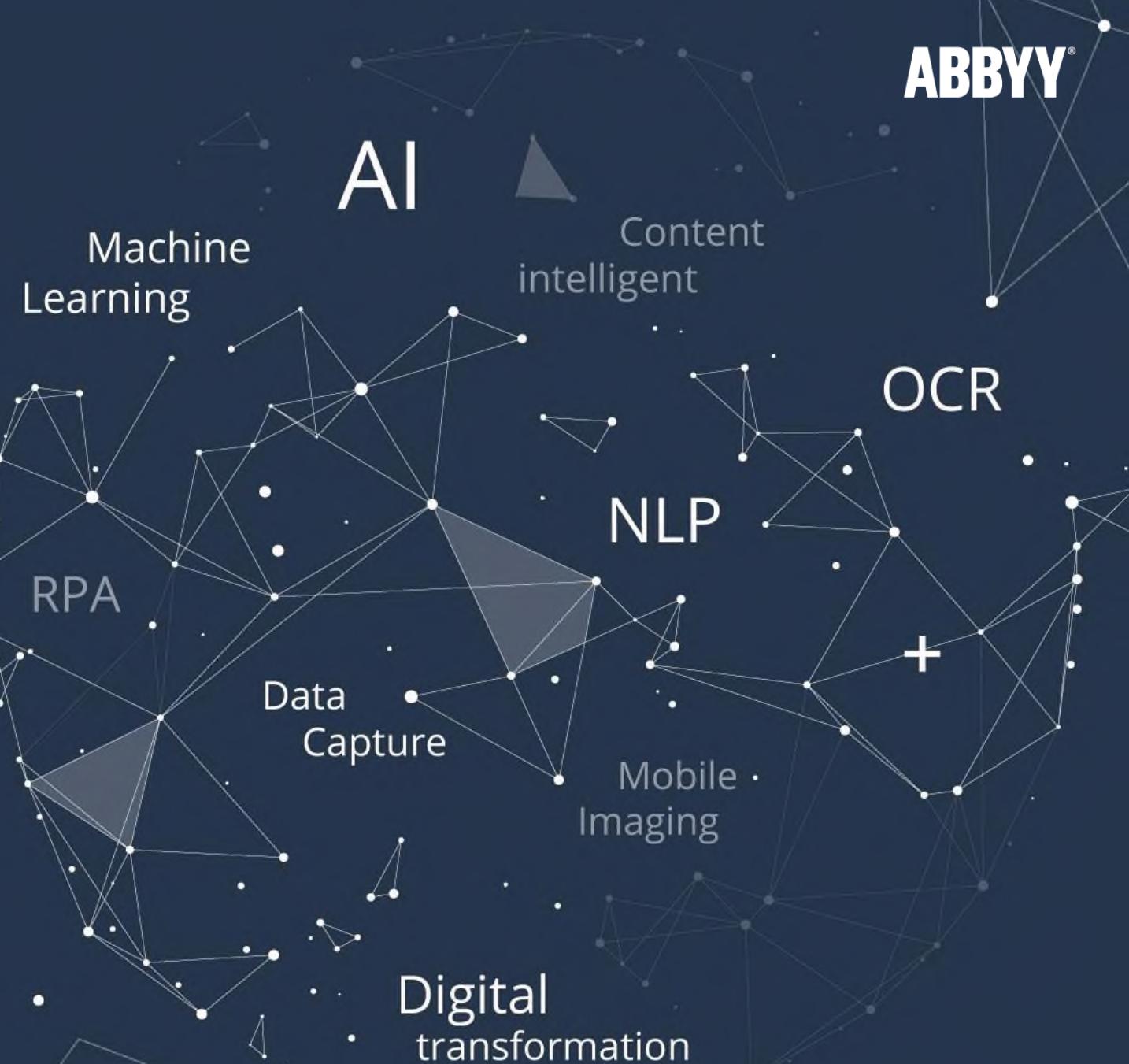
- 1 Documents arrived to the Verification station with poor recognition results due to bad scanning quality
- 2 Verification Operator sends documents to Rescanning
- 3 Scanning operator
  - Receives documents to be re-scanned
  - Completes the rescanning, sends back the task
- 4 Documents will be recognized and sent to verification again

Available for:

- Whole batch
- Separate documents

## CUSTOM SCRIPTS

---

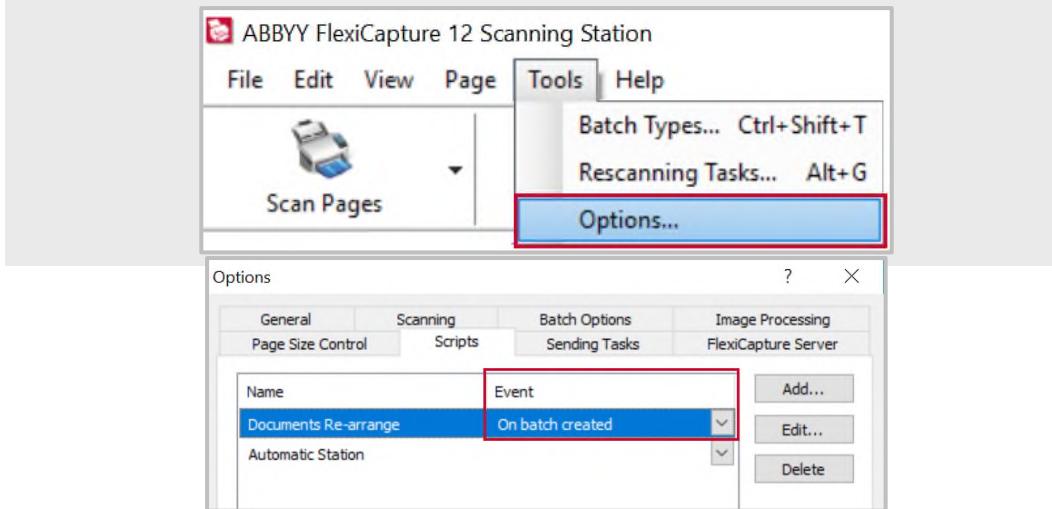


# Scripts on Scanning Station

ABBYY®

## Event Scripts

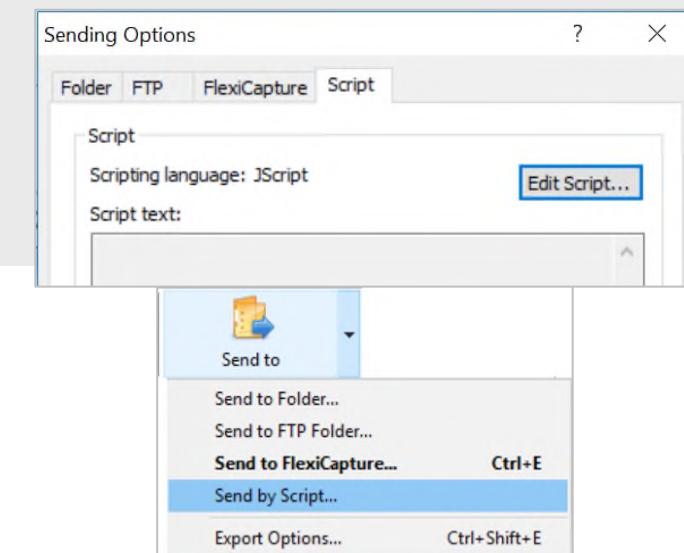
Create specific scenarios of working with Scanning Station



ABBYY FlexiCapture. Scanning Station

## Sending Scripts

Custom algorithm of sending scanned batches



# WEB SCANNING STATION

---



# Web Scanning Station

ABBYY®

## Scan images

- With TWAIN/ISIS/WIA compatible scanner
- Unified UI to setup any scanner/native UI
- Centralized predefined scanning settings

## Enhance images

- Automatic enhancement  
Deskew, Despeckle
- Manual editing  
Crop, Reduction, Rotation, etc.

## Add files

- All image file formats supported
- PDF support

## Document assembly

- Automatically by white papers, barcodes
- Automatically by scanning sessions
- Manually

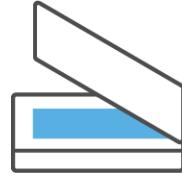
## Registration parameters

- Fill-in manually batch/document parameters

## Send batches

- Instant sending or on schedule
- Automatic re-sending and continue of sending in case of failure

# Summary



Batch Types  
are predefined  
scanning batch profiles



When batch is exported  
to a server:

- The project, batch type,  
task priority, and the  
registration parameters  
are defined



Scheduled sending of  
batches can be enabled



Custom scripts  
can be used in order to  
create specific scenarios  
of working with  
Scanning Station



## ABBYY FLEXICAPTURE

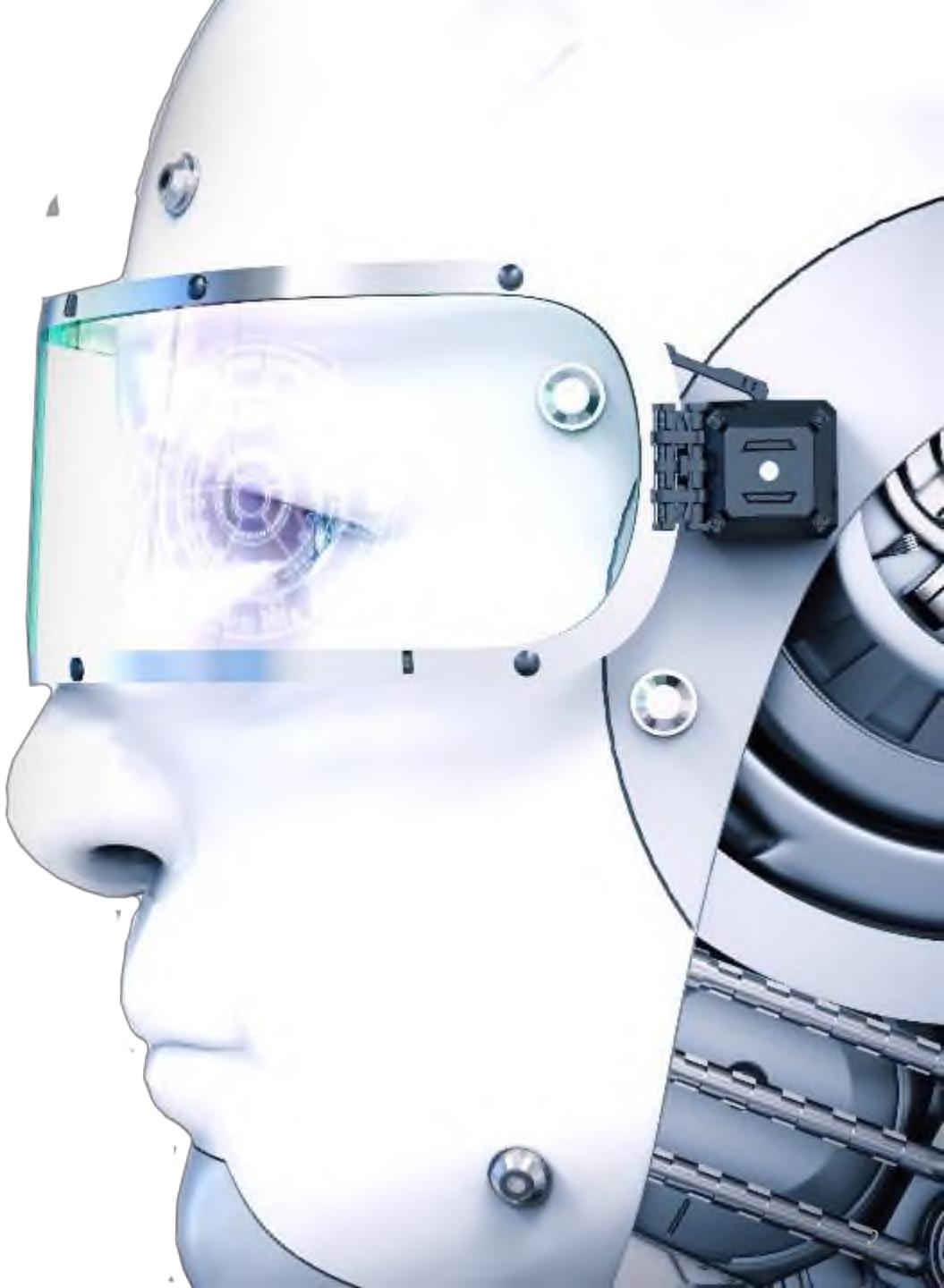
---

### IMPORT PROFILES CREATION

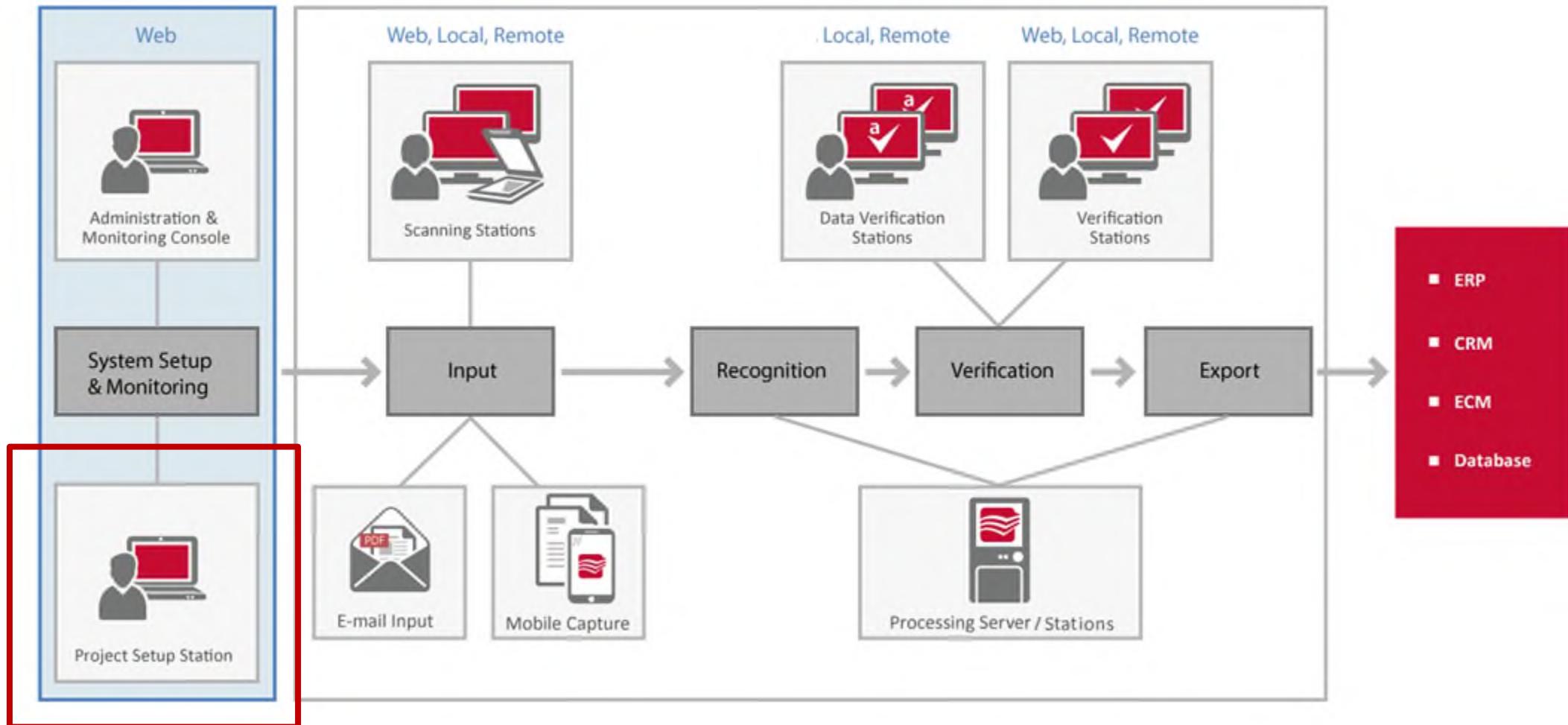


# Overview

- Image sources
  - Hot Folder Import Profile
  - Scanning Profile
- Hot Folder
  - Import profile creation
  - How to enable



# General Workflow





## Scanning of paper documents

### Scanning Station

- Add images with a scanner using pre-defined settings  
(scanning batch type)
- Import images from folder using pre-defined settings  
(scanning batch type)

## Automatic import of image files - Hot Folder

### Automatic images loading

FTP



IMAP



Folders



SFTP



Mailboxes



- MAPI
- POP3

Hot Folder is a storage  
of images to be checked  
by the system

Images in the folder  
are automatically imported  
to FlexiCapture batch

Several import profiles  
can be created

from different storages,  
for different batch types, etc.

Enable Hot Folders  
for the project

Administration & Monitoring  
Console

ID	Name	Hot folders
1	SingleEntryPoint	<input type="checkbox"/>
ID	Name	Hot folders
1	SingleEntryPoint	<input checked="" type="checkbox"/>

# Creating a Hot Folder Import Profile

ABBYY®

*Project Setup Station -> Image Import Profiles*

- Hot folder Setup
- Scanning profile (testing settings only)

Hot Folder image source:

FTP



IMAP



Folders



SFTP



Mailboxes



- MAPI
- POP3

# IMPORT FROM MAIL

---



## How to organize the processing workflow:

- Documents are scanned onsite
- Images are sent to a dedicated e-mail address
- Processing Station automatically checks the mailbox
- Image attachments or/and message bodies are imported
- Batches are created automatically and sent for recognition

*MAPI* – interface for import from Microsoft Exchange Mailbox

*POP3* – protocol for local e-mail clients to receive mails from remote servers

Special mailbox should be created for importing images

- With a specific user account e.g. *FlexiCaptureImportUser*

Processing Stations that perform import should be launched under this user account

# Import from email: Sender's address and subject

ABBYY®

It is possible to get sender's email address and subject via registration parameters

Named registration parameters will be filled in automatically during the import from email

Subject:  
fc\_Prefdefined>EmailSubject

E-mail:  
fc\_Prefdefined>EmailSender

This information can be used in routing scripts when export destination will change depending on a document sender

# Access to Files Attached to the Source PDF File

ABBYY®

Attachment of a PDF file can contain information about a document and fields values

Example: **ZUGFeRD** – a specific electronic format of invoices in Germany

Importing can be disabled via Windows Registry

# SCANNING PROFILE

---



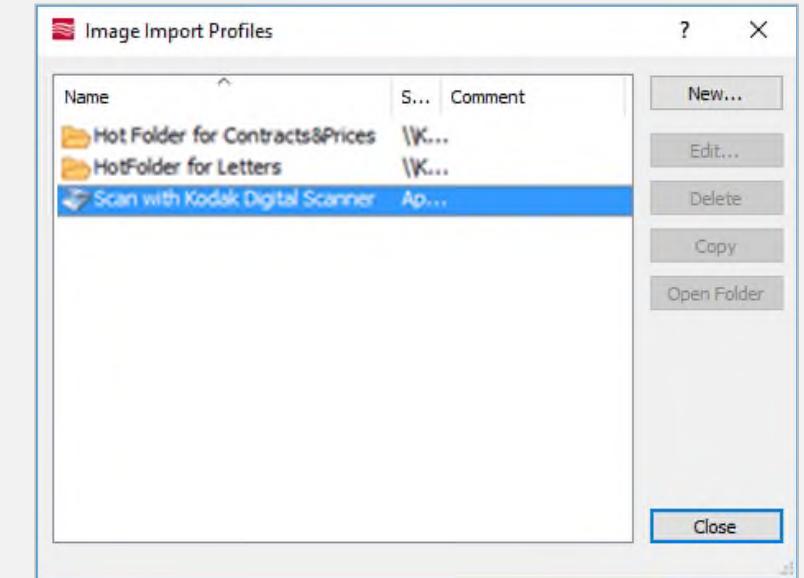
# Scanning Profile

ABBYY®

Used only for project debugging

Works only on the Project Setup Station

Several scanning profiles can be created



## Image sources

### **Hot Folder Import Profile**

- Created on the Project Setup Station
- Needed to be enabled on the Administration & Monitoring Console
- Sources of images – folders, ftp, sftp, imap, mailbox



### **Scanning import profile**

is used locally for testing scanner  
on Project Setup Station

# ABBYY®

## ABBYY FLEXICAPTURE

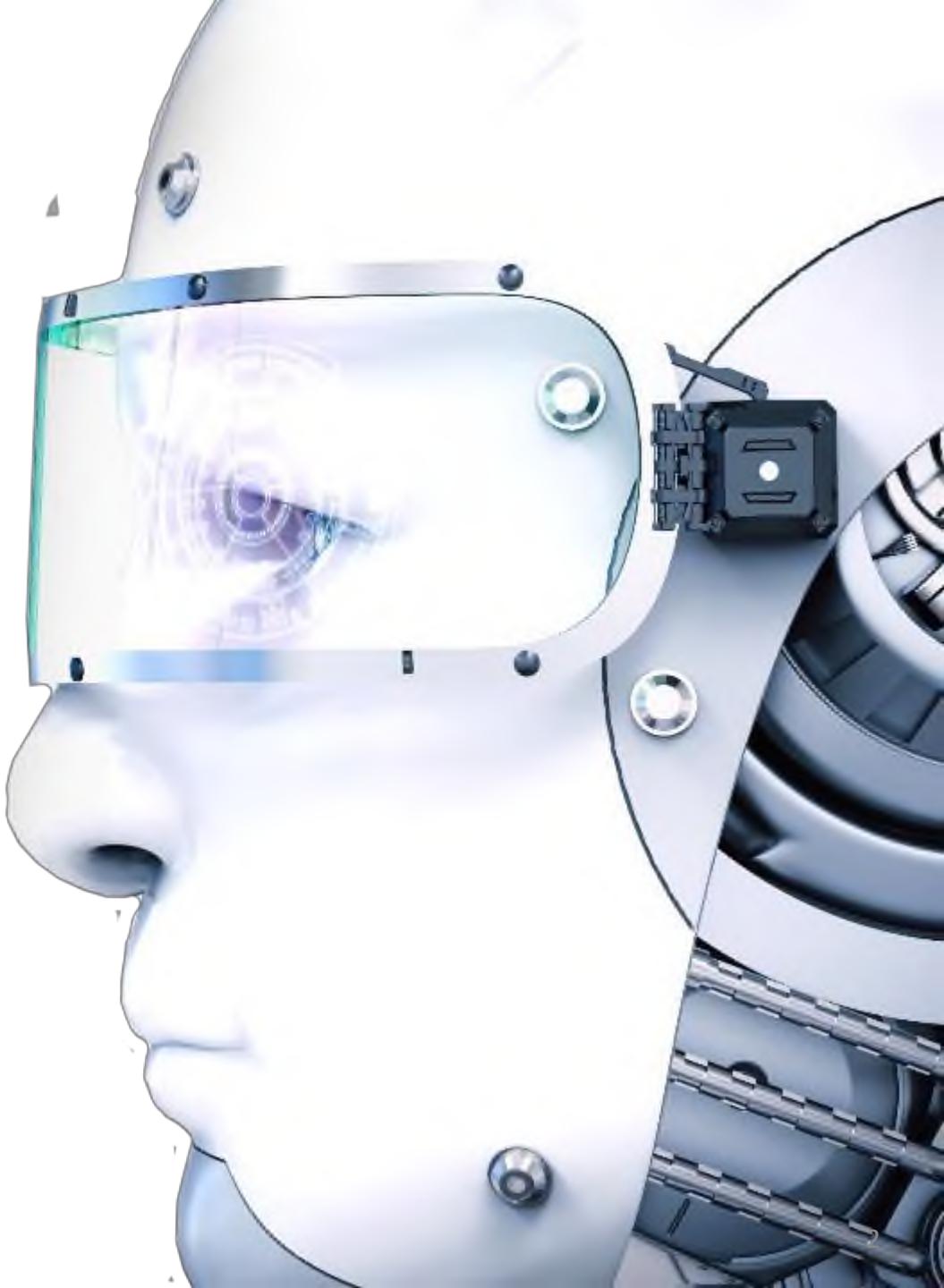
---

VERIFICATION



# Overview

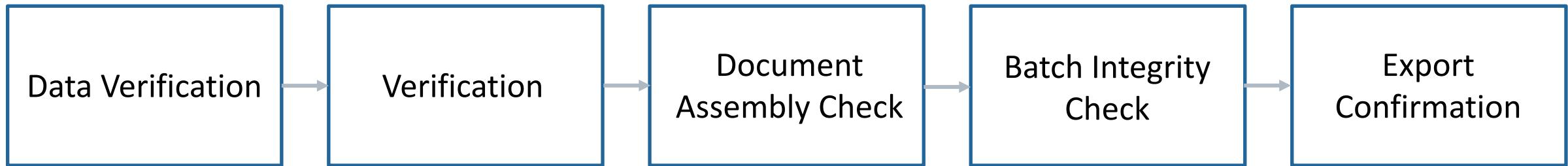
- Stages, Tasks, Queues
- Data Verification Settings
  - Verification in Document Window
  - Group Verification
  - Field Verification
- Web Verification Station



