

Session Title: Digital Transformation Starts at Capture: Converting Paper into AI-Ready Data

Session Description: Forward-thinking organizations recognize the untapped potential within their paper archives - both as valuable training data for AI and as a way to reduce physical storage costs. Yet the key to unlocking this value lies in getting data right from the start. This session reveals how intelligent capture workflows not only prevent costly downstream errors but transform decades of business information into AI-ready digital assets. By streamlining your capture workflow, you can validate data accuracy at the point of entry, eliminating expensive cleanup efforts while boosting productivity across your organization. Through practical examples, we'll explore proven capture solutions that address common business challenges like invoice processing, customer onboarding, and compliance documentation. By the end of this session, you'll have actionable strategies to harness the hidden value in your paper archives and build a capture foundation that delivers the clean, reliable data your AI initiatives need to succeed.

Digital Transformation Starts at Capture

Converting Paper into AI- Ready Data

April 2024



Poll Question: What industry segment are you from?

Government – (FED, ST, or Local)

Healthcare

Financial

Education

Manufacturing

Other



*"Study the past if you
would define the future"*

Confucius

c.5th Century BCE





The Promise of the Future...
Requires Data
(Accurate Data!)



Best Practices for Data Hygiene

Capture your data with the correct equipment & settings

Identify and remove redundant data from file archives

Classify & index your data

Implement metadata for access rights

Anonymize / Redact Sensitive Information

Chain of custody – Who, Where, & When, and How

Document Relevance/Ranking



Poll Question: Is your company or organization storing paper records?

Yes, we're storing them on-site

Yes, we're storing them off-site

Yes, we're storing them on-site and off-site

No, we're 100% Digital



The Cost of On-Site Document Storage



City	Average Cost (sq ft)	File Cabinet 44" x 21"
San Francisco	\$61	\$394
Seattle	\$38	\$246
Chicago	\$34	\$219
Dallas	\$27	\$172
Atlanta	\$27	\$173
New York City	\$80	\$514

What is the actual cost?

Storage

Retrieval

Risk

Insite

Poll Question: How much time do employees spend searching for data? (40 hour work week example)

4 hours

10 hours

20 hours

More than 20 hours



Poll Question: On average, what percentage of business decisions are based on data?

75-100% (Almost all of the time)

50-74% (Usually)

25-49% (Occasionally)

0-25% (Seldomly)



Is Your Company Making Data-Driven Decisions?

58% of respondents to a BARC survey¹ showed that managers are making decisions “On the cuff” or based on a “Gut Feel”

When your employees can't find data, they're left to their own devices to compensate.

This can cost your company money, drain productivity, and cause a loss of customer trust.

Not being able to find accurate data quickly is the top reason employees bypass it.



¹<https://barc.com/business-decisions-gut-feel/>

The Mystery of Dark Data

Data that you're storing, but not using

Unstructured and unclassified...and a potential value or risk to your organization

Includes data from employees who have left the company or moved departments

It could be 50% of an organization's data...or more. How much do you have?

How much of that data is on paper? In network file shares? In legacy systems?

Data is the key to future business success. Organizations that can harness dark data will have an advantage over their competition.



Poll Question: On average, how much does a data error cost an organization?

\$25 (About a dozen eggs)