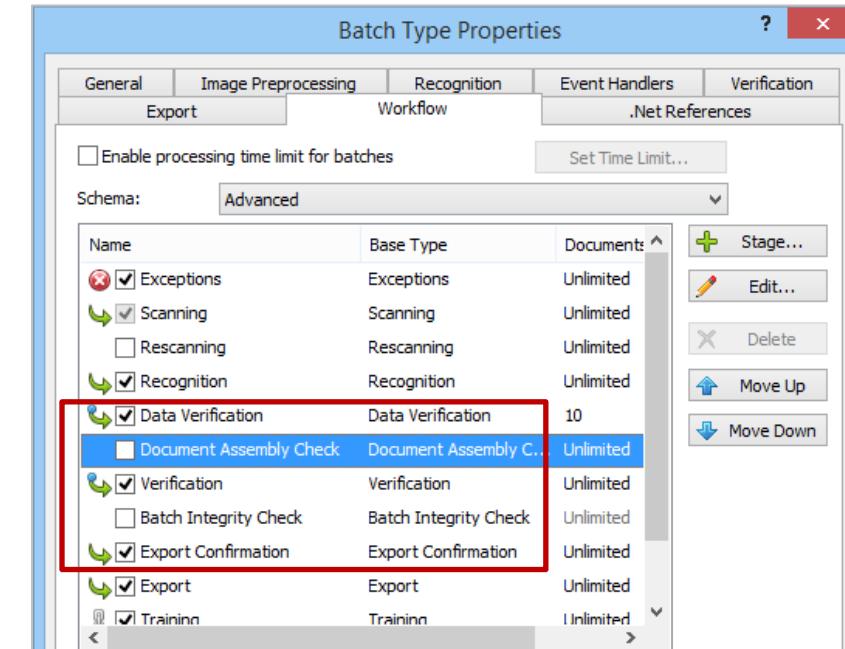
# Verification Stages & Queues

ABBYY®

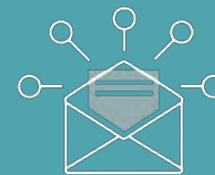
5 default queues for verification



Queue	Max Pri...	Personal...	Formed...
<Any>	Normal		
Exceptions			
<input checked="" type="checkbox"/> Data Verification	Normal	2	
<input type="checkbox"/> Document Assembly Check	Normal	2	
<input checked="" type="checkbox"/> Verification	Normal	1	
<input type="checkbox"/> Batch Integrity Check	Normal	1	
<input type="checkbox"/> Export Confirmation	Normal	2	
			1

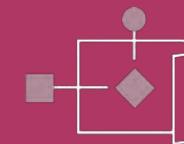


Batch can be sent to verification stage after data verification *by tasks* or in its *entirety*



BATCH

- Set of documents, scanned or imported together
- Can be divided to tasks

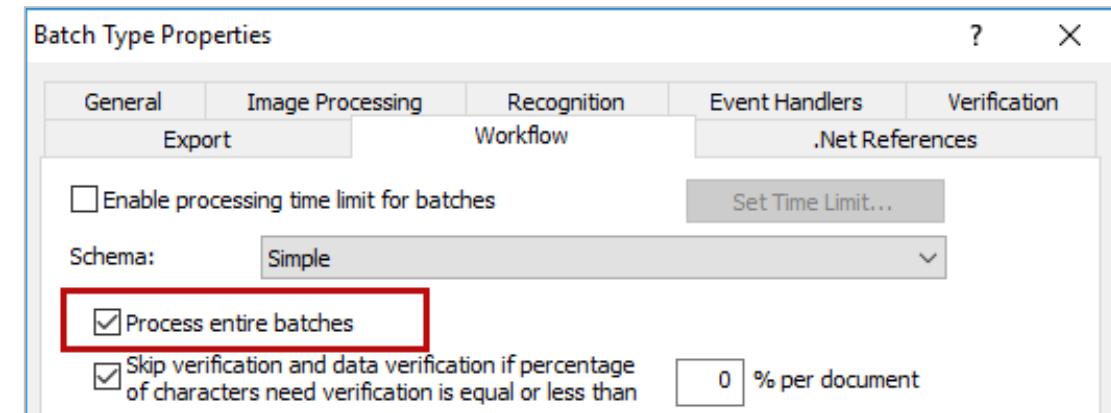
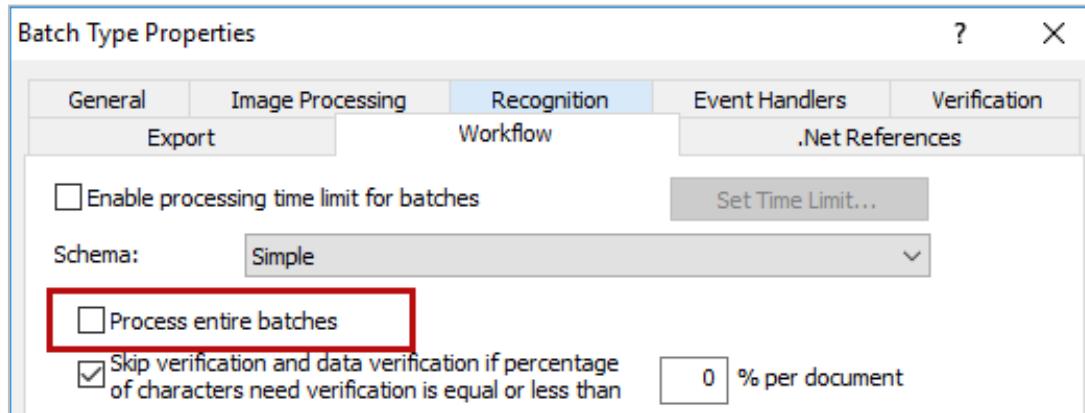


TASK

- Portion of batch documents to be processed together
- Can be created automatically or manually

# Whole Batch Processing

ABBYY®



- Batch is divided into verification tasks
- The task is sent to the *verification stage* and completed *separately*
- Tasks of one batch are kept in the background
- The task completes on the verification stage with all documents of the batch

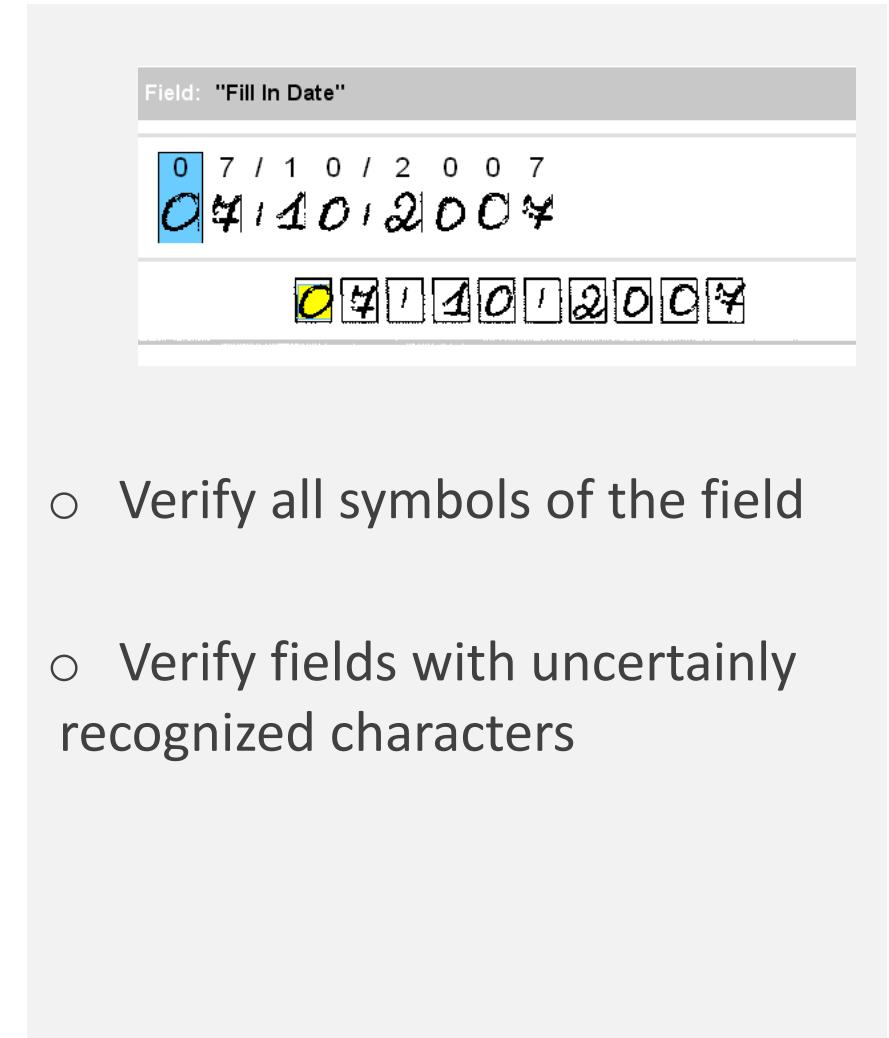
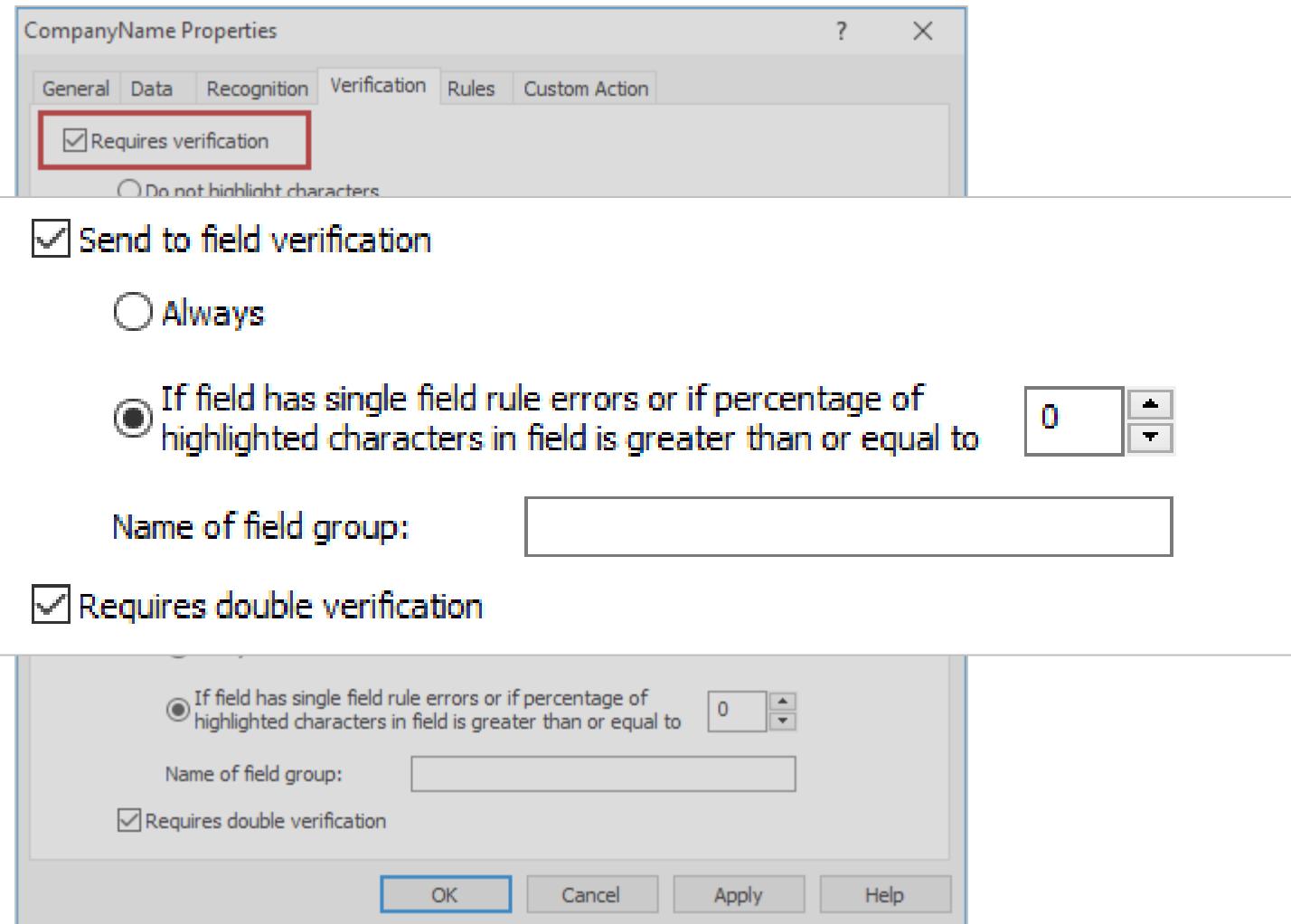
# DATA VERIFICATION SETTINGS

---



# Verification Settings

## Settings of each field in Document Definition Editor



# Double Verification

ABBYY®



Ensure the 100 %  
confidence in data  
correctness



The same data is checked  
by two different operators



Different results are sent  
to a 3rd person for final  
approval

Send to field verification

Always

If field has single field rule errors or if percentage of highlighted characters in field is greater than or equal to

Name of field group:

Requires double verification

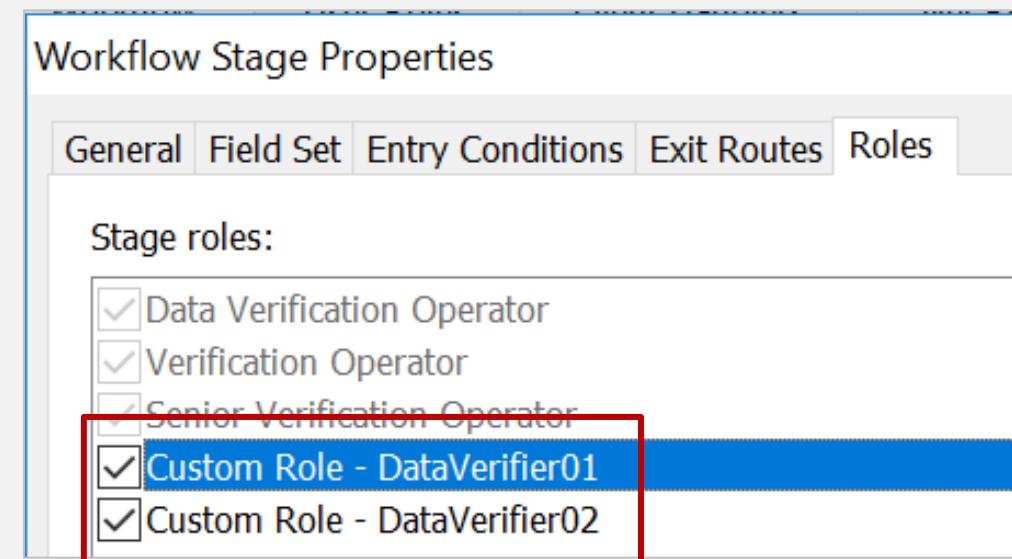
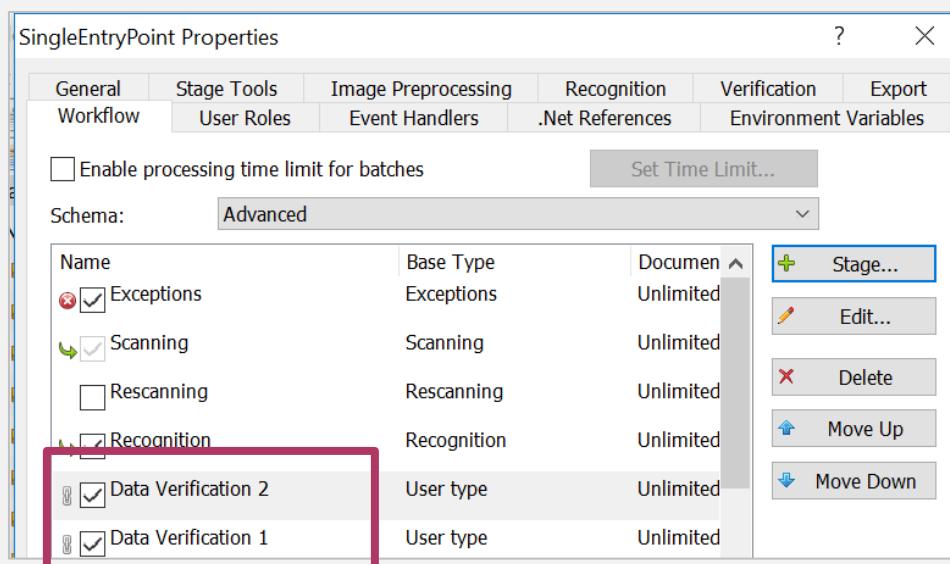
- In the properties of the field

- In the workflow settings

# Distributed Data Verification

ABBYY®

Special user roles for a customized workflow with additional data verification stages can be used



# Document Errors

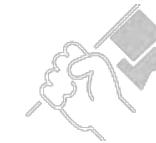
ABBYY®



## Field format check

- Date
- Number format
- Etc.

The field value is neither a date nor time.



## Validation rules

- DB check
- Relations between fields
- Sum calculation
- Etc.

Last Name, Last Name, Last Name

Compare Last Names:Fields from compare rule do not match.



## Document assembly check

- Order of pages
- Completeness

1. Banking\_eng

Order pages as follows: Page 1, Page 2, Page 3 + 2 annex pages  
Key values mismatch.



## Classification result check

- Class of page
- Class of document

# Document Errors Correction

ABBYY

Correct field value

Add missing pages:

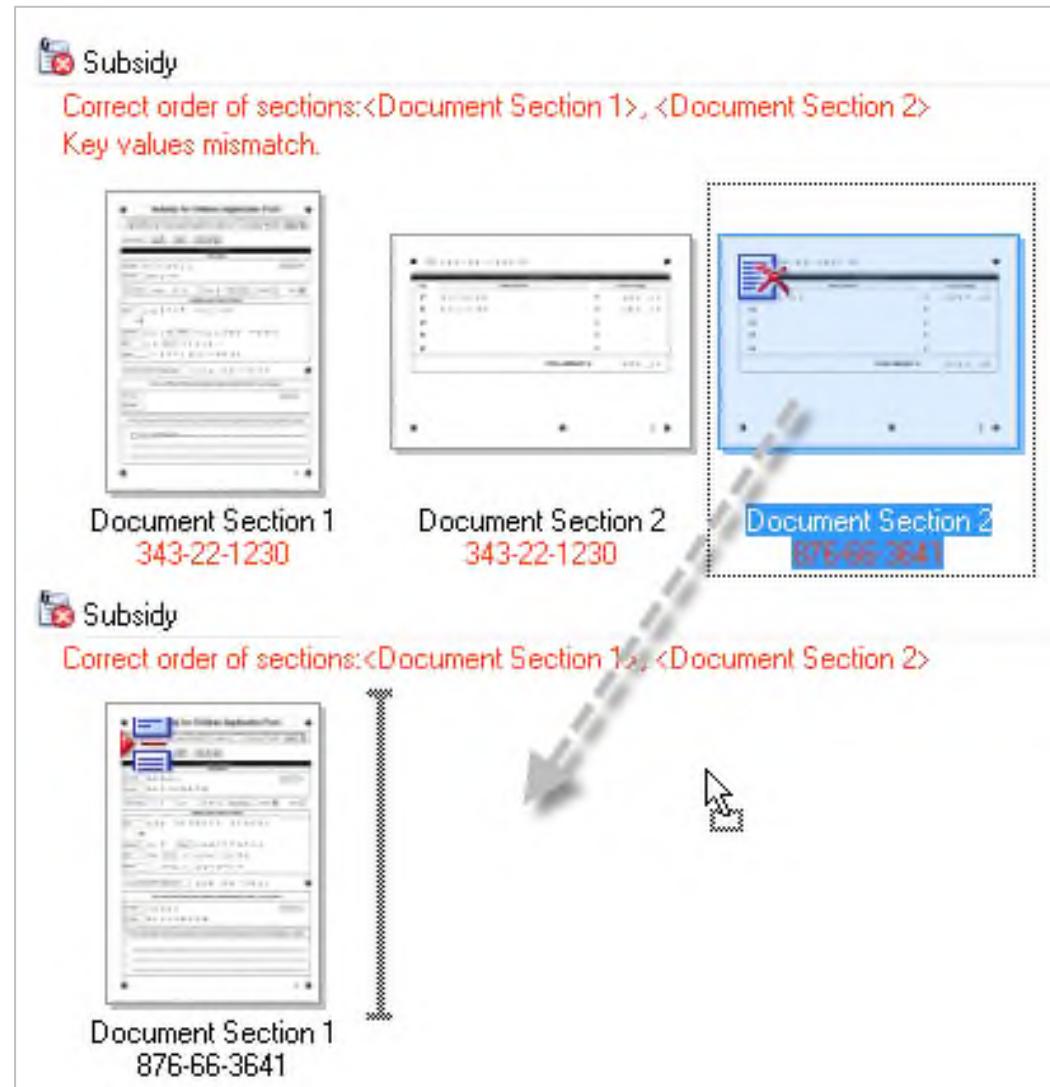
- Scan documents
- Load image files

If assembly error occurs:

Drag and drop page thumbnails  
in the batch with the mouse

If classification error occurs:

Confirm or change Document  
Definition



# WEB VERIFICATION STATION

---



## Check rules in browser

- For standard and most script rules

## Work with documents asynchronously

## Custom Data Form

- Includes all settings defined in the Document Definition

## Work with regions

- Highlighted field regions / group of regions / group of repeated fields
- Operator can draw a region
- Work with complex regions and regions with continuation

## Hot keys

- List is currently hardcoded and non-editable

## Line items and PO matching support

- Displayed as a table in the data form
- Operator can draw line items
- Continue/Reanalyze Line Items

## Workflow settings

- There are 5 standard queues for verification steps
- Verification queues are specified in the Batch Type workflow properties
- Batch can be divided into Tasks
- Tasks can be created automatically or manually

## Data verification settings

- Must be set for each field in Document Definition Editor
- Uncertainly recognized characters can be verified in Document Window and\or send to Group and\or Field verification

# ABBYY®

## ABBYY FLEXICAPTURE

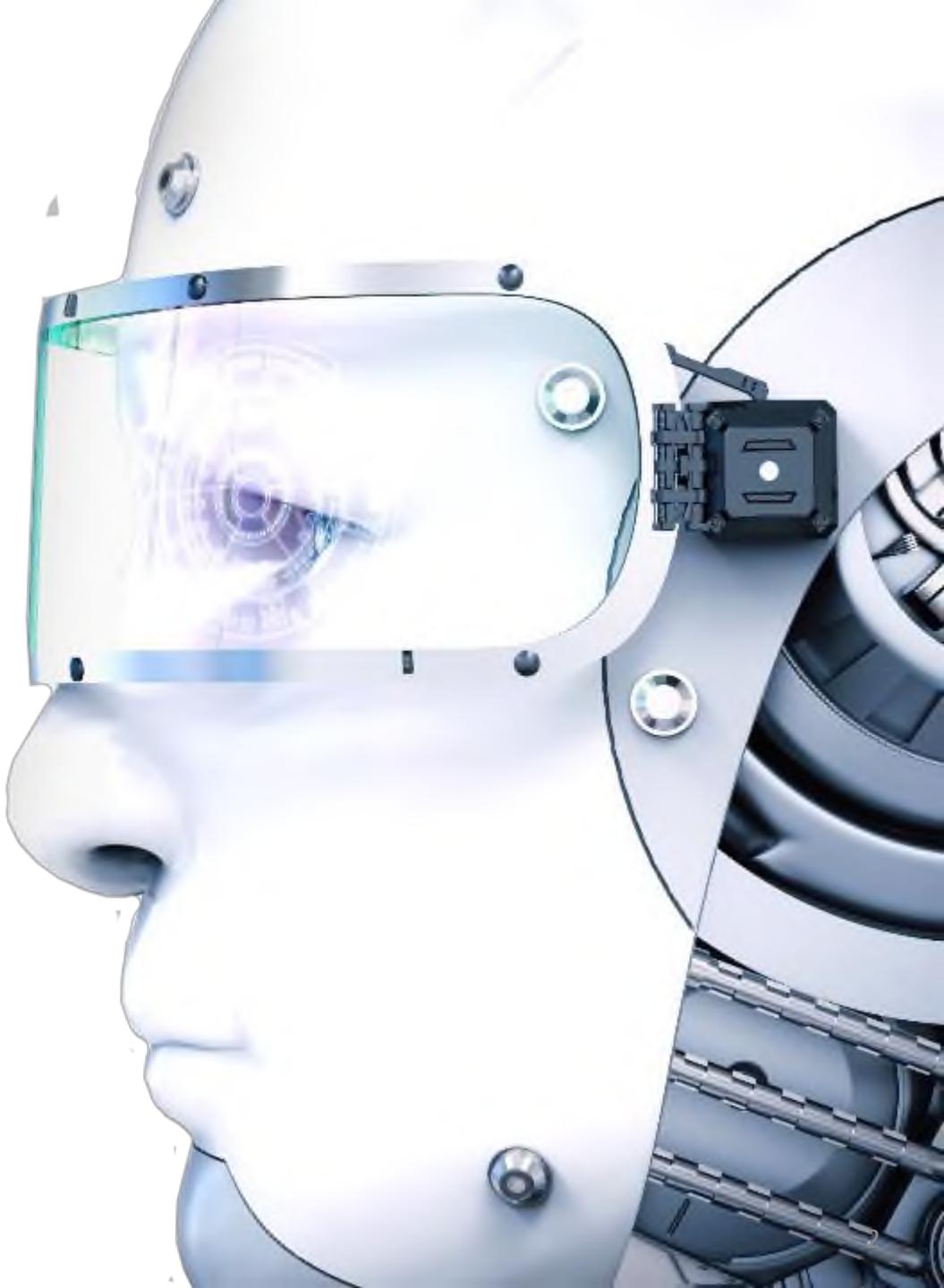
---

VERIFICATION  
TASK MANAGEMENT



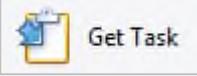
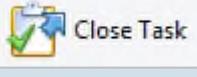
# Module Overview

- Tasks
- Task Management
  - Data Verification Operator activities
  - Verification Operator activities
  - Senior Verification Operator activities
- Monitoring



# Verification Task Management

ABBYY®

 Get Task	Get task	Perform a task from the queue
 Close task	Close task	Send task to the next stage after finish of processing
	Reject	Cancel a task either before or while doing it
	Postpone	Save the temporary results without returning the task into the queue and resume it later. Can be automatically returned to the queue after a certain time
	Mark for sending to exception	Send a task to the Exceptions queue if there is not enough information to verify and correct the data
	Mark for sending to rescanning	Send a task to rescanning if the recognition result is too low

# SENIOR VERIFICATION OPERATOR

---



## Switching between batches and tasks

- Batches view
- Queues view

## Sending batches, tasks, documents to stage

## Working with:

- Export Confirmation queue
- Exceptions queue
- Rescan queue

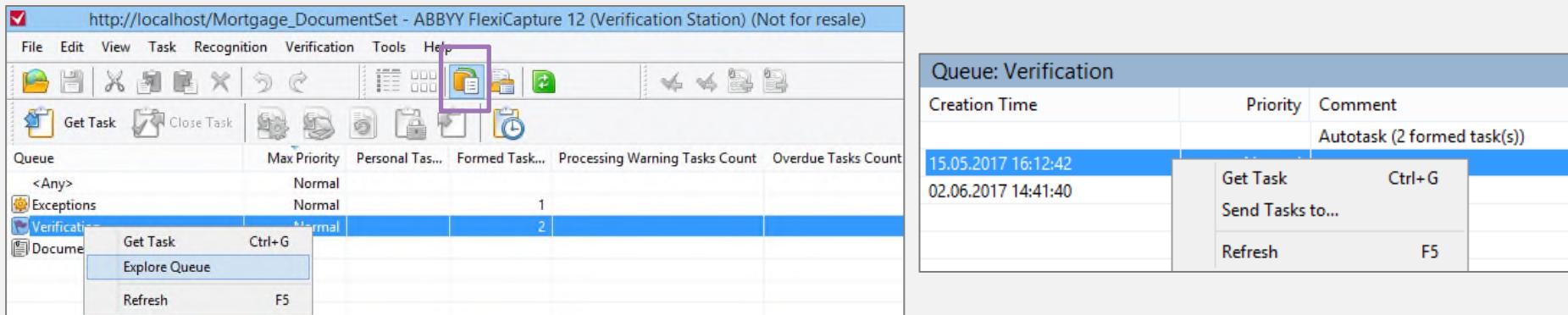
Batches can be displayed in their entirety (no tasks are displayed)

- View the current batch statuses
- Send batches for processing to a particular stage
- Apply filters to show particular batches

#	A...	Name	Type	Total Doc...	Total Pages	Stage	Creation Time	Expires on	Status	Confidence Level	Priority
5		14-392017-06-02	Default	0	0		02.06.2017 14:39:49		0% (0 of 0)	Normal	
4		19-042017-04-21	Default	3	27	Verification	15.05.2017 16:10:03		99% (1138 of 1149)	Normal	
3		19-042017-04-23	Default	3	27	Verification	24.04.2017 21:59:21		99% (1140 of 1153)	Normal	
2		19-042017-04-27	Default	3	27	Prerecognition f...	23.04.2017 21:02:15		99% (1141 of 1152)	Normal	
1		19-042017-04-27	Default	3	27	Exceptions	23.04.2017 19:24:23		99% (1154 of 1159)	Normal	

# Queues Exploring

- Any queue can be explored to see all the tasks
- Task from the queue can be opened for processing
- Task (s) can be sent to a particular stage and user



# Queues Exploring

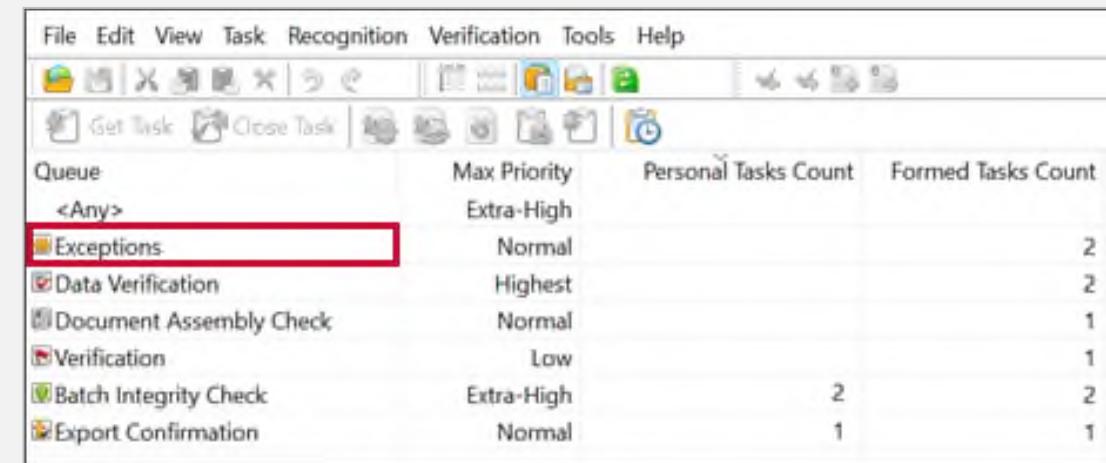
- Final approval to export the batch
  - Close Task button
- In cases of errors
  - send the task to the corresponding processing stage and specific operator
  - or correct all the errors

http://localhost/SingleEntryPoint - ABBYY FlexiCapture 12 (Verification Station)					
File Edit View Task Recognition Verification Tools Help					
#	A.	E.	Name	Status	Confidence Lev...
1			Banking_eng		88% (549 of 621)
1.1			Page 1		92% (305 of 332)
1.2			Page 2		86% (121 of 140)
1.3			Page 3		83% (123 of 149)

File Edit View Task Recognition Verification Tools Help		
Queue	Max Priority	Personal Tasks Count
<Any>	Extra-High	
Exceptions	Normal	
Data Verification	Highest	
Document Assembly Check	Normal	
Verification	Low	
Batch Integrity Check	Extra-High	2
Export Confirmation	Normal	1

Documents can be sent to exceptions:

- automatically, e.g. in case of export failure
- manually by operators



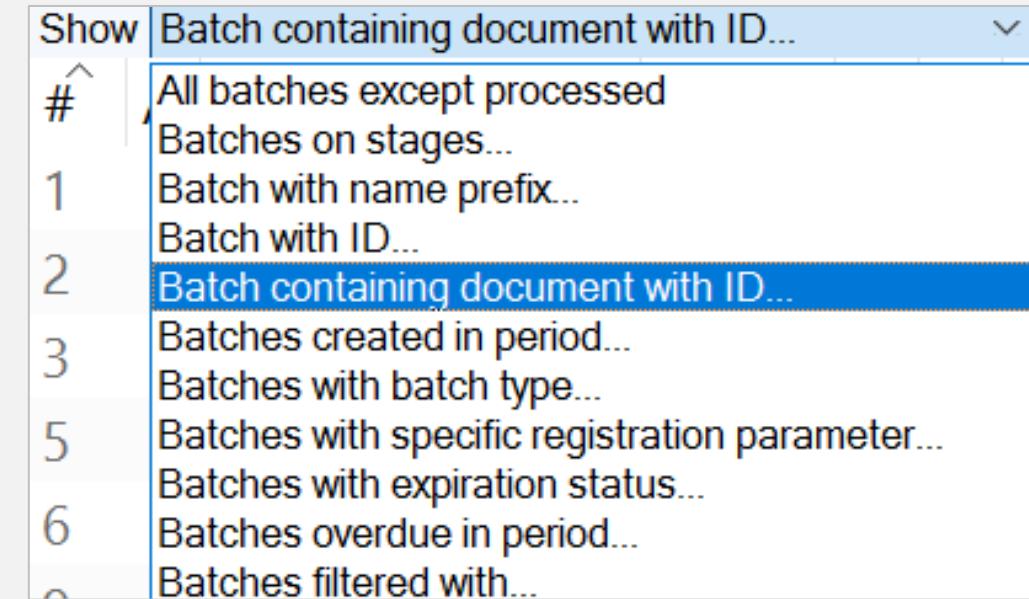
Queue	Max Priority	Personal Tasks Count	Formed Tasks Count
<Any>	Extra-High		
Exceptions	Normal	2	2
Data Verification	Highest	2	1
Document Assembly Check	Normal	1	1
Verification	Low	1	2
Batch Integrity Check	Extra-High	2	2
Export Confirmation	Normal	1	1

Read the comment for a task and see task priorities

Send a task to any processing stage, indicating a specific operator and comment if required

# Filtering Batches

- Simplify search and preview in Batch list
- Get fast access to necessary batch
- Filter batches



## Task Management

- Data Verification Operator
- Verification Operator
- Senior Verification Operator

In case of any problems, task or a set of documents can be sent to exceptions

- Exceptions queue is operated by Senior Verification operator



## ABBYY FLEXICAPTURE

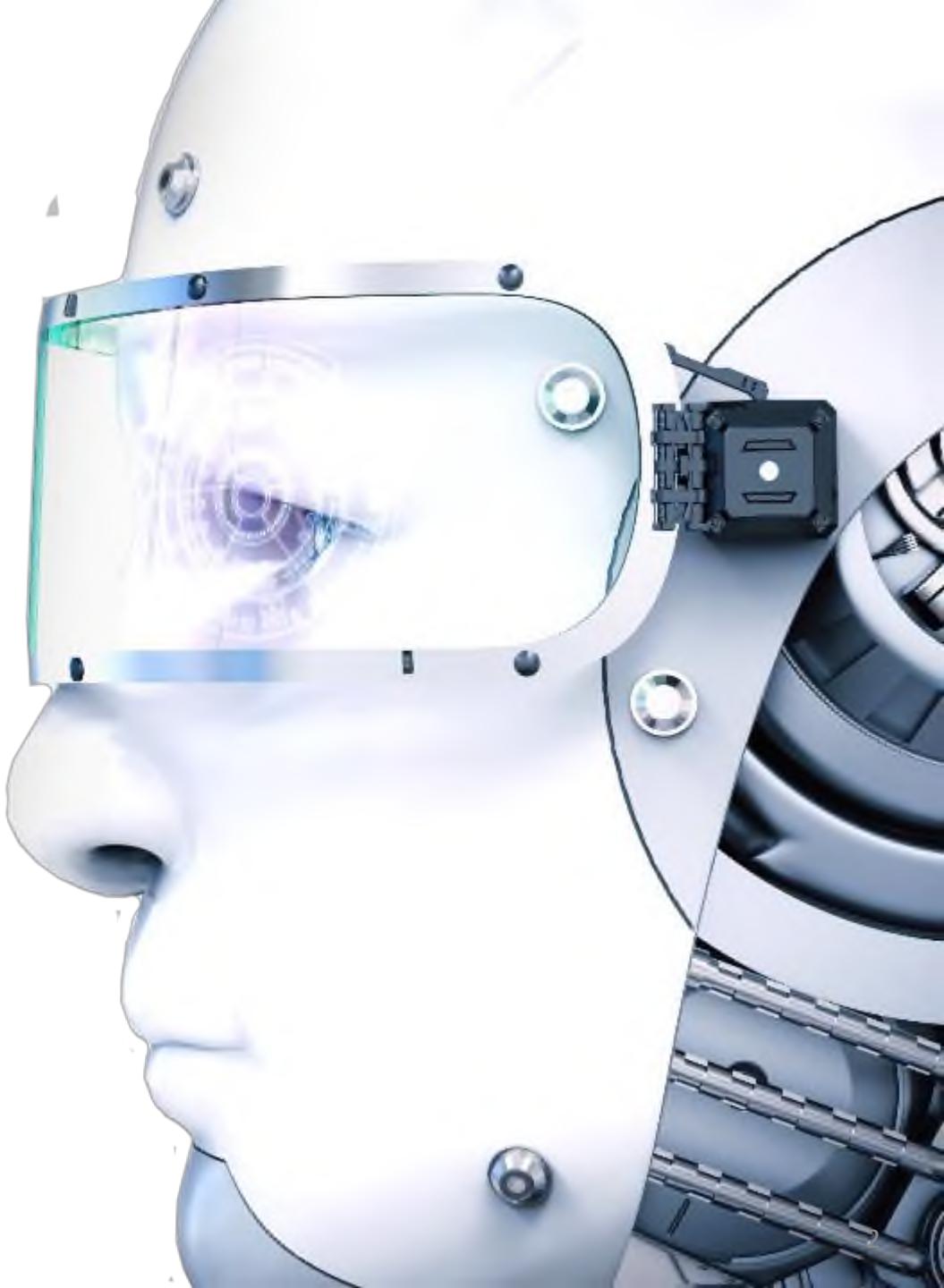
---

### EXPORT SETTINGS



# Overview

- Types of Export
- Data Export Step
- Document Definition: Export Settings
- Export format
  - Files
  - Databases
  - SharePoint
  - Custom export using scripts



# Types of Export

ABBYY®

## Export to data files

MS Excel Workbook (\*.xlsx, \*.xls)  
XML Document (\*.xml)  
JSON Document (\*.json)  
Comma-separated values file (\*.csv)  
Text file (\*.txt)  
DBase file (\*.dbf)

## Export to image files

TIFF  
JPEG  
PDF  
BMP  
PCX packbits  
PNG

## Custom export (script)

Export to ODBC-compatible database

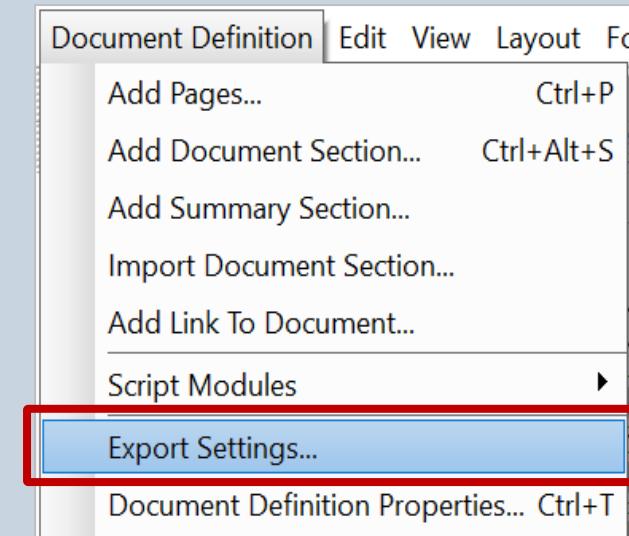
Export to SharePoint

Data is exported automatically  
on the Processing stations

Documents ready/confirmed for export

Export settings should be selected  
for each Document Definition

Exception: Document Set



## Project properties

Export *root* path

Example:  
\\127.0.0.1\Export Folder

## Document Definition properties

*Relative* path

Example:  
Export\_to\_data\_files

Batches are exported to  
\\storage\ExportFolder\Export\_to\_data\_files

# Deleting Exported Documents and Batches

ABBYY®

Processed documents  
can be automatically  
deleted from batches

Processed batches can  
be deleted from FC  
storage

- Deletion can be  
postponed

Specified in the *Batch  
type* settings

- If there are no Batch  
Types, the *Project*  
properties are used

# Export of Unrecognized Documents

ABBYY®

Batch can contain  
unrecognized documents  
(images)

- Bad quality of images for automatic classification
- Documents scanned by mistake
- Etc.

Unrecognized  
documents can be saved  
as images

Specified in the Batch  
type settings

If there are no Batch Types,  
the same Project properties are used

# Export to Multiple Destinations

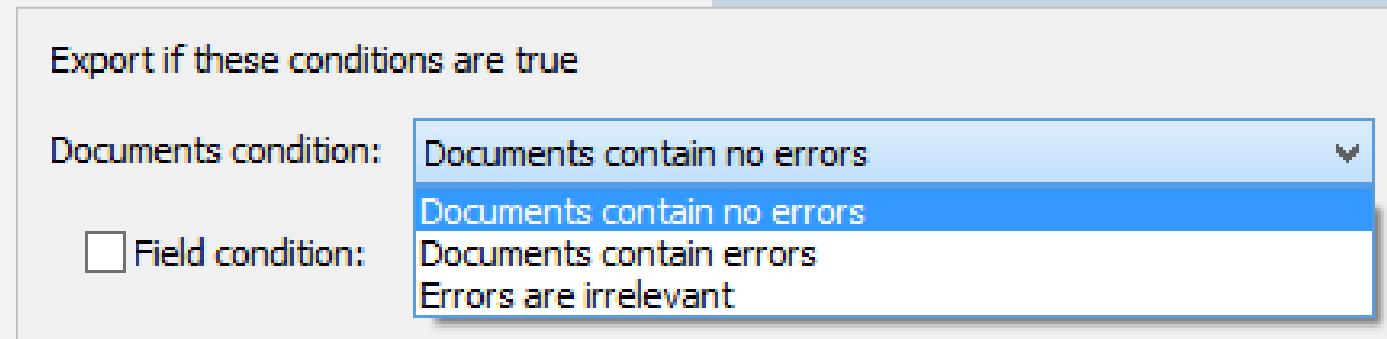
ABBYY®

Several export profiles for the same kind of document can be used

- Define how “critical” an export destination is (returns export failure if case of unsuccessful export)

Export profile for documents with errors can be created

- Select *Documents contain errors*



---

## Export of:

- **Data** to files, databases, using custom scripts
- **Document images** to TIFF, JPEG, etc.
- **Text** to searchable PDF, PDF/A

---

Export to multiple destinations is possible

---

Export is carried out automatically on the Processing Station



## ABBYY FLEXICAPTURE

---

PROJECT SETUP  
BATCH TYPES



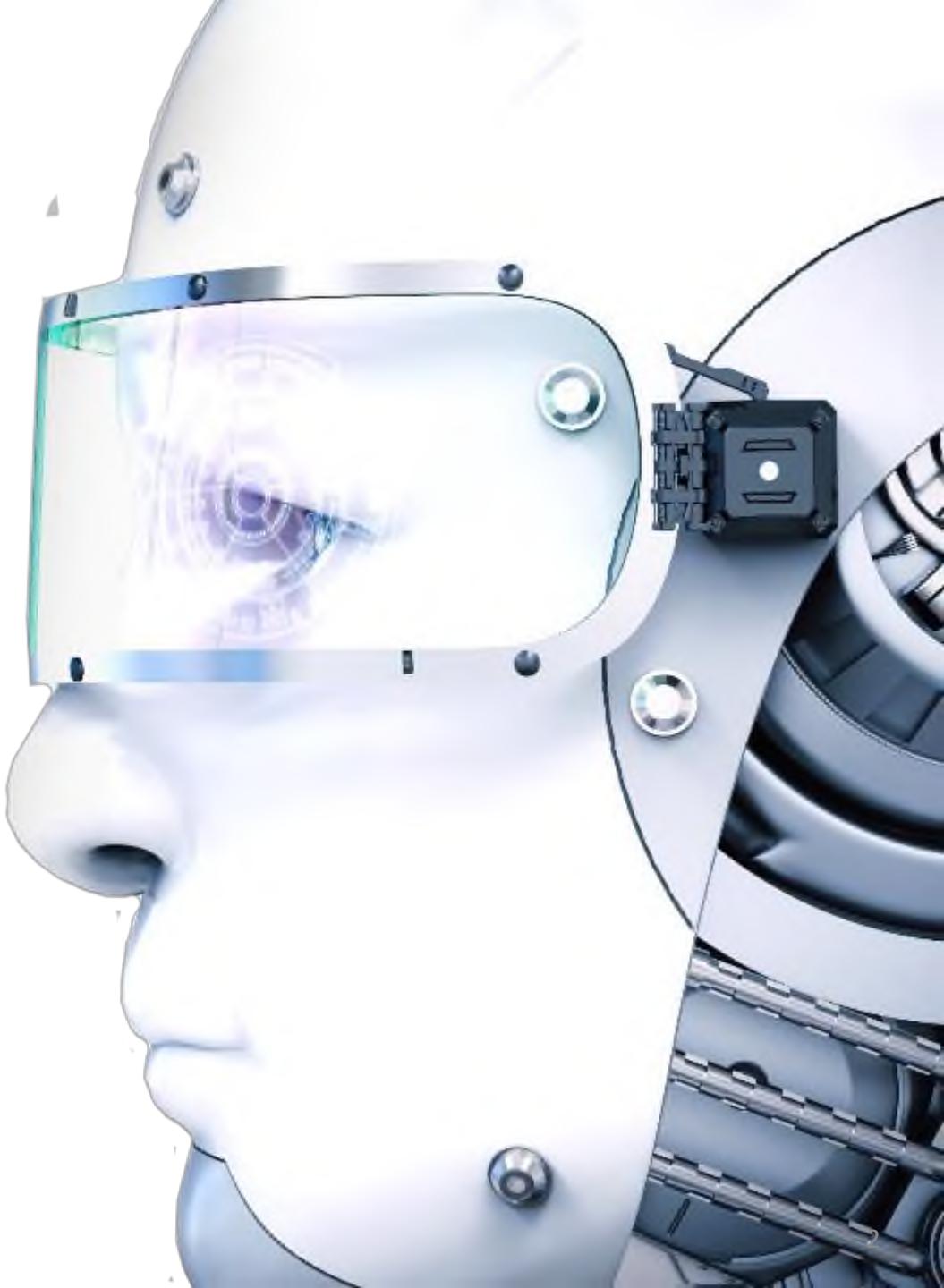
# Overview

## Batch types

## Batch type properties

- Recognition Options
- Document Separation and Image Pre-processing
- Registration Parameters
- Batch Type Event Handlers

## Operator Permissions



**Batch Type** is a set of batch processing options

Use of batch types:

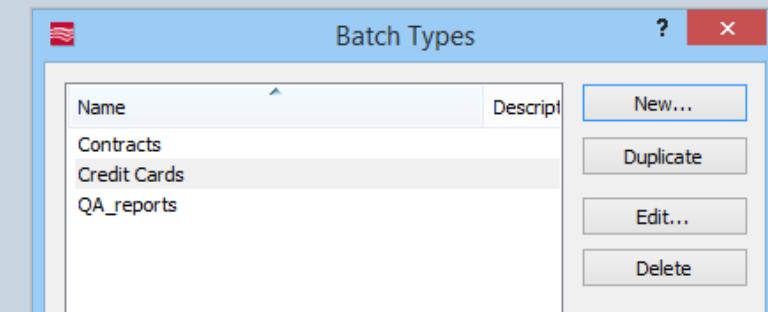
Multiple document definitions

Various schemas of verification  
for different sets of documents

Restricting permissions of operators  
for different document types

*By default there is no Batch types in the Project*

All documents are processed using  
the same processing settings and workflow

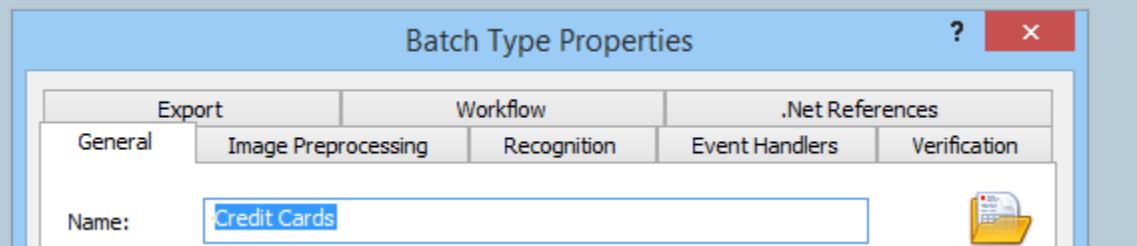


# Batch Type Properties

ABBYY®

## Batch Type properties:

- List and order of possible Document Definitions
- General image pre-processing and verification parameters
- Batch check (event handlers)
- Export settings
- Batch processing stages (workflows)
- Classifier (if used)



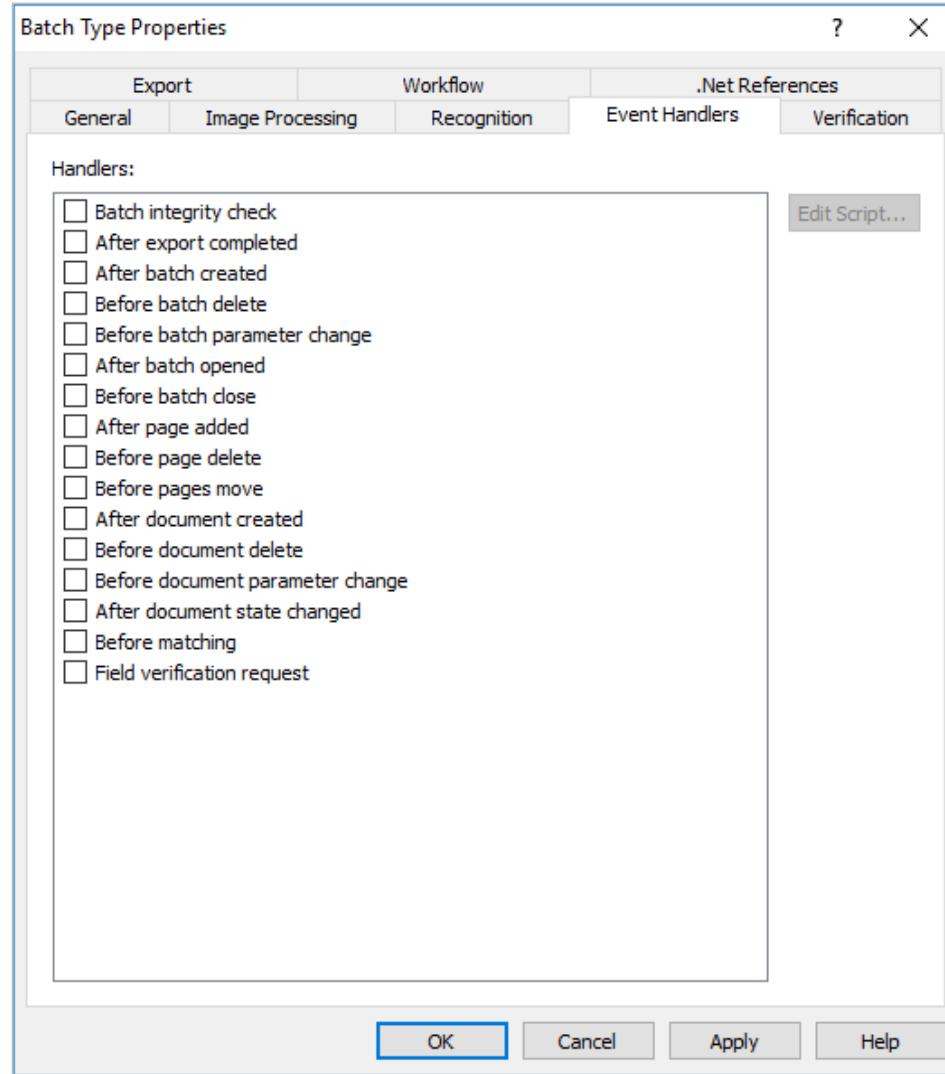
## Default Batch Type:

- Properties are defined by the *project properties*
  - Can be used if document processing does not require multiple workflows

At the Scanning Station you can select “Default” as a separate Batch Type

# Batch Type Event Handlers

ABBYY®



Custom handling of different document and batch events by scripting

Scripts serve for managing the documents and allow to write the processing log

# Event Handlers. Batch Integrity

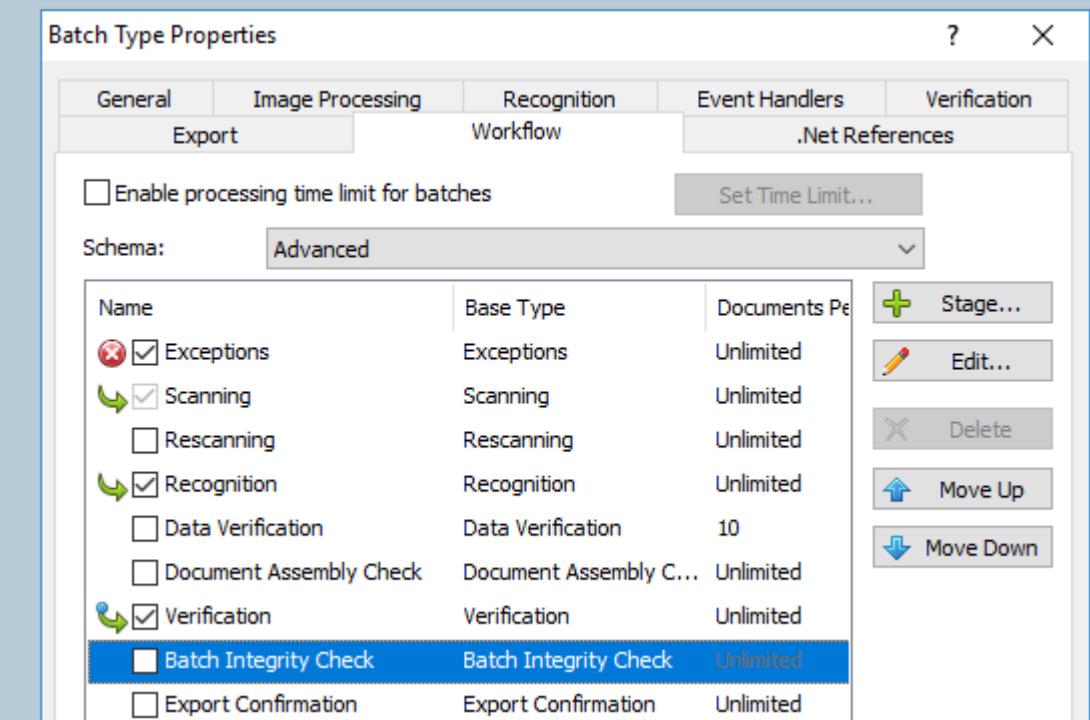
ABBYY

**Batch level rule** is defined by script  
and can be used to:

- check that nothing is missing in the batch
- ensure that a particular field is unique on all document of the given type

Batch Type Event Handler –  
Batch integrity check

Script rule shows errors for documents  
on a separate workflow stage



# Operator Permissions

ABBYY®

The operators' rights must be specified  
for the selected project

Available batch types can be set for each role

Operators can only work with the tasks  
of the allowed batch types

Permission	
Roles	Projects
<input type="checkbox"/> Scanning Operator	<input checked="" type="checkbox"/> Cust_Classification
<input type="checkbox"/> Data Verification Operator	<input checked="" type="checkbox"/> SLA_Script_WeekDays
<input type="checkbox"/> Verification Operator	<input checked="" type="checkbox"/> Low_SLA_1_day
<input type="checkbox"/> Senior Verification Operator	<input checked="" type="checkbox"/> Normal_SLA_2_hours
<input checked="" type="checkbox"/> Processing Server	<input checked="" type="checkbox"/> Important_SLA_1_hour
<input type="checkbox"/> Project Settings Editor	<input checked="" type="checkbox"/> VIP_SLA_15_minutes
<input checked="" type="checkbox"/> Monitoring Operator	<input checked="" type="checkbox"/> SingleEntryPoint
<input type="checkbox"/> Administrator	<input checked="" type="checkbox"/> <Default>
<input type="checkbox"/> External User	<input checked="" type="checkbox"/> MyInvoiceProject
<input type="checkbox"/> Web Capture Operator	<input checked="" type="checkbox"/> <Default>
	<input checked="" type="checkbox"/> Invoices
	<input checked="" type="checkbox"/> Letters
<input type="checkbox"/> SLA	<input type="checkbox"/> SLA
	<input checked="" type="checkbox"/> <Default>
	<input type="checkbox"/> SLA-5-3

---

A batch type defines document processing parameters

---

Batch type specifies the set of document definitions, workflows, cleanup settings, etc.

---

A batch may be of only one type

---

Operators are given rights to process certain batch types

---

The “Default” batch type is used by default and its properties depend on the properties of the project



## ABBYY FLEXILAYOUT STUDIO

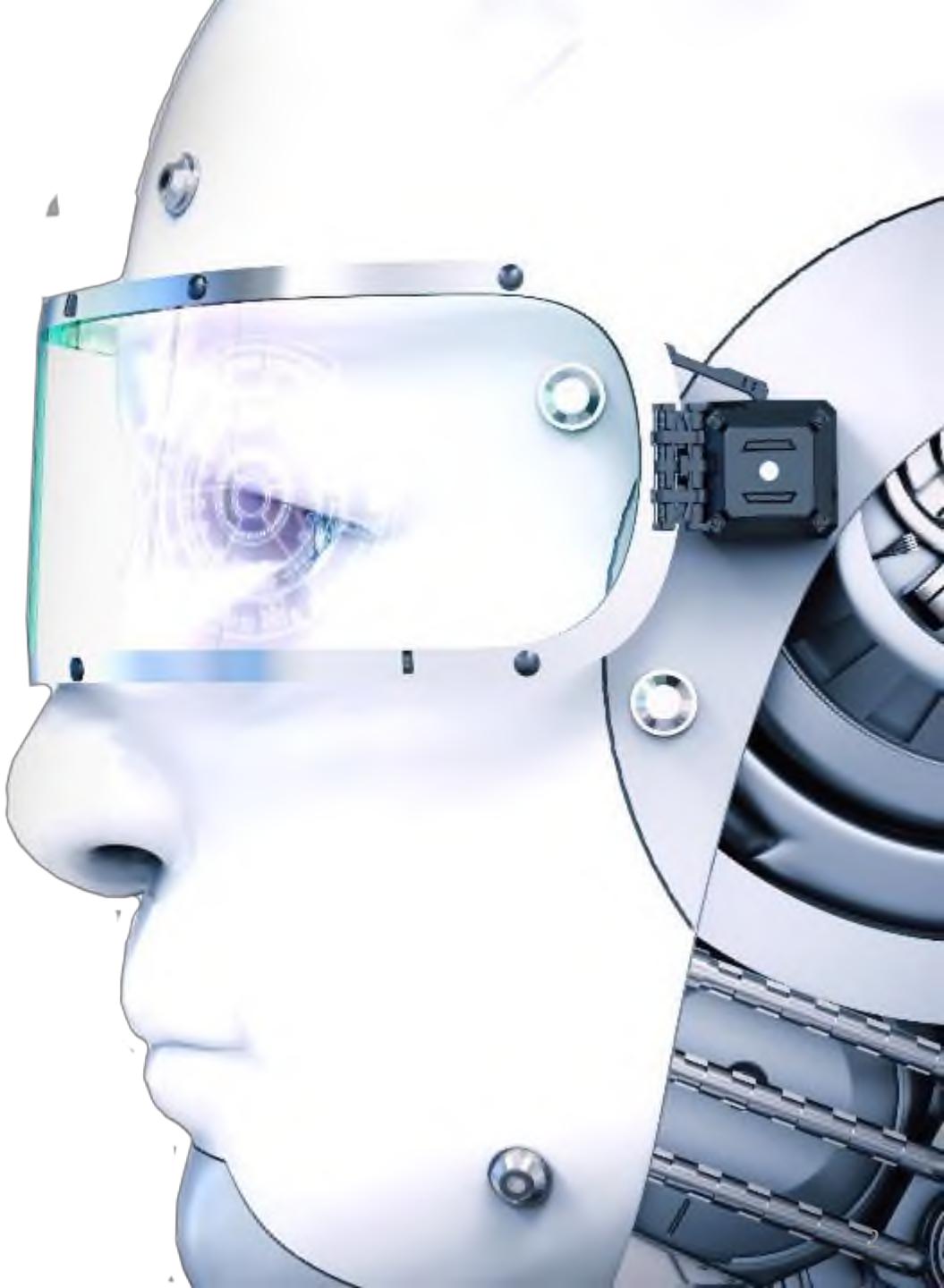
---

### OVERVIEW



# Overview

- Flexible Documents Processing
- Description of Flexible Documents
- FlexiLayout
- Creating FlexiLayouts
  - FlexiLayout with alternatives
  - FlexiLayout



## Semi-structured documents

- invoices,
- purchase orders
- reports, etc.

## Unstructured documents

- business cards
- checks
- standard contracts
- business letters, etc.

## Slightly shifted fields

- passports,
- air tickets, etc.

# Challenges of Flexible Documents Processing

ABBYY®

1

Locations of fields vary  
from one document  
to another

2

The exact location  
of field is unknown  
in advance

3

Not all fields may be  
present on all  
of the documents

# Challenges of Field Detection

ABBYY®

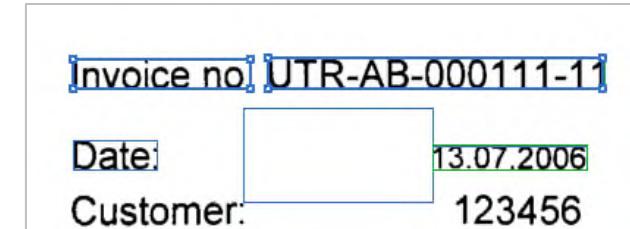
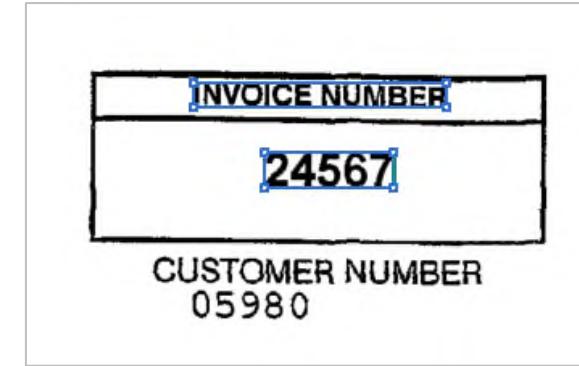
Various position on the page

Various (or missing) keywords

Various surrounding fields

Various position relative to the keyword (to the right, below, etc.)

- Various content
- Fixed coordinates can not be used
  - Special flexible description is required



# Flexible Layout - FlexiLayout

ABBYY®

## Fields search principles:

*Format of data*  
(date, phone, e-mail, etc.)

- regular expressions

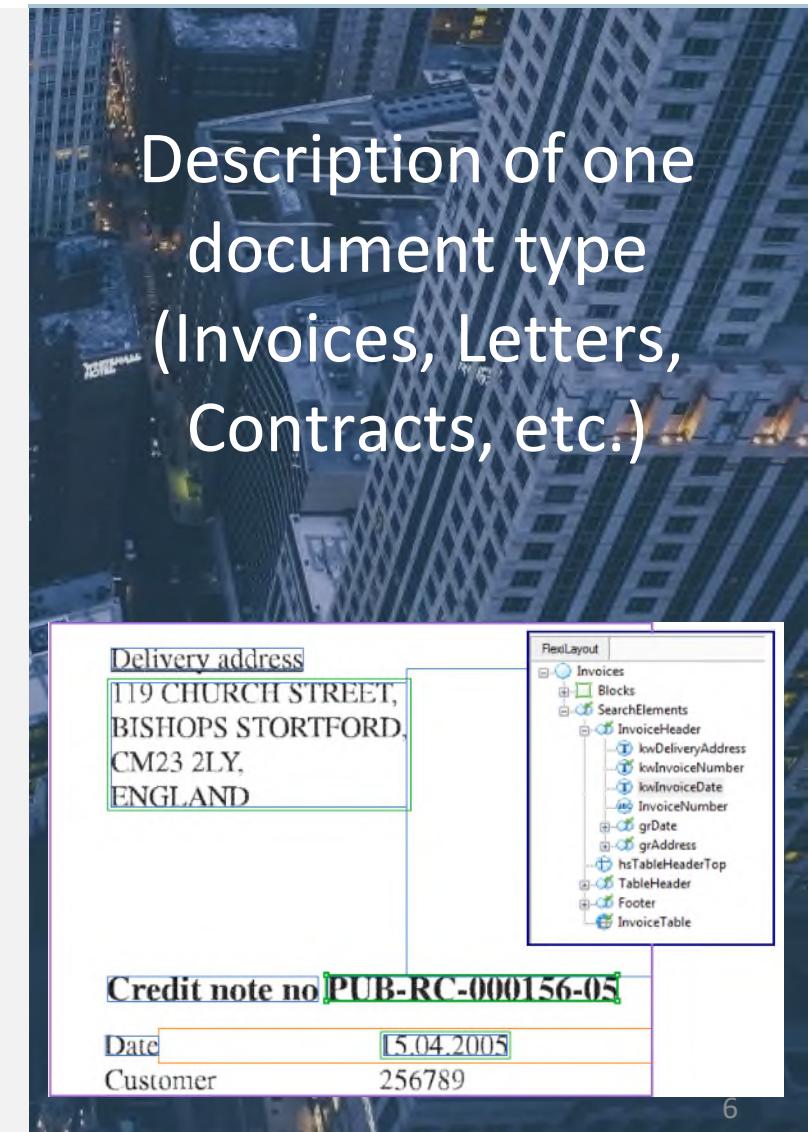
*Relationships* among the objects on a document

- right of, lower than, closest to, within given distance from, etc.

Approximate *location* of the object on the page

Search by *dictionaries*

Advanced possibilities for data search



# Recognition of Flexible Document

ABBYY®

- 1 Preliminary recognition of images
- 2 Searching for ID elements
- 3 Detecting the regions of fields
- 4 Recognition of data from fields regions

## *Identifiers of document type*

## Objects for automatic multi-page assembly

- first page and last page detection

## Fields to be captured from the document

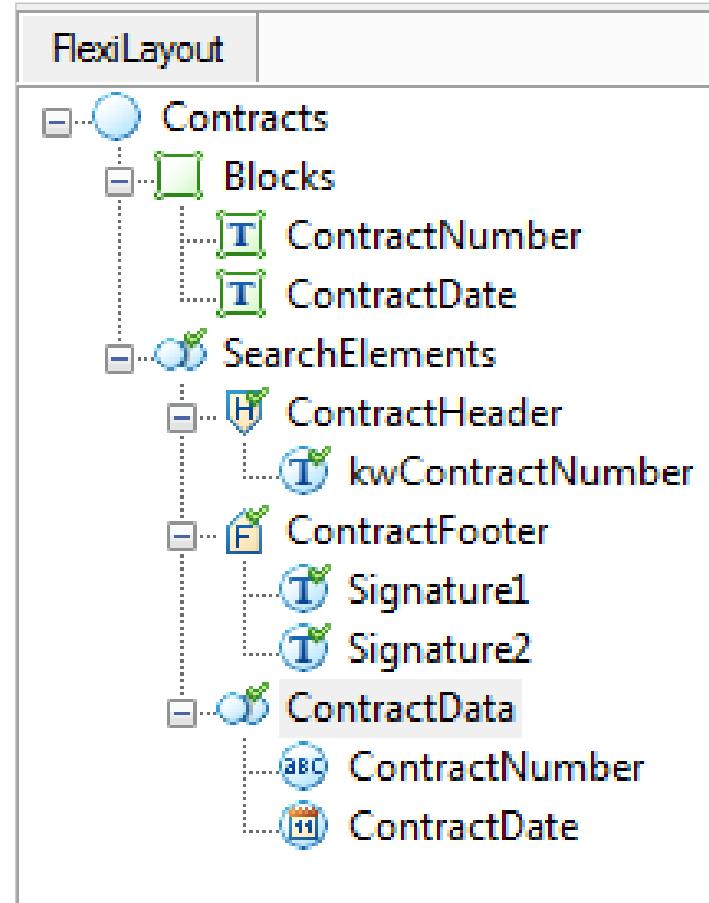
- “Blocks” in terms of FLS = future FC fields

## Elements that describe objects necessary to locate the fields

- Nearby objects— keywords, separators, white space, pictures, etc.

## Elements that describe fields' data

by format, location, relations to other objects, etc.



# ABBYY FlexiLayout Studio Interface

ABBYY®

Batch with test images

FlexiLayout Tree

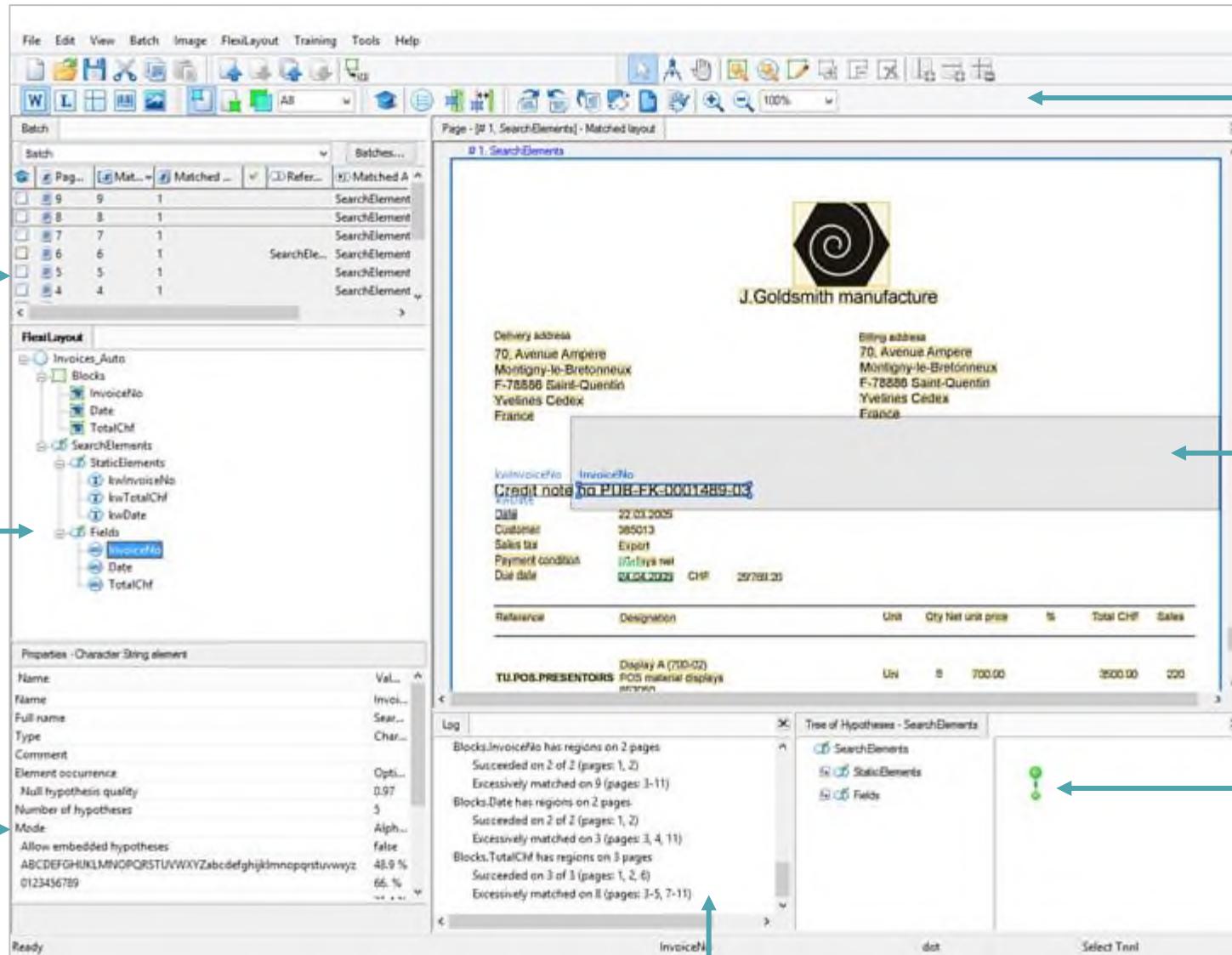
Found element's properties

Toolbars

Active test image with selected elements

Tree of hypotheses

Log Window



# FlexiLayout with Alternative Layouts

ABBYY®

## Suitable for:

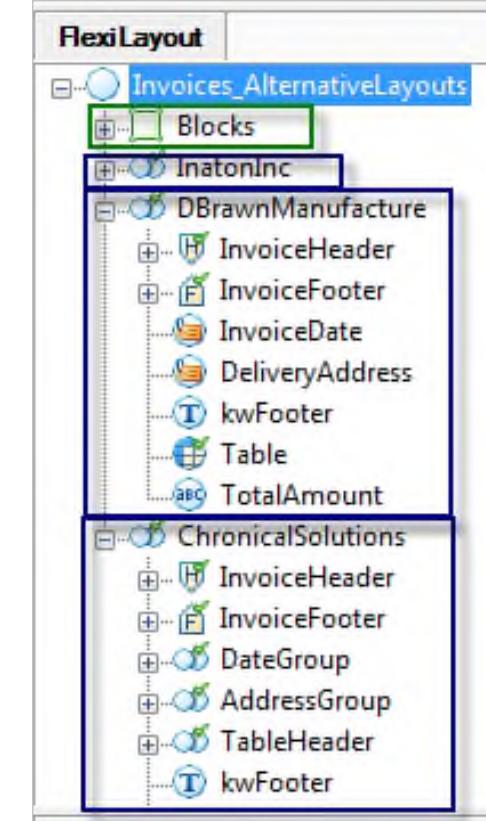
- Documents of one type with greatly varying field positions
  - e.g. driver licenses from different states

## Advantages:

- Easier structure of each alternative, less elements
- No need to describe various complicated search criteria

## Method:

- One list of fields (blocks), several independent trees of elements



# Types of Projects

ABBYY®

*FlexiLayout* – description of layout  
of one document type

Goal – to detect fields of particular  
document type

*Classifier* – description  
of classification methods of several  
types of documents

Goal – to distinguish several  
document types

---

Semi-/unstructured documents require adding a special FlexiLayout description to a document definition

---

Flexible Layout is a formalized description of the document structure which detects image areas corresponding to data fields

---

To capture data from flexible documents create Flexible layout in ABBYY FlexiLayout Studio and then export it into ABBYY FlexiCapture



## ABBYY FLEXILAYOUT STUDIO

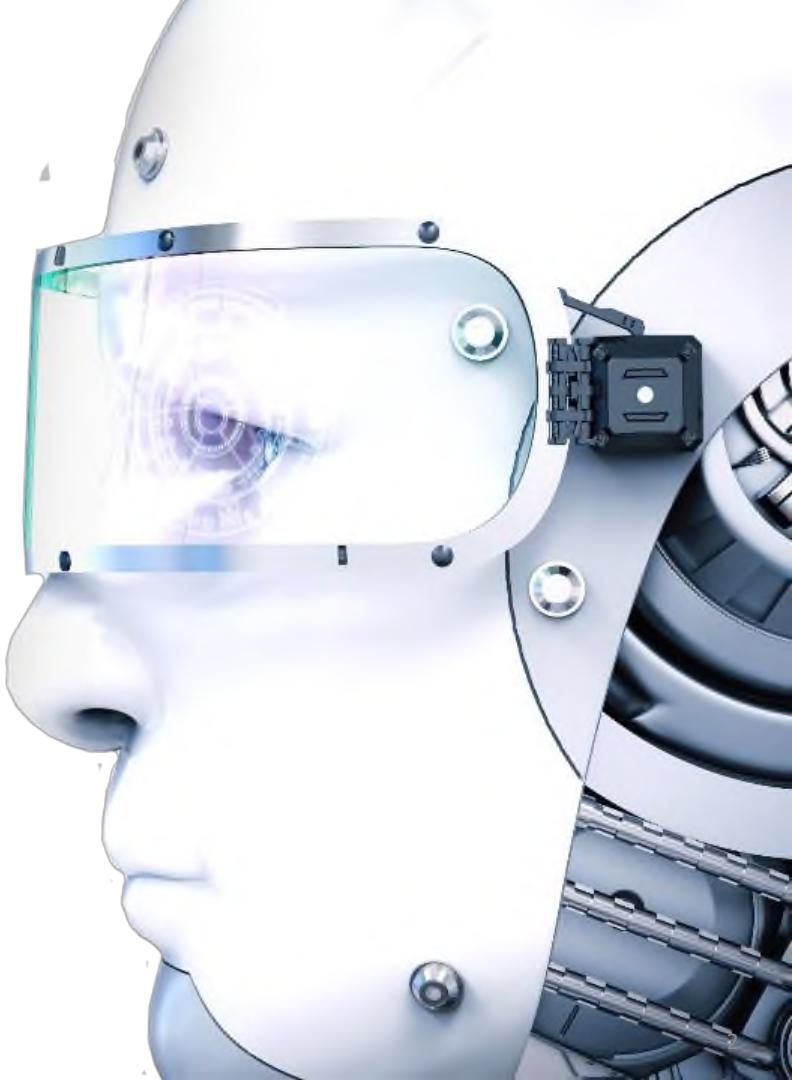
---

### FLEXILAYOUT CREATION STEPS



# Overview

- Creating a FlexiLayout
- FlexiLayout Project
- Identifiers and Anti-identifiers
- Matching of Flexilayout



## Preparation

1. FlexiLayout project
2. Batch of sample images
3. Pre-recognition of images
4. Analyzing pre-recognition results
5. FlexiLayout blocks

## Layout description

1. Identification elements
2. Header/footer elements for multi-page documents assembly
3. Basic elements to detect the required data positions
4. Source elements for blocks



## Required element

- Must be found
- If not found, matching stops
- Used for document identification!



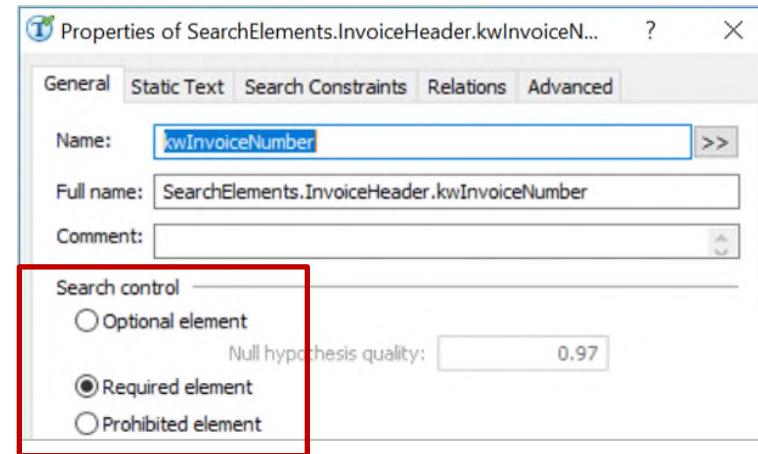
## Prohibited element

- If found, the chain is stopped
- Used to prevent matching!



## Optional element

- If not found, null hypothesis is generated
- FlexiLayout matching and search for other elements continues



Can be used if it is difficult to find the reliable objects for making them ID elements

E.g. “Total Amount” | “Total Sum” words should not be found in document of Price List type



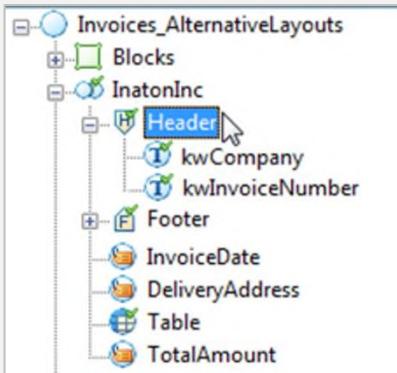
*Prohibited Elements - elements that are never present in this document type*

# ID Elements and Header Group

ABBYY®

Header Group contains elements for:

- Describing the ID objects for a document type
- Describing the first page of multi-page document



It is recommended  
to create  
*ID elements*  
(required  
elements)  
inside *Header group*



- After creating the FlexiLayout, project is exported to a \*.afl file
- The \*.afl file is added to a Document Definition within a FlexiCapture project

# FlexiLayout Folder

- Back up entire folder including subfolders
- The .fsp file can't be opened without the subfolders

Name	Date modified	Type	Size
ProfitLossStatementsBatch	6/22/2018 6:13 PM	File folder	
ProfitLossStatementsTemplates	6/22/2018 6:12 PM	File folder	
ProfitLossStatements.afl	6/22/2018 6:13 PM	FlexiLayout descri...	1 KB
ProfitLossStatements.fsp	6/22/2018 6:12 PM	FlexiLayout Studio...	1 KB

# Matching Order in FlexiCapture

ABBYY®

1

STRUCTURED  
DOCUMENT  
DEFINITIONS

2

FLEXILAYOUTS WITH  
HEADER GROUPS

3

ALL OTHER  
DESCRIPTIONS

The first suitable  
structured definition  
will be chosen

Or descriptions  
in the classifier project

Depending on the option “Use first  
acceptable FlexiLayout”:

- From 1 in the list till 1 **successfully matched**  
or
- All: resulting definition is selected by quality

# Matching Order in FlexiCapture

1. Structured Document Definitions are matched  
the first suitable structured definition will be chosen

2. FlexiLayouts are matched

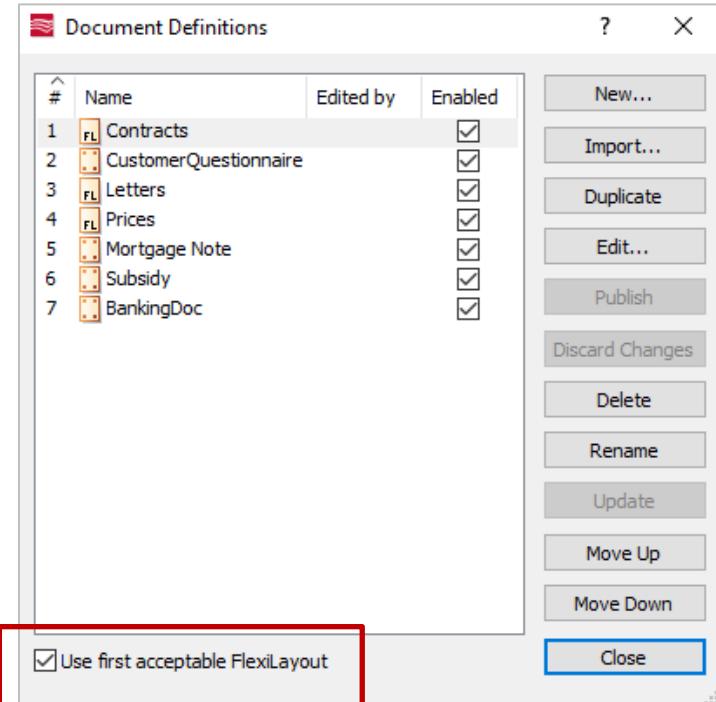
Descriptions with a Header groups or descriptions  
in the classifier project

All other descriptions:

Depending on the option “Use first acceptable FlexiLayout”:

From 1 in the list till 1 successfully matched,  
or

All. The resulting definition is selected by comparing  
the matching quality



Steps for FlexiLayout creation are:

1. Project creation
2. Adding and pre-processing text images
  - Reference assembly of multi-page documents
3. Pre-recognition
4. Document analysis
5. Blocks creation



## ABBYY FLEXILAYOUT STUDIO

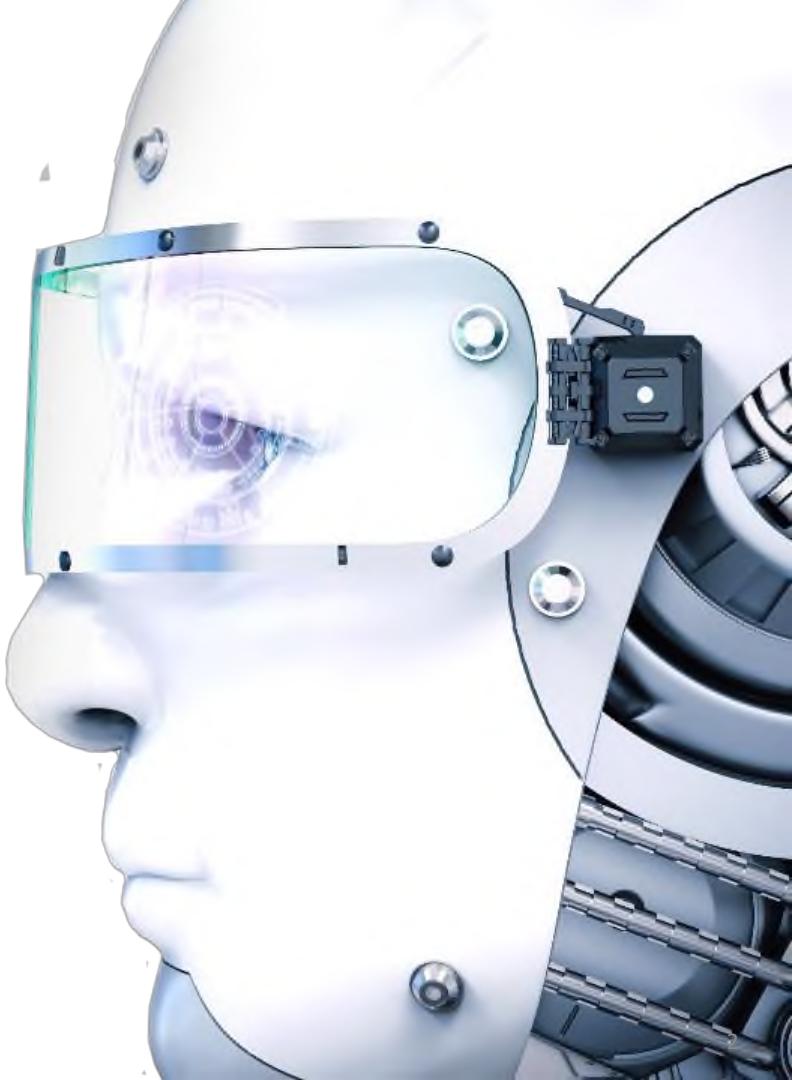
---

ELEMENTS  
SEARCH BY FORMAT



# Overview

- Elements Types
- Format settings



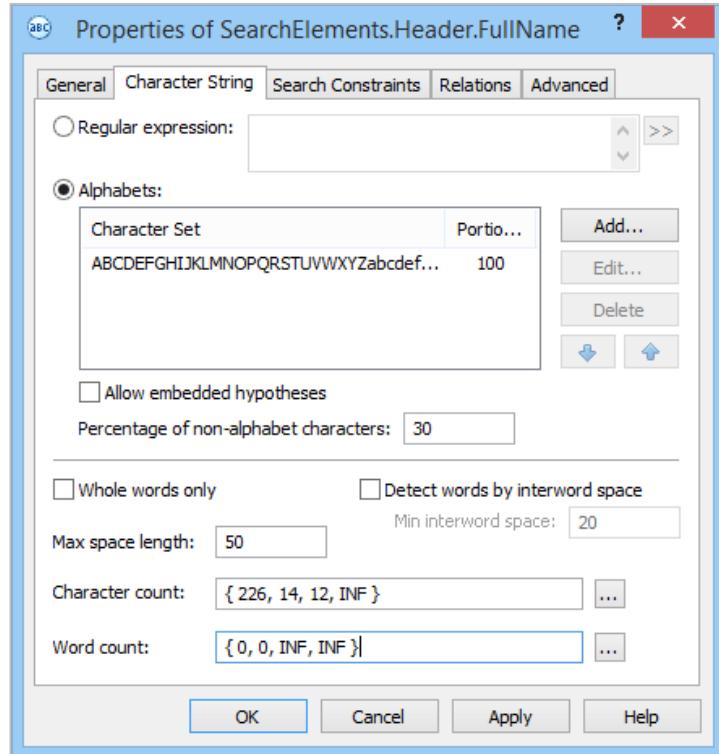
# Format Properties of Elements

Influence on:

- Selection the correct objects from the image
- Assigning the quality for an object

Define the penalization coefficients:

- To reduce the quality of slightly inappropriate objects but not to ignore it



# Static Text



Text variants are separated by the “|”

Phrase variants are put into braces - { }|{ }

- “A B”, “A C” -> A B|C or {A B}|{A C}

Take spaces into account

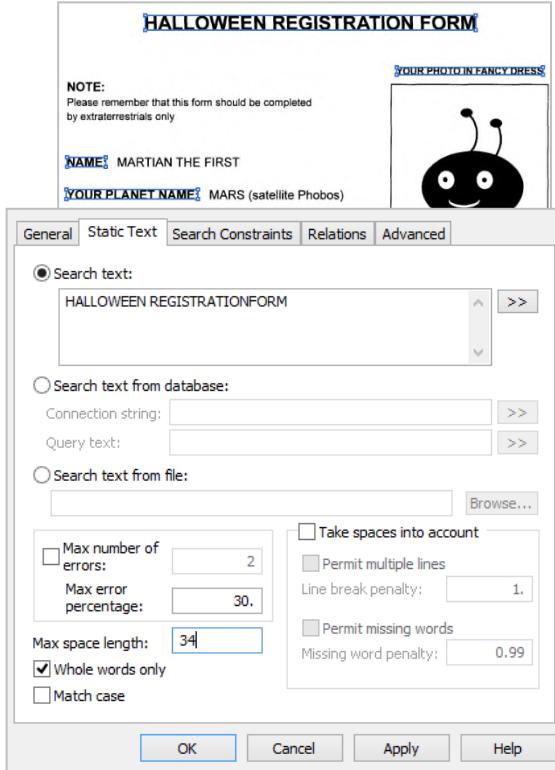
- don't forget to type the spaces!

Max space length (in dots by default)

Whole words only

- can't be a part of another word

Match case



# Errors

Errors are calculated for the whole search text as for *one word*

Search text with spaces + *Take spaces into account* enabled:

**YOUR PLANET NAME:** – errors are allowed in *each word*

Example:

- “YEAR”, 30% of errors allowed -> only 1 error allowed
- “DAY”, 30% of errors allowed -> no errors allowed

<input checked="" type="checkbox"/> Max <u>n</u> umber of errors:	1
Max <u>e</u> rror percentage:	30

# Labeled Field

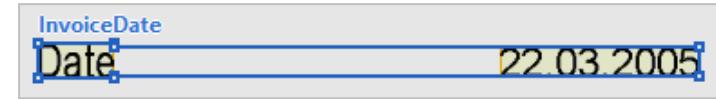
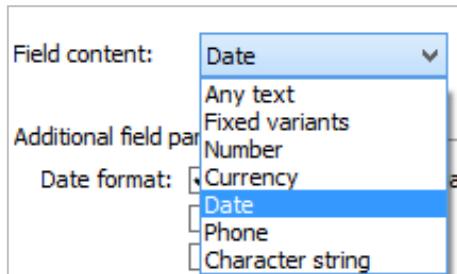


Is used to search the fields with key words (labels) nearby

Sub-elements:

- *Label* – field header
- *Gap* – space between label and value
- *Field* - field value

Expand to Group



# Paragraph



ABBYY®

Line count

Alignment and Indentation

Spacing

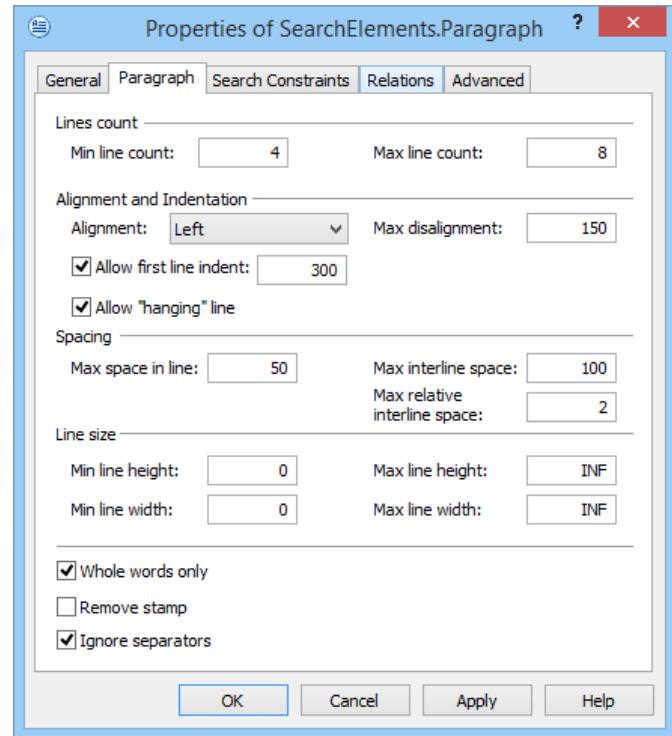
Line size

Min\Max line height

Min\Max line width

Remove stamp

Shipping address:  
"Clear Sky"  
Mr. Brian Jones  
USA, Columbia, Washington  
DC,  
Roman Street,  
Building 16, Office 2  
Co-11167





For images of good quality!

Badly recognized dates might not be detected

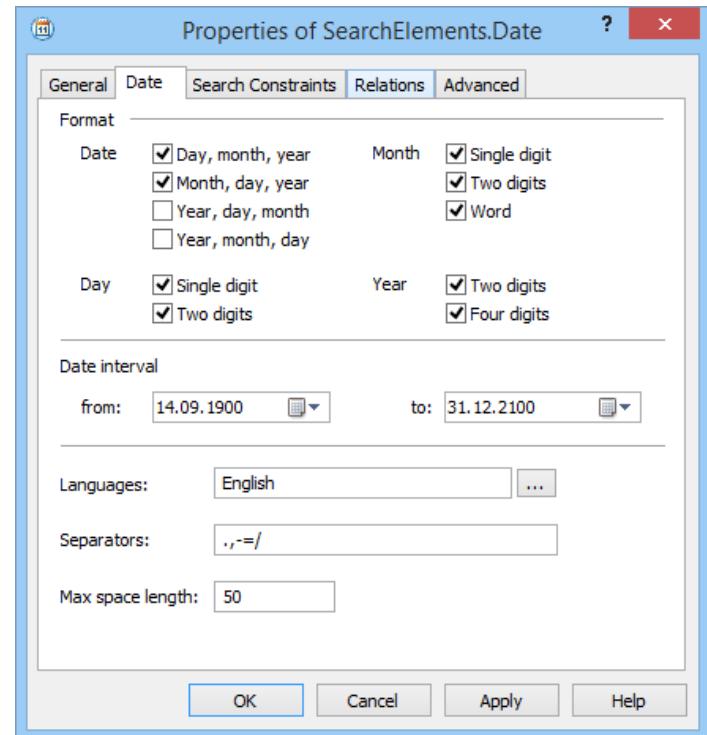
## Format

- Day-month-year order
- Day, Month, Year – number of digits
- One, several or all variants

## Date interval

- Values exceeding the interval are ignored

## Language (in case of Month->Word)



# Currency



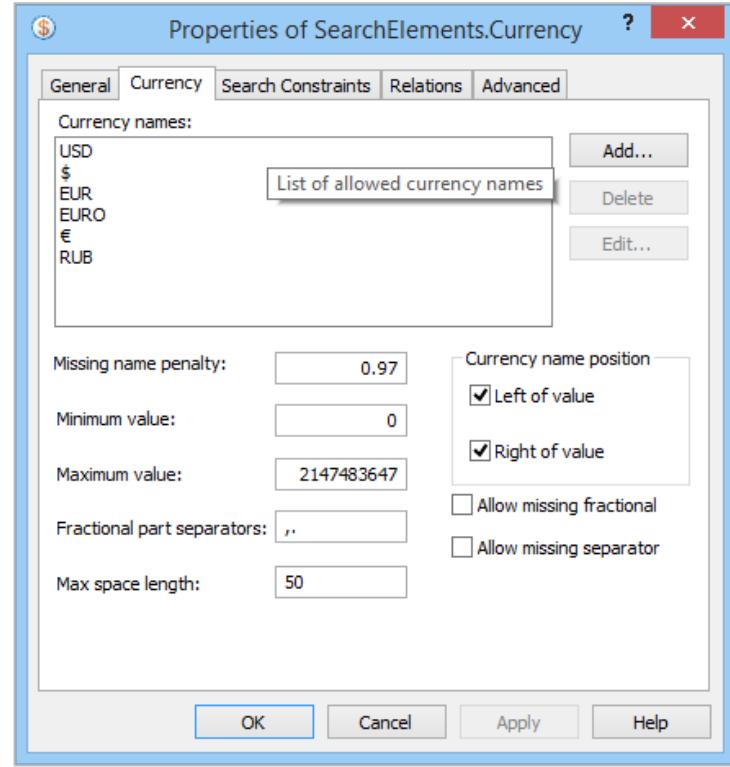
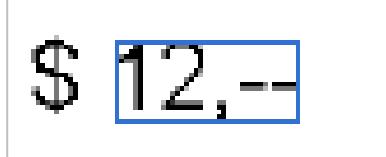
ABBYY®

## Currency names

List of allowed currency names

Set in words or by a symbol (\$, ¥)

*No errors are allowed during currency name search*



# Phone



ABBYY®

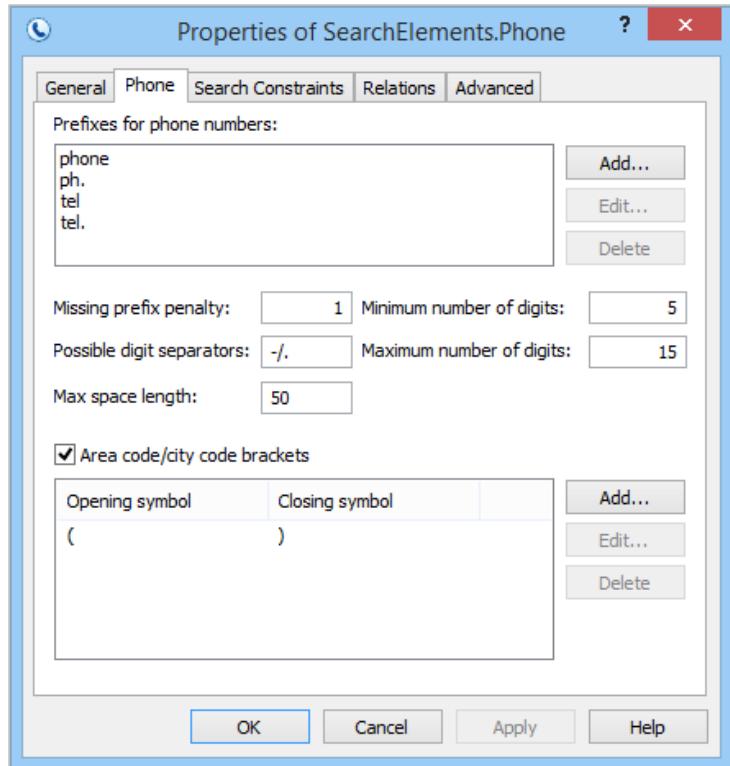
Prefixes for phone numbers

Words and word combinations are allowed

20% of errors are allowed

Missing prefix penalty

Tel: +7 (095) 783 3700 / 2723



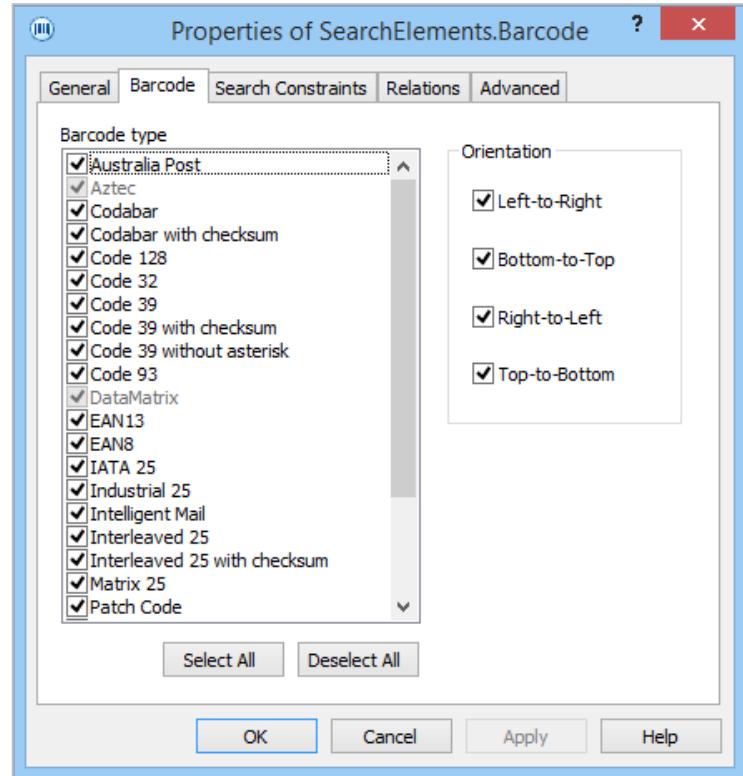


## Barcode type

Several variants can be selected

## Orientation

Several variants can be selected



# Separator



ABBYY®

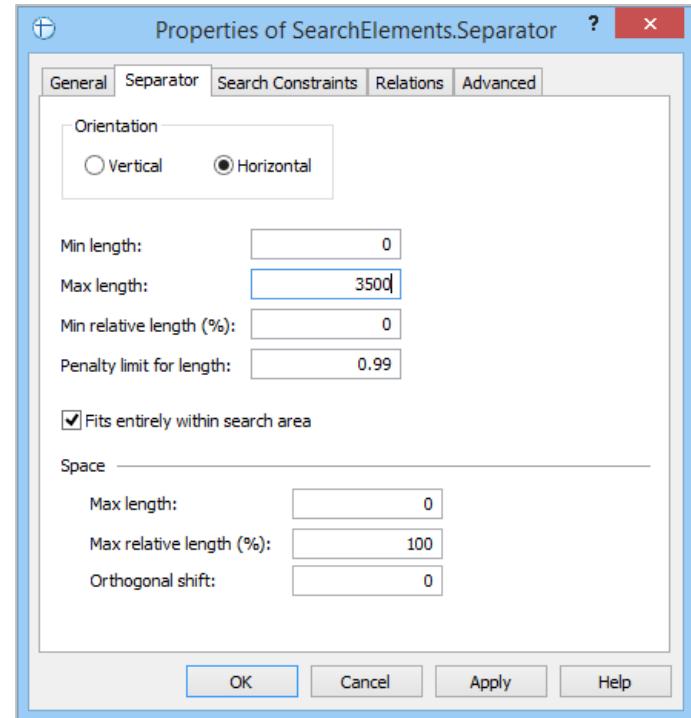
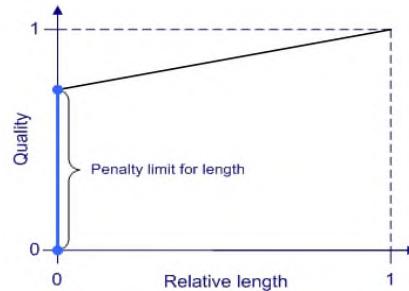
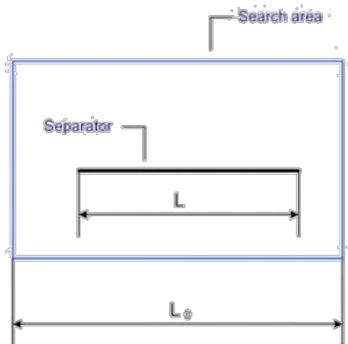
## Orientation

### Min length & Max length

### Min relative length (%)

- $=L/L_0$

### Penalty limit for length



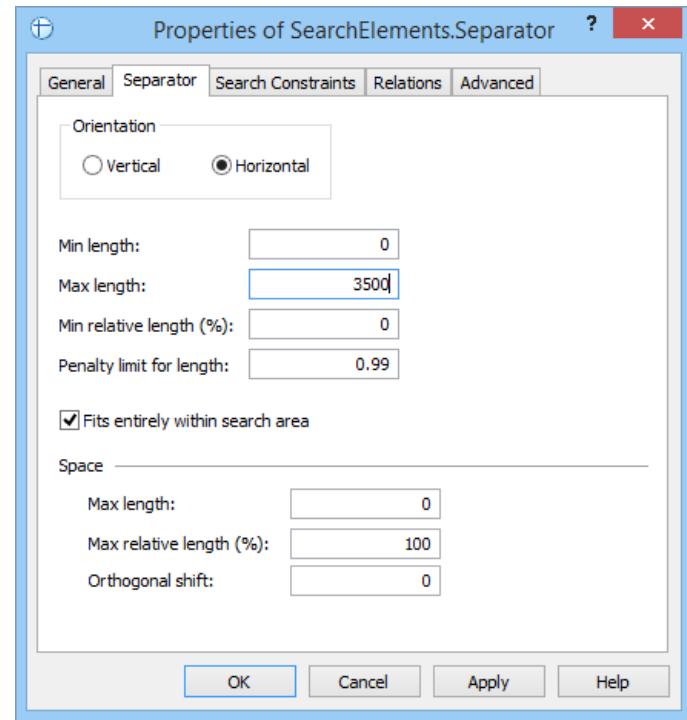
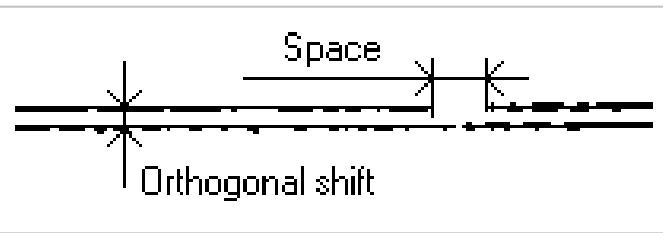
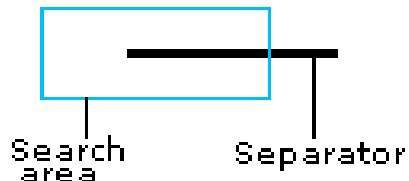
# Separator



ABBYY®

Fits entirely within search area

Space





## Orientation

## Min height / weight

Depends on the orientation

The screenshot shows a document page with a white gap at the top. The page contains text and a barcode. Below the page is the 'Properties of SearchElements.WhiteGap' dialog box. The 'White Gap' tab is selected. The 'Orientation' section has 'Horizontal' selected and 'Min height:' set to 20. The 'Search among objects of type:' section includes options for Text, Punctuation mark, Picture, Inverted text, Checkmark, Separator, and Barcode, with 'Separator' checked. The 'Noise' section has 'Lower threshold limit:' at 37, 'Upper threshold limit:' at 150, and 'Threshold coefficient (%):' at 20. At the bottom are OK, Cancel, Apply, and Help buttons.

## Search among objects of type

## Noise

Threshold coefficient (%)

Lower threshold limit (in dots)

Upper threshold limit (in dots)

# Object Collection



ABBYY®

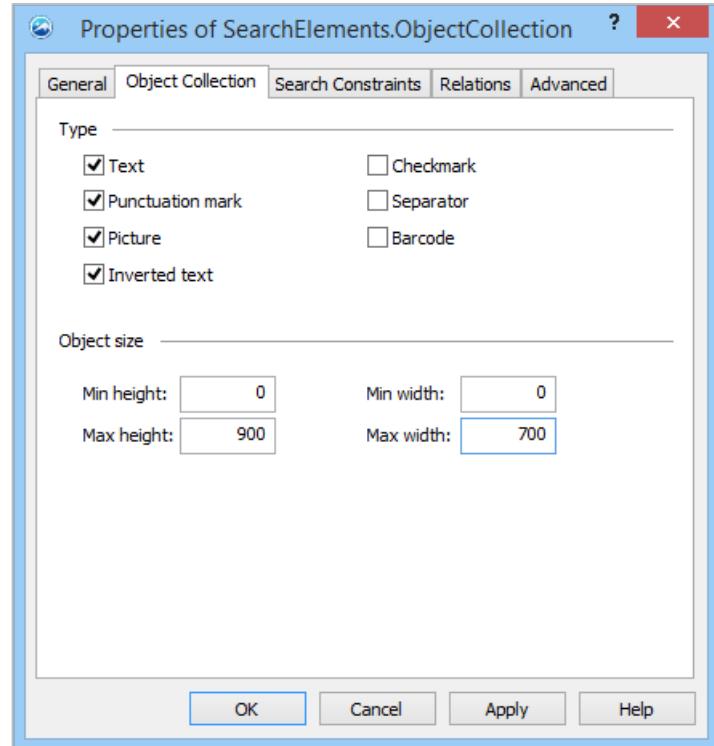
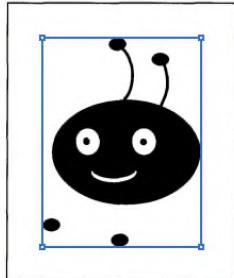
## Type

Type of searched objects

## Object size

The object's length and height in dots

*For each of the objects*





## Logical element grouping

## Subelements

Simple

Composite

Reference	Designation	Net Unit price	%	Total CH
Reference	Designation	Quant	UnitPrice	Total
TU.POS.PRESENTOIRS	Display A (700-02) POS material displays 653050	5	700.00	3500.00
TU.POS.PRESENTOIRS	Display C (703-02) POS material displays 653050	3	480.00	1440.00
TU.POS.PRESENTOIRS	Winch (711-02) POS material displays 653050	1	120.00	120.00
TU.POS.MATERIALS	Mirrors (716-00) POS Material diverse 653010	15	210.00	3150.00

## Used to:

Simplify navigation through FlexiLayout

Decrease the number of hypotheses

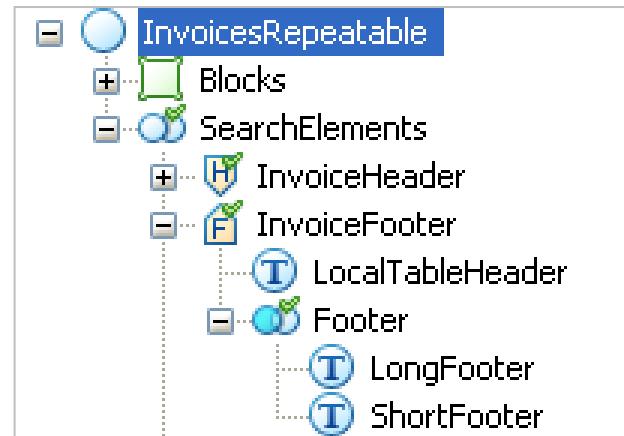
Search grouped elements more steadily



Is used when several elements are created to find the same block

It searches for sub elements until the first one is found

No need to specify that the next element in group should be found only if the previous is not found





## Line Of Elements (LineGroup)

Search of subelements of various types positioned in the same line

Any text in between the objects is allowed

Vertical shift is allowed

LineGroup	Reference	Designation	Unit	Qty	Net unit price	%	Total CHF
TU.POS.PRESENTOIRES	Display A (700-02) POS material displays 653050		Uni	5	700.00		3500.00

Properties of SearchElements.LineGroup

General Line of Elements Child Elements Search Constraints Relations

Child Elements:

**kwUnitPrice**

Add New SubElement

Name: kwDesignation

Type: Static Text

Element parameters

Name: kwUnitPrice

Variants: Unitprice

OK Cancel

Max number of errors: INF

Take spaces into account

Permit missing words

Missing word penalty: 0.99

Max space length: 50

OK Cancel Apply Help

# Summary

- The purpose of using elements is to find objects on the images which regions can be used as regions for blocks
- Element properties allow specifying search criteria for image objects
- Element format settings influence:
  - selection of the correct objects from the image
  - assigning the quality of an object



## ABBYY FLEXILAYOUT STUDIO

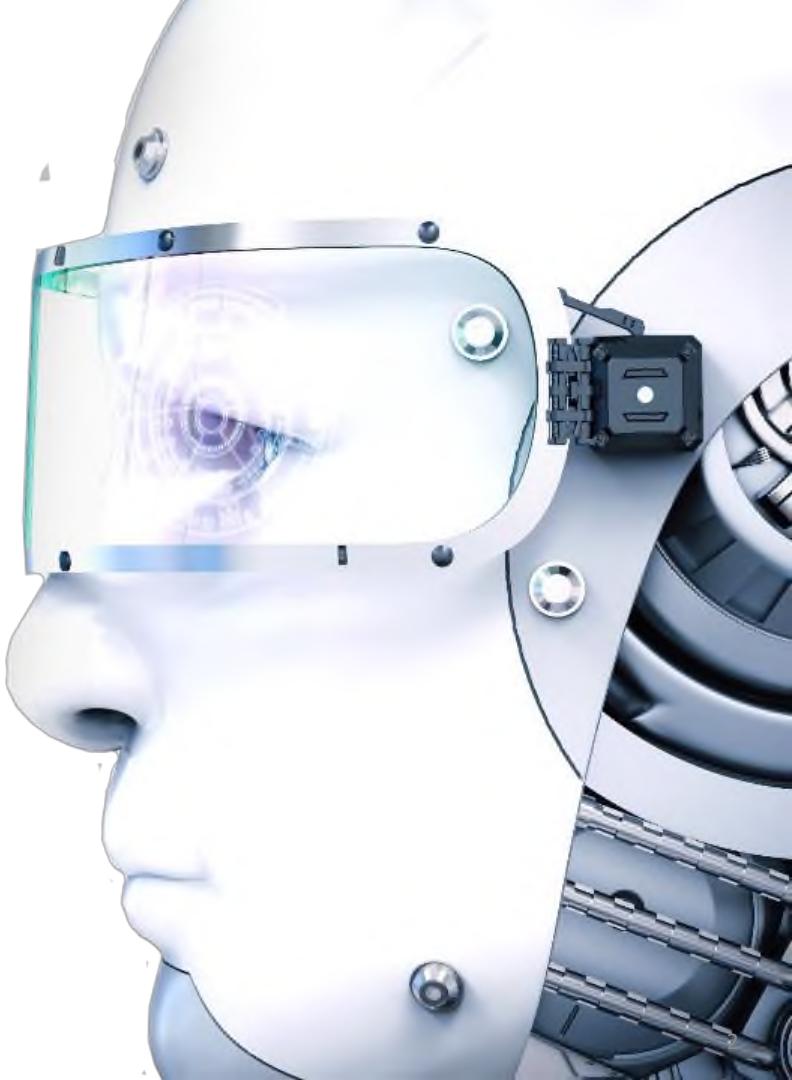
---

ELEMENTS SEARCH AREA



# Overview

- Element properties
- Element search area description
  - Absolute
  - Nearest
  - Relations, offsets
- Elements References





J.Goldsmith manufacture

**DeliveryAddress**

Delivery address:  
70, Avenue Ampere  
Montigny-le-Bretonneux  
F-78886 Saint-Quentin  
Yvelines Cedex  
France

Billing address  
70, Avenue Ampere  
Montigny-le-Bretonneux  
F-78886 Saint-Quentin  
Yvelines Cedex  
France

**kwInvoiceNumInvoiceNumber**

Credit note no: DUR-FK-0001489-0

Invoice Date: 22/04/2005

Customer 385013  
Sales tax Export  
Payment condition 30 days net  
Due date 24.04.2005 CHF 29769.20

**InvoiceTable Header****kwDesignation**

kwReference	kwDesignation	Unit	kwQnt	kwUnitPrice	kwTotal	kwSales
			Quantity	UnitPrice	%	Total
TU.POS.PRESENTORS	Display A (700-02) POS material displays 553050	Uni	5	700.00		3500.00
TU.POS.PRESENTORS	Display C (703-02) POS material displays 553050	Uni	3	480.00		1440.00
TU.POS.PRESENTORS	Winch (711-02) POS material displays 553050	Uni	1	120.00		120.00
TU.POS.PRESENTORS	Sets of 6 blocks (704-02) POS material displays 553050	Uni	10	200.00		2000.00
TU.POS.PRESENTORS	Set of 2 cordials (E10-96) POS material displays 553030	Uni	5	110.00		550.00
TU.POS.PRESENTORS	Counter Display (702(n)-06) POS material displays 553030	Uni	10	360.00		3600.00
TU.POS.PRESENTORS	Display B (701(a)-02) POS material displays 553060	Uni	10	530.00		5300.00
TU.POS.PRESENTORS	Mirrors (716-00) POS Material diverse 553040	Uni	15	210.00		3150.00

A logical part of a FlexiLayout  
that is used to:

- Identify documents
- Locate blocks
- Find another element

# Search of Elements

ABBYY®

Where to search for an object (Search area)  
What kind of data it is (Format of element)

Search Area can be described by:

Absolute coordinates

Relations to other objects

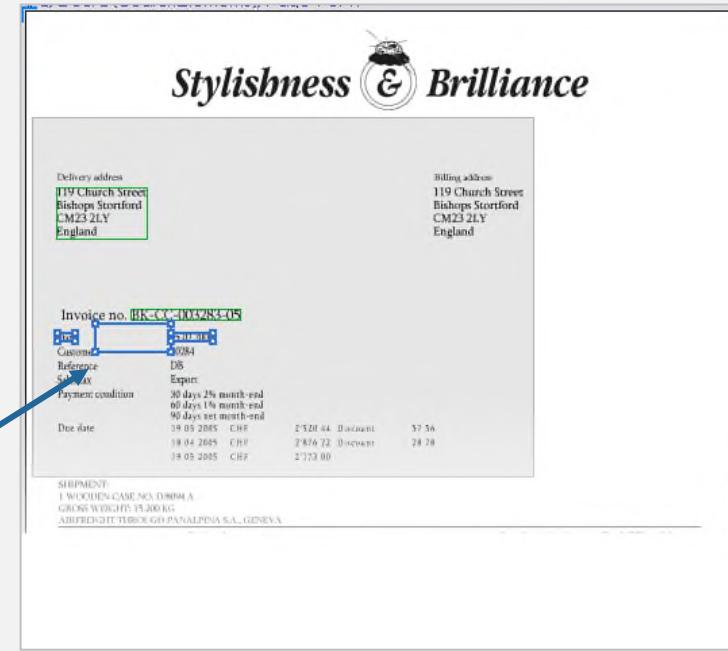
The screenshot shows the ABBYY FlexiLayout Studio interface. On the left is a tree view of search elements for a credit card statement. The root node is 'Credit\_card\_statement\_one\_page', which contains 'Blocks' and 'SearchElements'. 'SearchElements' further contains 'Header' and several text nodes: 'kwCreditCard', 'kwCreditCardTop', 'CreditCardNoTop' (which is selected and highlighted in blue), 'kwMS', and 'CardholderName\_Top'. Below this tree is a properties dialog for 'SearchElements.CardholderAddress'. The dialog has tabs for General, Paragraph, Search Constraints, Relations, and Advanced. The Relations tab is active, showing a table titled 'Relations (Hint):'.

E...	Relation	Reference Element	Offset
<input checked="" type="checkbox"/>	Below	SearchElements.CardholderName	0
<input checked="" type="checkbox"/>	Above	...rchElements.CardholderName.Bottom	-300
<input checked="" type="checkbox"/>	LeftOf	SearchElements.Header.kwStatement	0

## Absolute search area constraints

- Range of pages for searching
- Area in absolute coordinates
- Format: [left, top, right, bottom]
- Several constraints can be set

Absolute search area  
[19, 370, 1787, 1640]

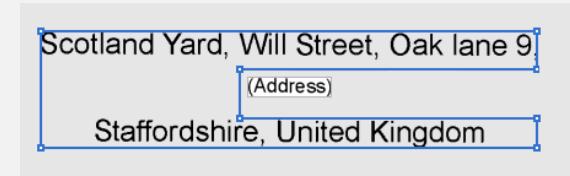


# Excluded Regions

Exclude the region of found element from the search area

Only the elements that located *above the current one* can be excluded

If excluded element is not found, nothing is excluded



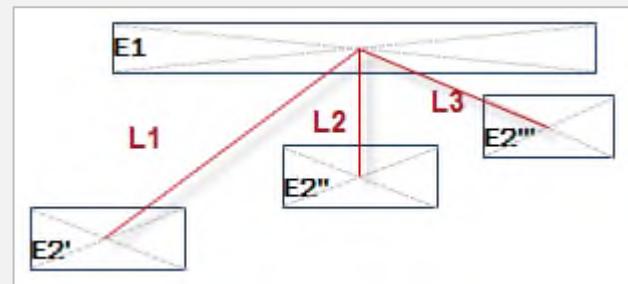
Excluded element  
region

# Relative Search Area: Nearest

ABBYY

The closest object is chosen

The shortest distance between the centers  
is used



**Nearest**

To: **Element**

X:

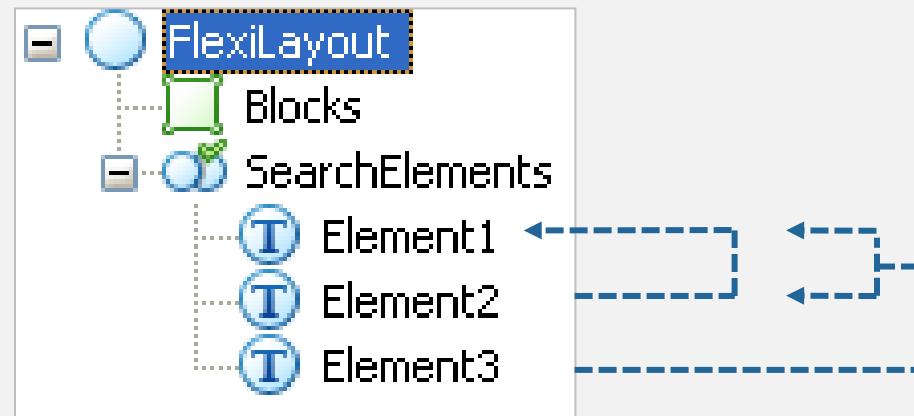
Reference element:

Minimum quality:

- Element
- Element
- X coordinate
- Y coordinate
- Point
- Page top edge
- Page bottom edge
- Page left edge
- Page right edge

# Relative Search Area: Relation

- Sets the object search area in relation to other elements
- Use elements above the current element



# Direction and Borders

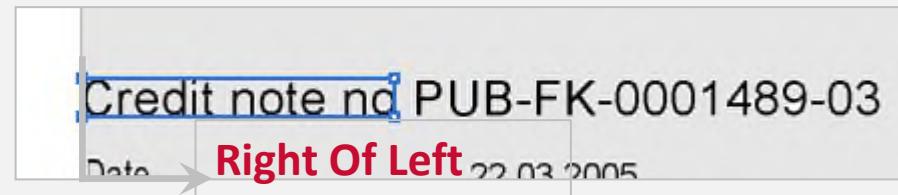
ABBYY®

## Direction

- Left Of
- Right Of
- Above
- Below

## Borders

- Left
- Right
- Top
- Bottom
- Whole

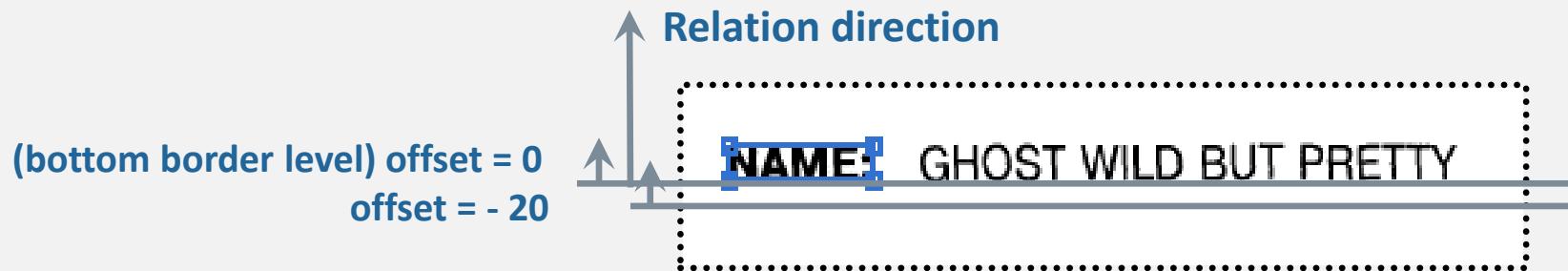


# Negative Offset

*Negative Offset (shift) – shifting of the half-plane opposite the relation direction*

Above ...GuestNameHeader.bottom 0

Above ...GuestNameHeader.bottom -20

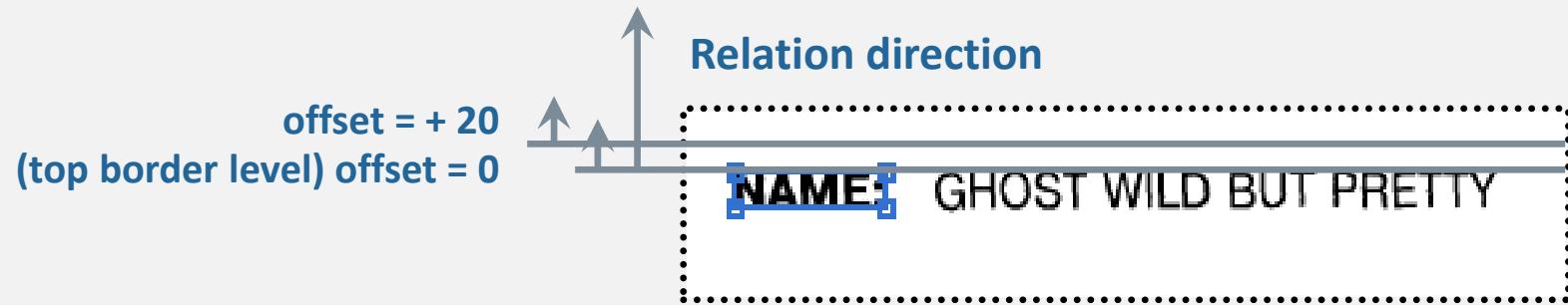


Which of these two offsets is better for production use?

*Positive Offset (shift) – shifting of the half-plane in the relation direction*

Above	...GuestNameHeader.top	0
-------	------------------------	---

Above	...GuestNameHeader.top	20
-------	------------------------	----



Are these settings ideal for production use?

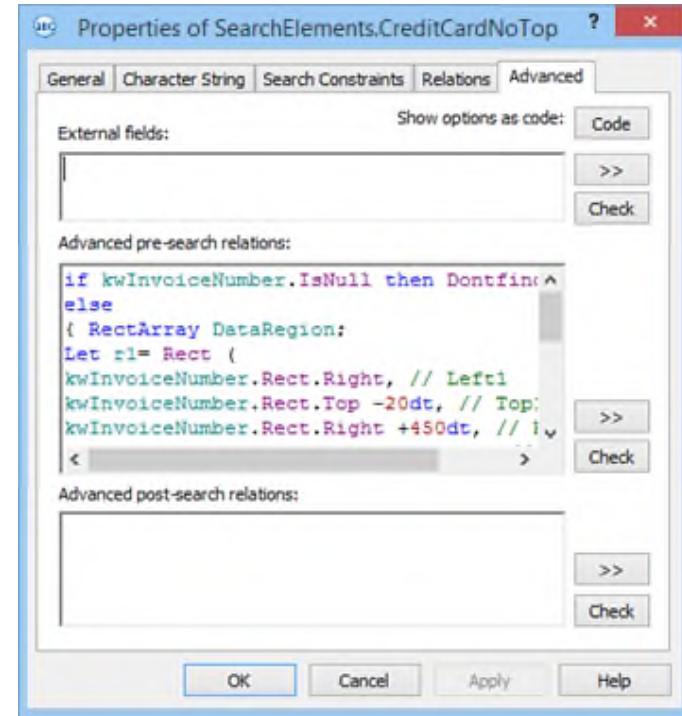
# Additional Search Conditions

## Advanced pre-search relations

- Element properties and search conditions

## Advanced post-search relations

- Hypothesis properties check
- Hypothesis quality settings



Elements are auxiliary objects used for defining block locations (document areas to capture data from)

The element search area can be set by coordinates and relation to other elements

Directions, borders and offsets can be used to specify search area

Element properties allow specifying search criteria for image objects



## ABBYY FLEXILAYOUT STUDIO

### TABLE DETECTION

---



# Overview



Table



Creation steps

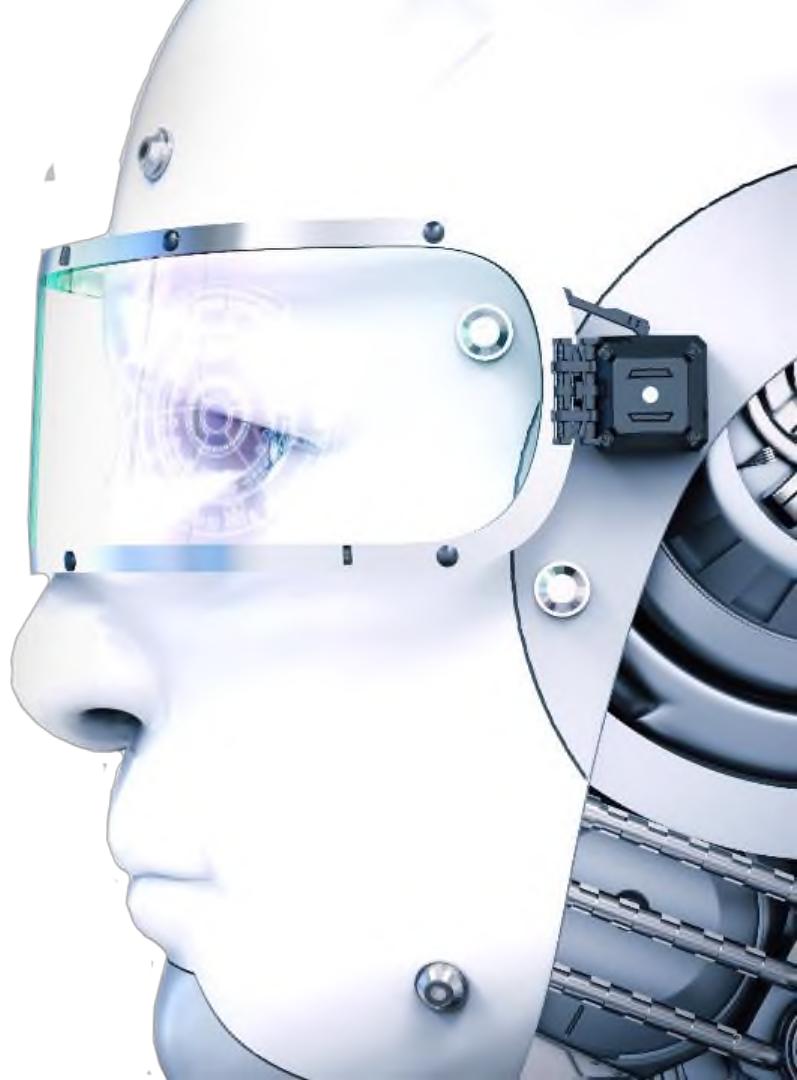


Table Element

- Columns
- Columns order
- Header
- Footer
- Rows
- Search Area



Block Location Description



# Table

ABBYY®



Block



Element

Reference	Designation	Qty	Unit price	TotalCHF	Sales tax
<b>Work:</b>					
SERVICE A	PERIODIC COMPLETE SERVICE	1	1200.00	1200.00	220
<b>Exterior parts :</b>					
PU.256.015.BA	Y.G. PUSHERS	3	55.00	165.00	220
CO.156.598.BA	Y.G. WATERPROOF CROWN	1	52.00	52.00	220
CL.566.123.SA	SAPHIRE CRYSTAL	1	70.00	70.00	220
IA.156.013.NO	PAIR OF HANDS	1	98.00	98.00	220
AI.157.012.BU	STEEL BLUE COUNTER HANDS	3	30.00	90.00	220
<b>Discount</b>					
		30%	-560.00		
<b>Total CHF</b>					
				1115.00	

Header

Body

Footer

# Table Element Creation Steps

## The **Table** block

Table name

Columns names and types

## The **Table** element

- Link with the corresponding Table block
- Each column: header and content
- Order of columns
- Header search parameters (top border)
- Footer search parameters (bottom border)
- Division into rows
- Search area and additional search conditions

## The **Table** block

Location

# Column Order

## Fixed

- All table columns are necessary;
- The program will choose one of the fixed orders

## Rule

- The required columns are *mixed with unnecessary columns*
- All rules function at the same time!

Fixed column order

Order name	Column set
Order1	Reference; Designation; Quantity; U...
Order2	Designation; Reference; Quantity; U...

.....

Reference	Designation	Quantity	Unit price	Total
<b>Work:</b>				
SERVICE A	PERIODIC COMPLETE SERVICE	1	1100.00	1100.00
SERVICE D	REPAIRING THE BEZEL AT BEST	1	110.00	110.00
SERVICE D	REFRESHING COMPLETE CASE	1	160.00	160.00
Exterior parts :				
O.ISS.230.BA	Y.G.WATERPROOF CROWN	1	60.00	60.00

*Use column order from previous instances – for multi-page tables described within repeated groups*

## HEADER

The top border of the table body

Has an auxiliary function, no data  
is extracted

Contains column names

Search according to column parameters

## FOOTER

The bottom border of the table body

Has an auxiliary function, no data is  
extracted

Search:

- By keywords
- By element of *any type*

# Table Rows

## *Automatic row detection:*

- By black separators and white spaces in all columns
- Text line = table row

## *By rows of the Master column:*

- Multiple selection is possible
- Text alignment in fields
- Text line = table row

*Master column  
should contain  
one line of text  
per row*

# Block Location Description

## Source element

- Based on the table element region

## Expression

- Based on several table elements with different block search algorithms

### Example

```
let T1 = SearchElements.TableElement1;  
let T2= SearchElements.TableElement2;  
if T1.RowsCount > T2.RowsCount then OutputTable = T1;  
else OutputTable = T2;
```

*TableElement1  
and  
TableElement2  
must refer  
to the same  
block!*

# Multi-Page Table Types

ABBYY®

## One multi-page table

- the header is on the first page, the footer is on the last page

## One multi-page table

- column names are repeated on each page

## A repeatable table (multiple instances on one page)

To locate the table:

- Create the table block and provide column names and column data types
- Create the table element and inherit the column structure from table block
- Configure the element properties using element property tabs as needed
- Set the table block location - a.k.a. - mapping the table block

# ABBYY®

## ABBYY FLEXILAYOUT STUDIO

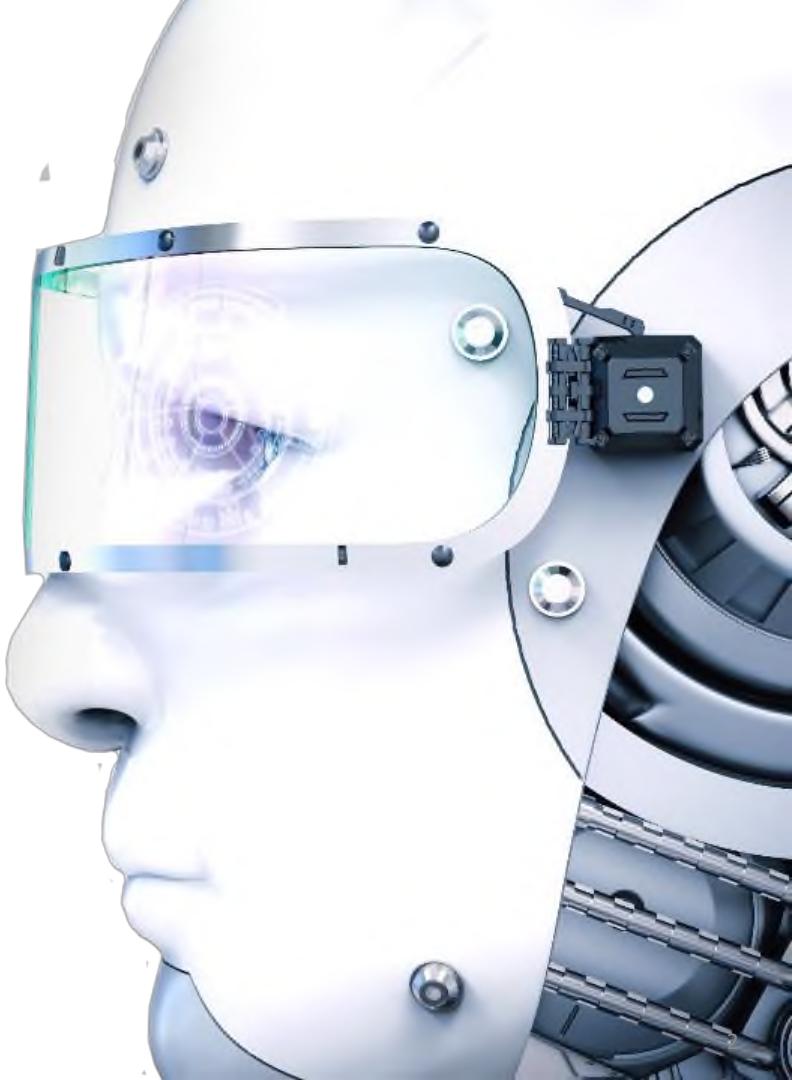
---

### MULTI-PAGE DOCUMENTS



# Overview

- Multipage Support
- Multipage Parameters
- Reference Assembly
- Document Assembly



*Header\Footer*

Multi-page parameters

Multi-page  
element/block areas

## INTERFACE:

Reference assembly

View all pages at once

Page parameters

Must contain at least one *Required* element

*Header* defines the beginning of multi-page document



*Footer* defines the end of multi-page document  
*Can be optional*

Specify the range:

- Minimum number of pages
- Maximum number of pages

Manage assembly using the FlexiLayout language

- *SetPageCount* function in the Header and Footer elements
- Returns the current number of pages in the document

# Multiple FlexiLayout Matching

ABBYY

## FlexiLayout matching:

The selected pages are pre-recognized

The first page of the document is defined

The last page of the document is defined

The template is matched with found pages

#	Page No.	Matched	Matched	Matched
1	1	✓	✓ 1 (H) (F)	
2	2	✓	✓ 1 (H)	
3	3	✓	✓ 2 (F)	
4	4	✓	✓ 1 (H)	
5	5	✓	✓ 2	
6	6	✓	✓ 3 (F)	
7	7	✓	✓ 1 (H)	
8	8	✓	✓ 2	
9	9	✓	✓ 3	
10	10	✓	✓ 4 (F)	

*Header* is used to detect  
the first page

*Required Footer* is used  
to detect the last page



No Header, but:

- Required Footer
- Optional Footer:
  - if found => it will be used
  - if not found => the top page limit will be used



There is neither  
a Header nor a Footer

- top page limit will be used  
(multi-page parameters)



# Summary

ABBYY®

## FlexiLayout for multi-page documents:

- Header\Footer
- Multipage parameters
- Multi-page element/block areas

## Document assembly:

- Multi-page parameters:  
from 1 to N
- Header and Footer  
(presence\absence)

FlexiCapture first of all matches the Layouts with Header element



## ABBYY FLEXICAPTURE

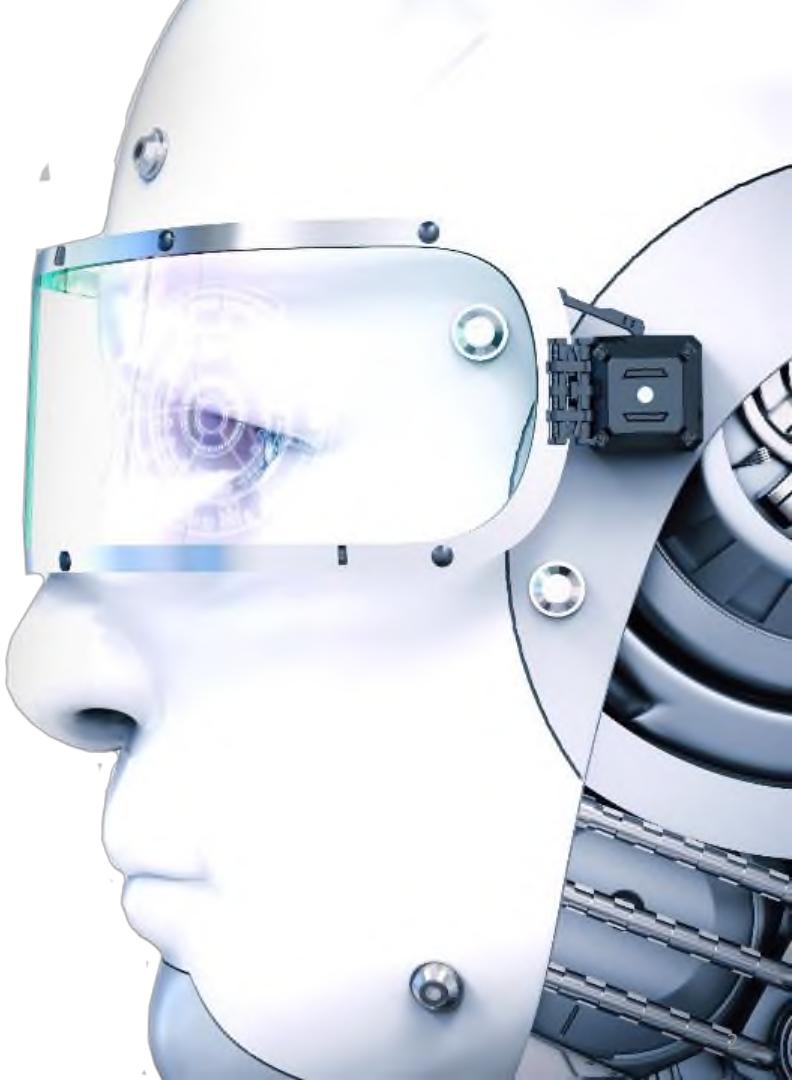
---

### TRAINING IN FLEXICAPTURE



# Overview

- Introduction
- Setup a Project by Training
- Add Training to Existing Project
- Auto-learning



# INTRODUCTION

---



# Training of Standalone Fields

- Example: invoice date, invoice number
- Enable *Allow field extraction training* in the section's properties
- Training works from 1 document, 3-10 is recommended



Program creates  
FlexiLayout  
to detect field  
according  
to its surrounding  
area

# Training of Standalone Fields

ABBYY®

Program creates FlexiLayout to detect field according to its surrounding area

The image shows a document with several text blocks highlighted by different colors:

- T. van LEFT D. Heid**  
GmbH & Co. KG
- SCHRAUBEN- UND  
MUTTERNFABRIK  
UMFORMTEILE
- Abnahmeprüfzeugnis 3.1 nach EN 56789**  
Inspection Certificate 3.1
- Zeugnis-Nr. / certificate No.: 1789555**
- Auftrags-Nr. / order No. + Pos:** **Alphabet Match**
- Auftragsdatum / order date: 15.07.15**
- Werks-Nr. / work No. + Pos: 782256.3**
- F. ReyherYtrgf. GmbH&Co. KG**  
Postfnnr 33 22 66  
98100 Hamburg
- Kennzeichnung / identificationmark:**  
**LEFT 10.45**
- Prüfbedingungen / condition of test:**  
**EN ISO 898-1:2012**

Red arrows point from the "Alphabet Match" label to the "Auftrags-Nr." field and the "Werks-Nr." field, indicating that the program uses contextual matching to identify specific fields.

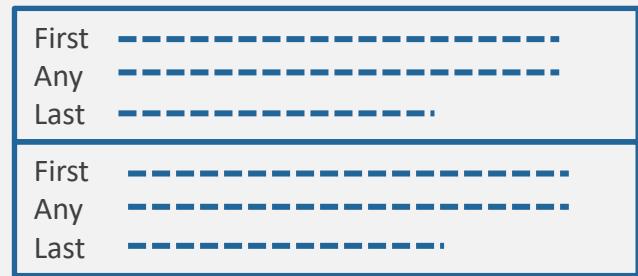
# Training of Repeating Groups

Program indicates rows in table

3 types of lines are recognized inside the row:

- First line
- Any line
- Last line

Each line is recognized according to surrounding area (like any standalone field)



# Semi-Structured Documents – Table Training

ABBYY®

*Supplier*

ITV Business Service Centre  
PO Box PO BOX 429  
Manchester  
Greater Manchester  
M60 1DX

Tel:  
Fax:  
Email:

02 AUG 2011

**Invoice**

Information	
Invoice number	90186336
Invoice Date	29.07.2011
Customer No.	1015808
Customer Order No.	2000317339
Delivery Note No.	80290714
Delivery Note Date	29.07.2011
Currency	GBP
Invoice Amount	102.84
Payment terms	Proforma invoice

Page 1 of 1

Item	Article	Description	Quantity	Unit price	Value
1	10 101093	S 102 Z052-130+	5 PCE	15.14 GBP	75.70
any	last	Quantity ordered:	5 PCE		
any	last	Quantity to follow:	0 PCE		
1	50 CARRIAGE 1 UK Delivery		1 PCE	10.00 GBP	10.00
any	last	Quantity ordered:	1 PCE		
any	last	Quantity to follow:	0 PCE		
		Sub Total			85.70
		VAT	20.000 %		17.14
		Invoice Amount			102.84

# Training if Fields Locations Vary Greatly

ABBYY®

If fields locations vary greatly within one document type:

1. Differentiate document variants

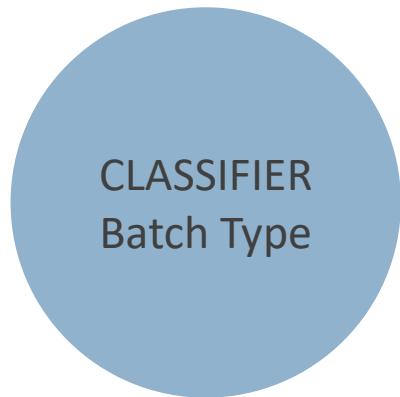
Document Definition has special data set *Variants*

2. Train Flexilayout for each variant



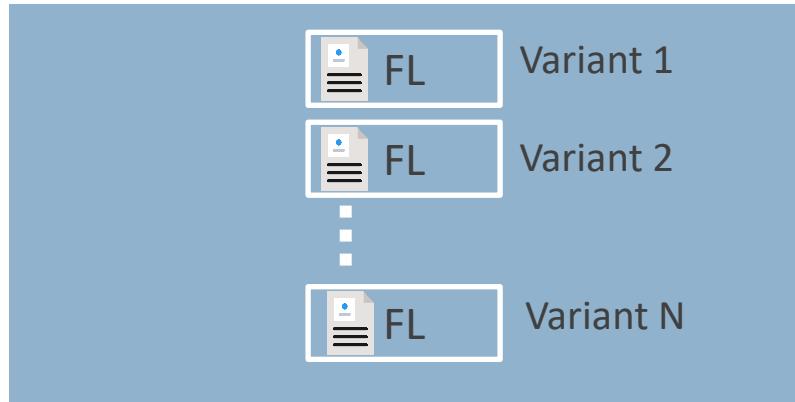
The result  
of training  
is a Flexilayout  
for standalone  
fields + repeating  
groups

# Training if Fields Locations Vary Greatly

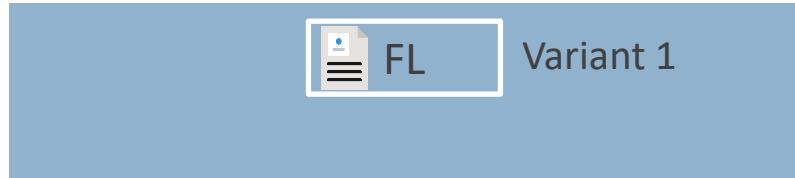


Document Definition 1  
Variant 2

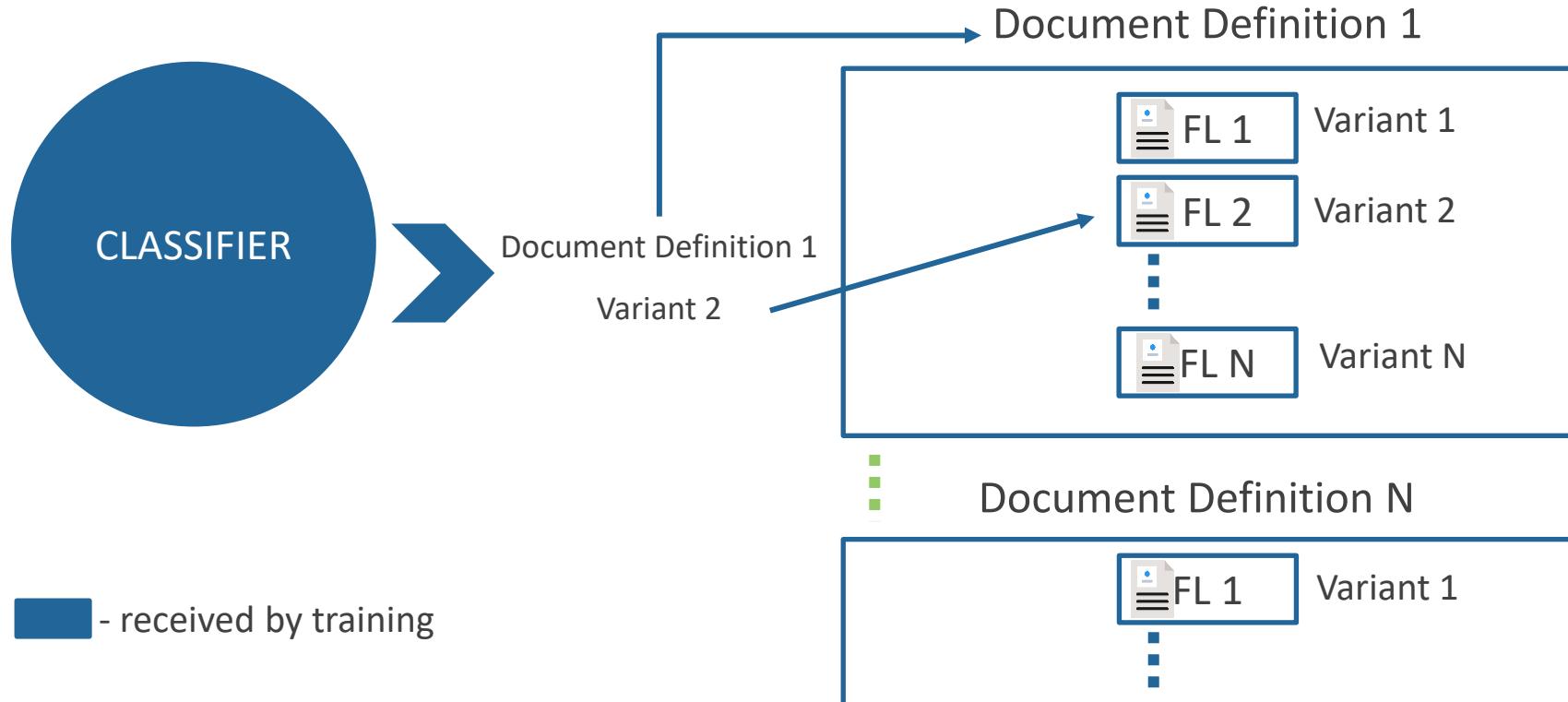
Document Definition 1



Document Definition N

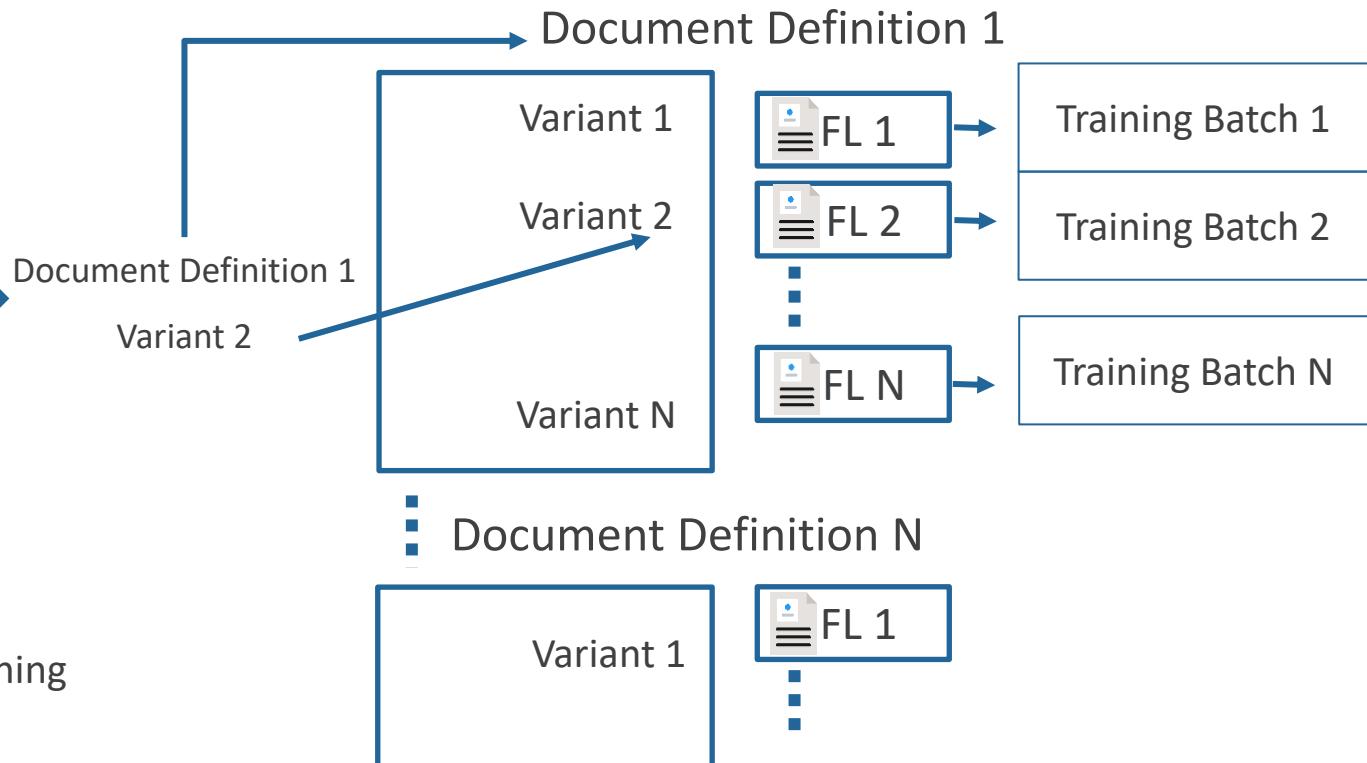
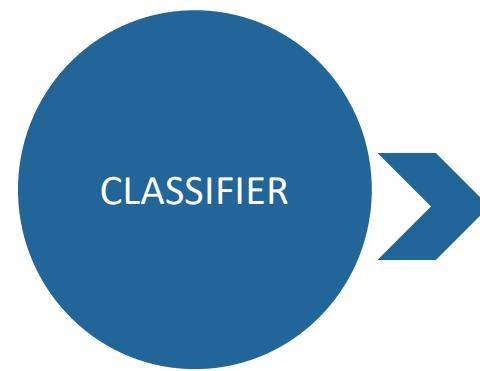


# Training if Fields Locations Vary Greatly



# Training if Fields Locations Vary Greatly

ABBYY®



# SETUP PROJECT BY TRAINING

---





- Simple project can be adjusted to be demonstrated to customer
- Such projects can be used as proof of concept or – after testing – in production

# Setup Project with One Variant

1. Create Document Definitions
2. Enable the *Allow field extraction training* option for the Document Definition section
3. Create necessary fields without a region
  - Choose the *Can have region* option in the field properties
4. Create a new batch in *Field Extraction Training Batches*
5. Select the Document Definition to be trained
6. Select *Default Variant* from list of variants
7. Load images and recognize
8. Correct field locations
9. Train the Field Extraction Training Batch

# Setup Project with Multiple Variants

1 Add variants to document sections of Document Definitions

2 Train the classifier:

1. Open the *Classification Training Batches* view and load images to new batch
2. Assign reference classes to the documents

Each variant of the document section can be used as a separate class

3. Train the Classification Training Batch

3 Train ABBYY FlexiCapture to detect field regions on each variant:

1. Open the *Field Extraction Training Batches*
2. Create a new batch
3. Select the Document Definition section whose fields the program will be trained to detect
4. Select the variant whose fields the program will be trained to detect
5. Load and recognize images
6. Correct field locations
7. Train the Field Extraction Training Batch

# MANUAL DEVELOPMENT

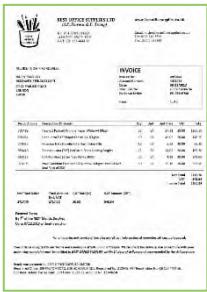
---



# Manual Development in FC11

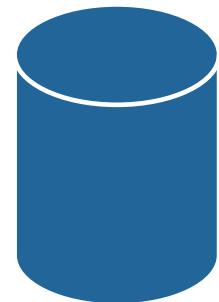
Document Type  
Identification

Fields Extraction



Fixed Template / Flexilayout

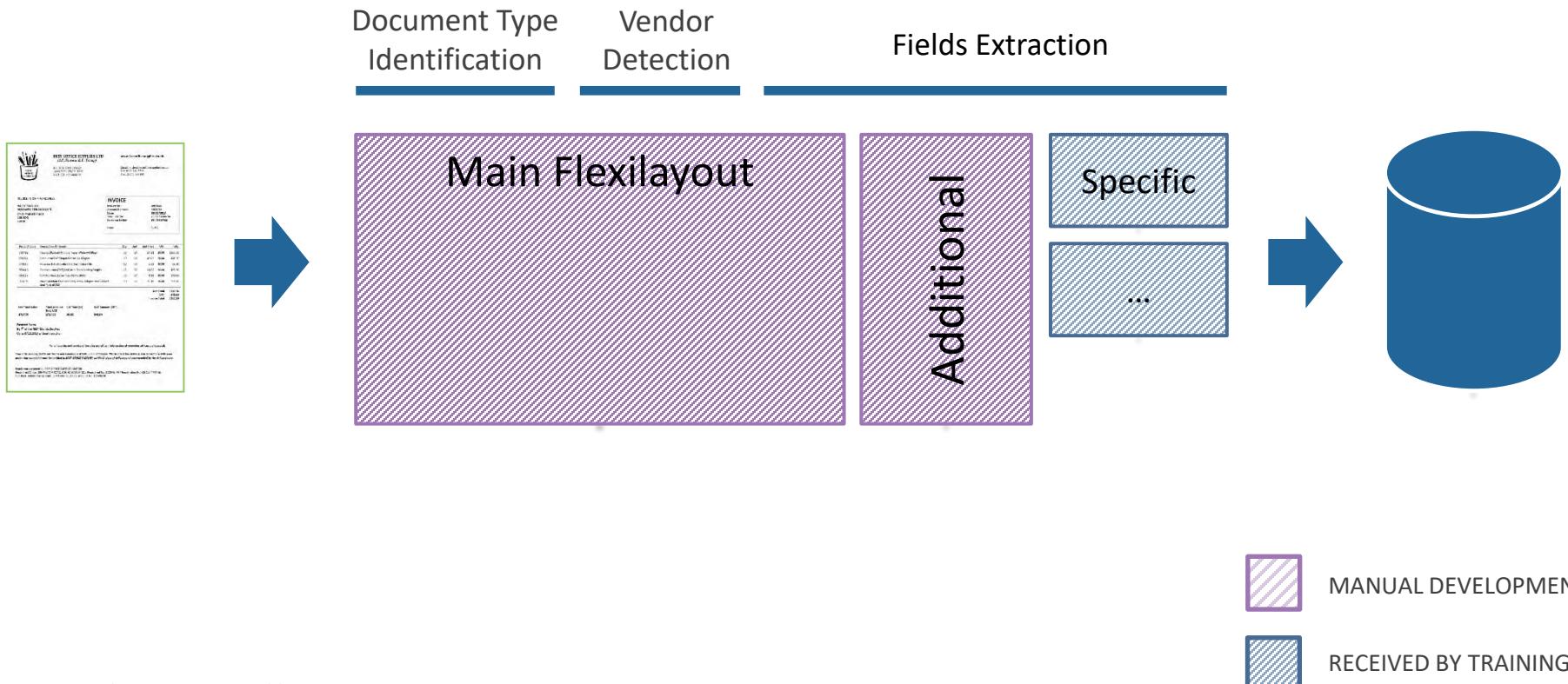
Header – Body – Footer



MANUAL DEVELOPMENT

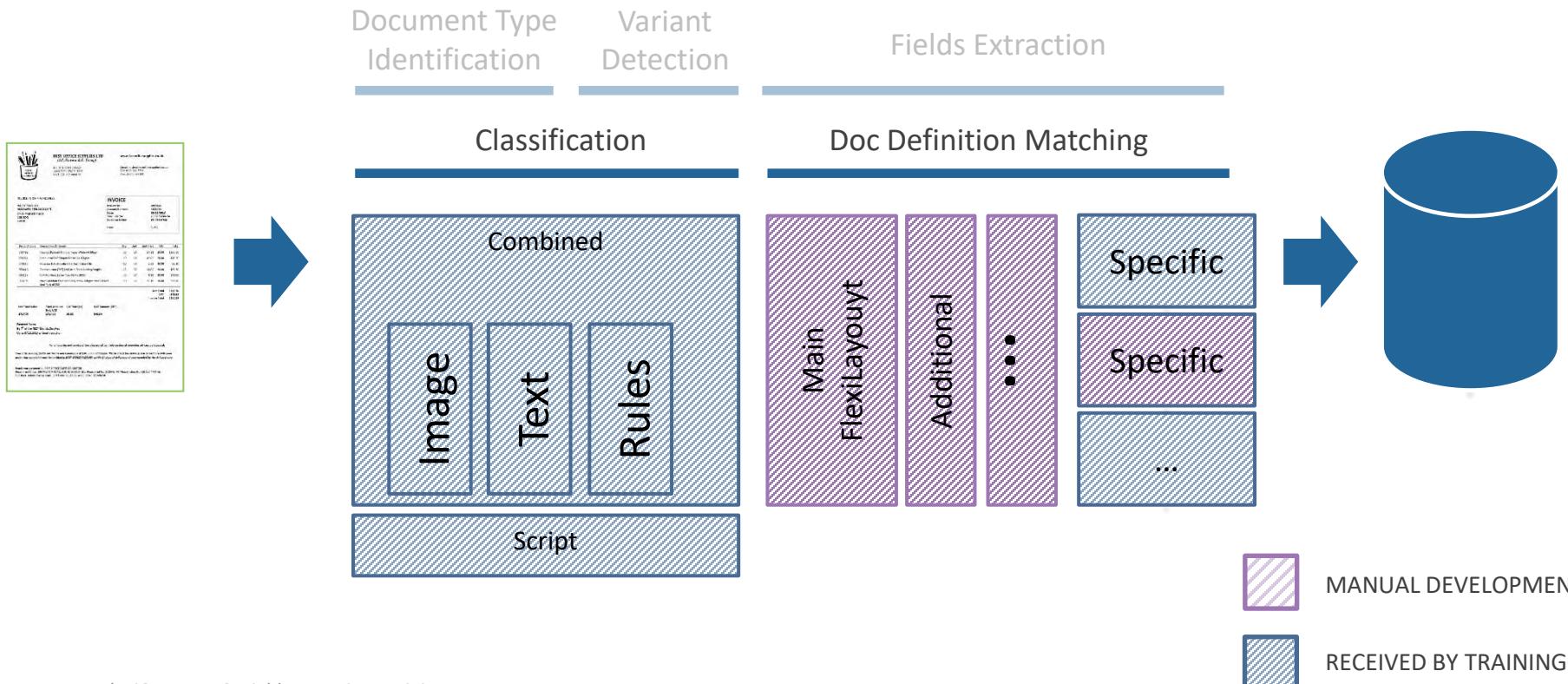
# Manual Development in FCI

ABBYY®



# Manual Development in FC12

ABBYY®

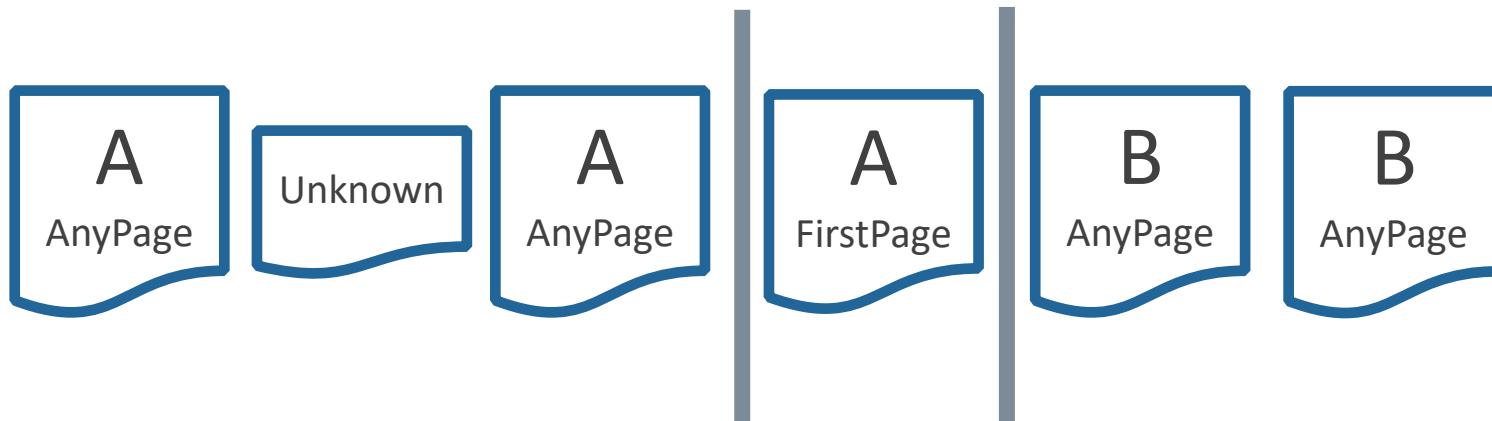


# Identification and Assembly while Training

Trained classifier identifies and assists in assembling documents

Class can correspond to one or several sections

Class can correspond to the first page or any page of the section



# Identification and Assembly with Manual Development

First page is classified, FlexiLayout is selected and applied

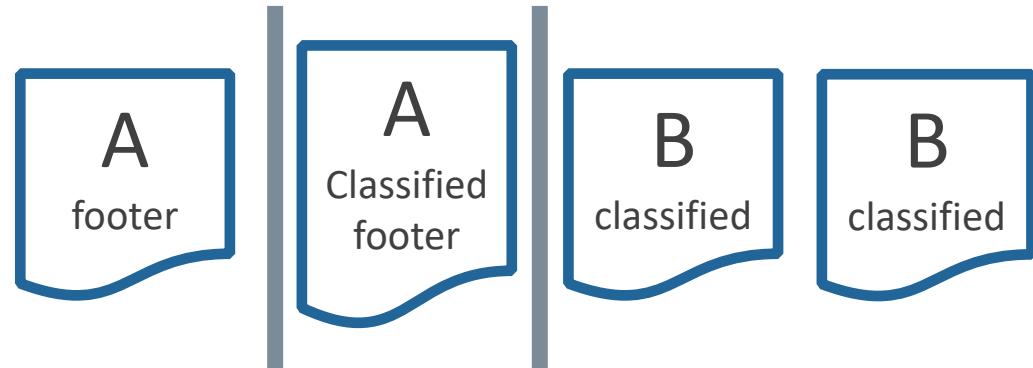
FlexiLayout has *no* obligatory footer

next page is classified to define if the FlexiLayout  
still can be applied



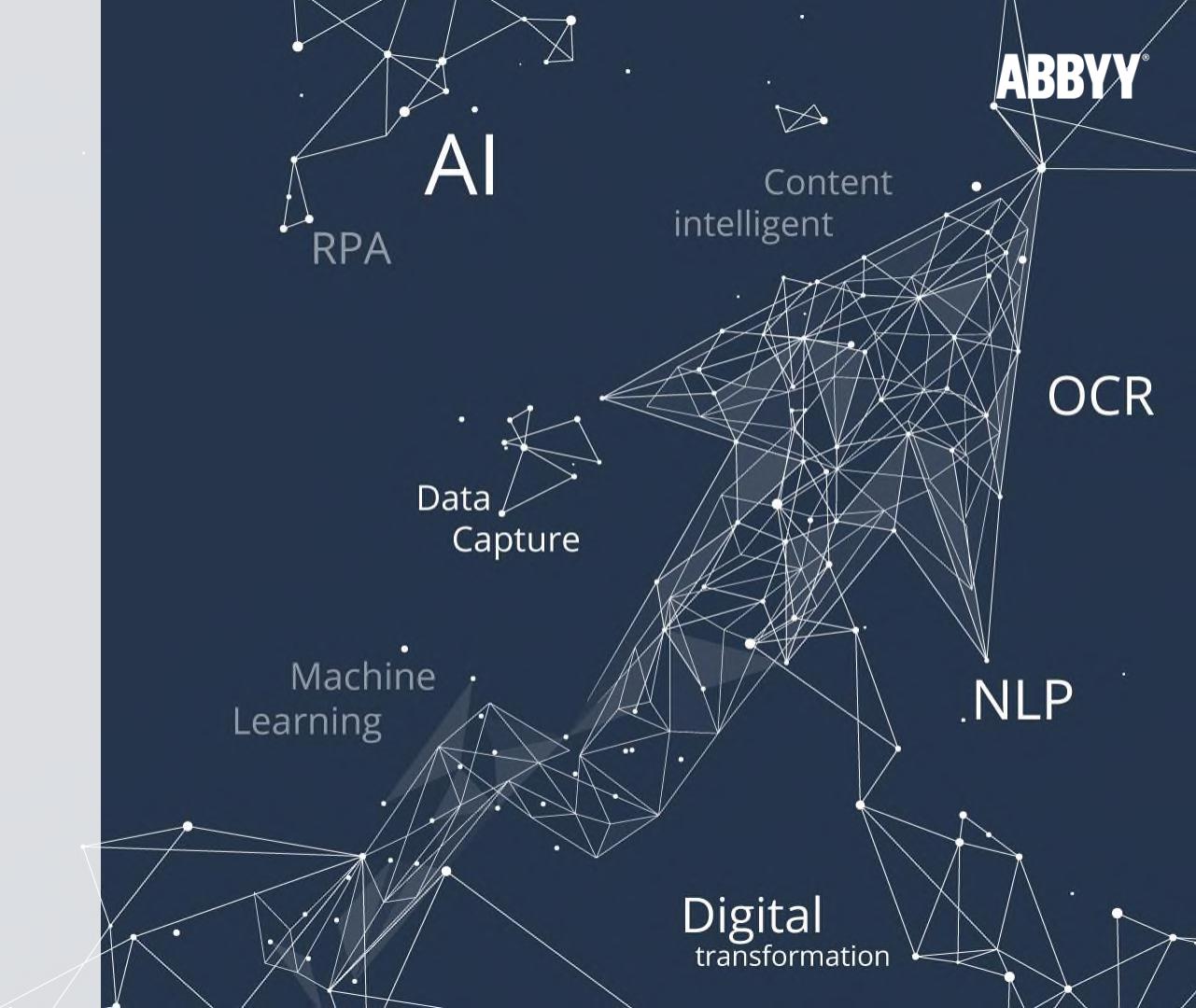
FlexiLayout has *obligatory* footer

FlexiLayout is matched until it encounters a  
footer



# AUTO-LEARNING

---



# Auto-learning

System learns from user's input

---

Auto-learning is applied only after documents export

---

Section auto-learning can be locked

---

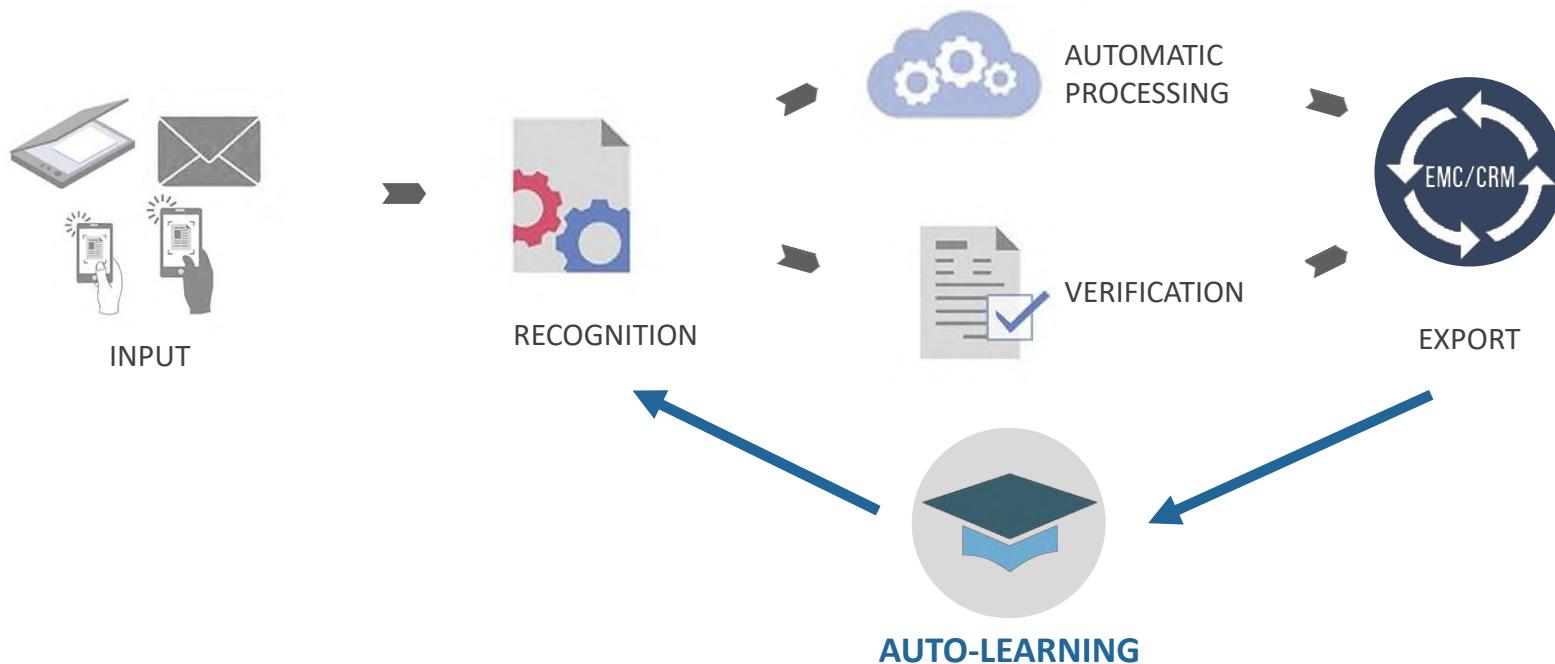
Variant auto-learning can be locked from batch properties

---



*Allow field extraction training option in section's properties*

# Auto-learning



# Summary



All types of fields can be trained:

- Standalone fields
- Repeating groups
- Fields whose locations vary greatly



Classifier is trained to detect variants



Auto-learning is applied after document export



## ABBYY FLEXICAPTURE

---

### SYSTEM SETUP: QUICK START



# System Setup: Main Steps



Creating new DB/connecting to existing DB

Creating new project

Uploading project to server



Assigning roles and rights to users/processing service

4

Setup Processing Server

- Add processing stations
- Start Server and Stations

5

Start working

Administration and Monitoring Console

Project Setup Station

Administration and Monitoring Console

Processing Server

Different Operator Stations

# 1. Database Settings

FlexiCapture Application Server settings => Database creation

Users  
Groups  
Requests

Application Server

SMTP settings  
Email notifications  
Event logging mode  
Cleanup

Create new database

MS SQL  
 Oracle  
 Azure

Server instance name (MS SQL): BAIKAL-SQL\SQL2014

Database name: Smith\_1234

Server access authentication

Authentication settings: Windows Authentication

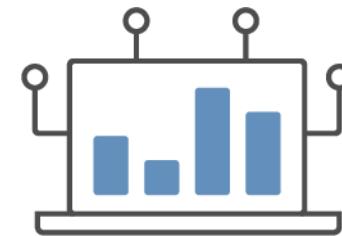
User name: (redacted)

Password: (redacted)

Use external file storage

Path: C:\FileStorage

OK CANCEL TEST CONNECTION



Administration and Monitoring Console

## 2. Creating a Project

ABBYY®

1

PROJECT

2

PROJECT SETUP

3

UPLOAD PROJECT TO  
SERVER

- 
- Create new
  - Open an existing one
  - Document Definitions
  - Batch types
  - Processing options
  - Application Server address
  - Tenant name
  - Credentials

### 3. User Roles and Permissions

- One account after installation:  
*Administrator* who installed servers and created database
- Add users and grant them access rights



Administration and Monitoring Console

<b>NEW USER</b> <b>IMPORT...</b> <b>DELETE</b> <input type="text"/> <b>Search</b>				
Showing: 1 to 2 of 2				
<input type="checkbox"/>	Full name	E-mail	Groups	Permissions
<input type="checkbox"/>	Login			
<input type="checkbox"/>	<a href="#">WORKGROUP\FC12-VM\$</a>			Processing Server
<input type="checkbox"/>	<a href="#">fc12-vm\student</a>	Student	<a href="#">s@m.com</a>	Administrator

# 3 (1). Processing Server – Account

By default “*Network Service*” account is used for Processing server and stations

Specify the account to run the *Processing Server* at the *Administration and Monitoring Console*

The screenshot shows the user configuration interface for the account 'fc12-vm\student'. The 'Login' field contains 'fc12-vm\student', 'Full name' is 'Student', and 'E-mail' is 'student@fc12.com'. A 'RESET PASSWORD' button is present. Below these fields is a section titled 'Belongs to groups' with an 'Add' link and a note 'No groups'. Under 'Permission', there are two sections: 'Roles' and 'Projects'. The 'Roles' section lists several operator roles with checkboxes: 'Scanning Operator', 'Data Verification Operator', 'Verification Operator', 'Senior Verification Operator', 'Processing Server' (which is checked), 'Project Settings Editor', 'Monitoring Operator', 'Administrator' (which is checked), 'External User', and 'Web Capture Operator'. The 'Projects' section notes that the 'Processing Server' role is given to all projects.

Roles	Projects
<input type="checkbox"/> Scanning Operator	The role given to all projects
<input type="checkbox"/> Data Verification Operator	
<input type="checkbox"/> Verification Operator	
<input type="checkbox"/> Senior Verification Operator	
<input checked="" type="checkbox"/> Processing Server	
<input type="checkbox"/> Project Settings Editor	
<input type="checkbox"/> Monitoring Operator	
<input checked="" type="checkbox"/> Administrator	
<input type="checkbox"/> External User	
<input type="checkbox"/> Web Capture Operator	

## 4. Setup Processing Server

1

ADD PROCESSING  
SERVER AT *PROCESSING  
SERVER MONITOR*

2

SPECIFY APPLICATION  
SERVER LINK

3

ADD PROCESSING  
STATIONS

4

START SERVER  
AND STATIONS

## 5. Start Processing

The system is ready to process documents!