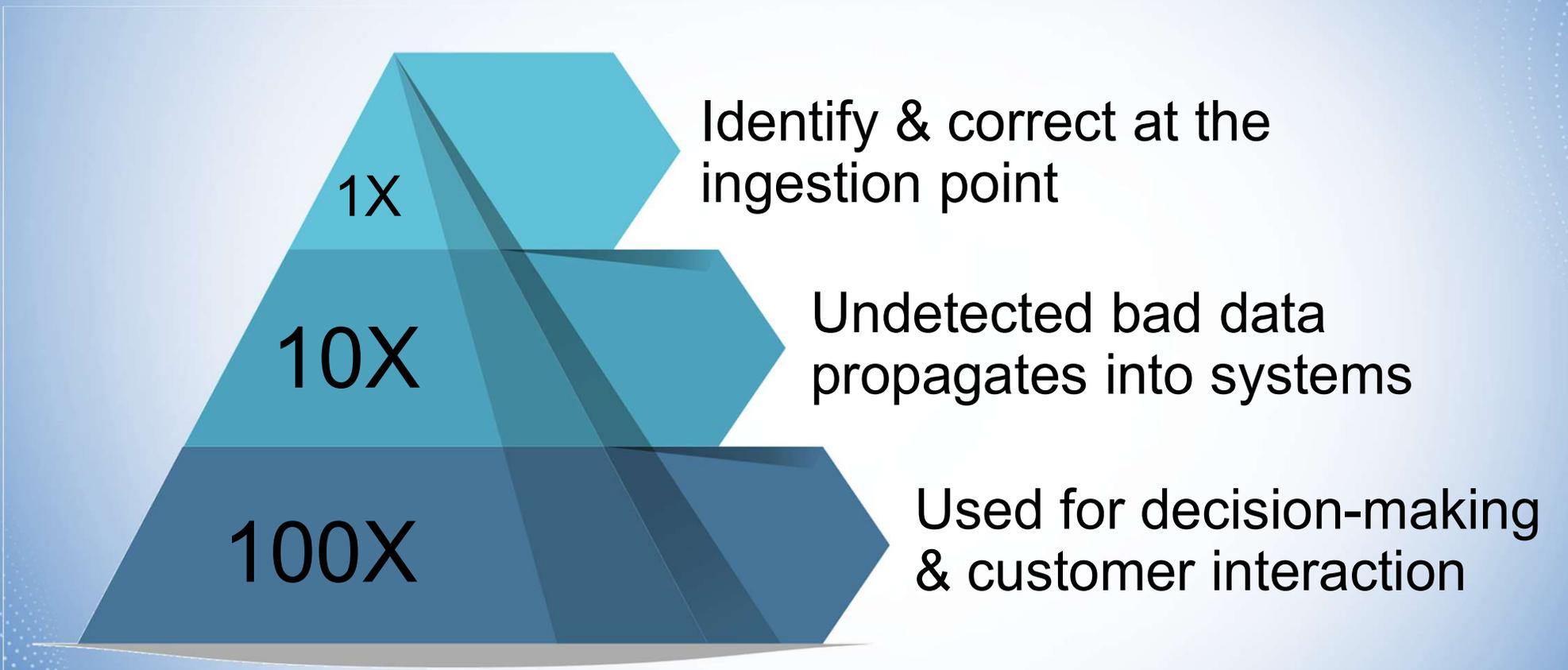
\$50 (Half tank of gas)

\$75 (2 drinks & a hot dog at the game)

100\$ or more!



The Rule of 1:10:100 - The Cost of Data Errors



1X

Identify & correct at the ingestion point

10X

Undetected bad data propagates into systems

100X

Used for decision-making & customer interaction

The True Cost of Data Errors For Scanning Capture Workflows

Significant productivity loss, especially if the error is detected downstream

AI data hallucinations

Customer dissatisfaction and loss of trust / reputation

Regulatory fines for PII breach

Bad business decisions based on inaccurate data



“Conversion projects are labor intensive and take much longer than expected”



These are some of the challenges with document scanning workflows and archive projects



Pre- and post-scan processing is labor-intensive and costly



Manual classification and data entry waste time, resources, and are error-prone.



Forms are data-intensive and may have sensitive data



Workflow bottlenecks can stall information processing

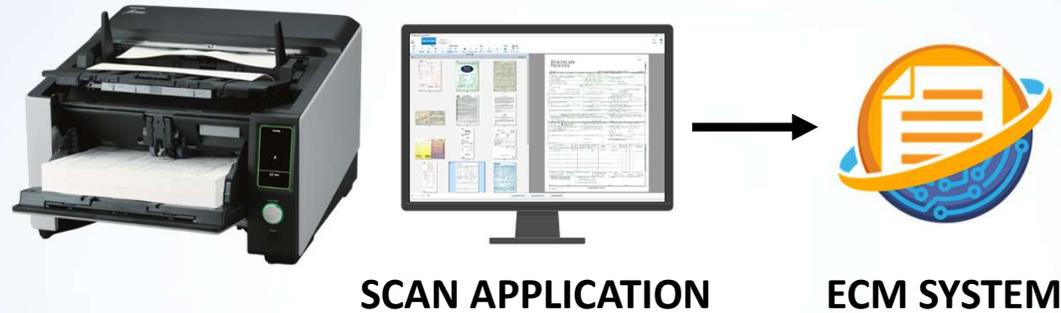


Capture solutions can be cost-prohibitive



Not all capture devices are optimized for high-quality image output and volume scanning

Capture Workflow Bottlenecks Can Stall Information Processing



- Too many steps for scanner operator
- Potential for errors
- Reduces throughput
- Hinders customer engagement



- 1 SCAN
- 2 QC
- 3 Index
- 4 Name File
- 5 Choose File Location

Poll Question: What type of devices do you use to scan paper documents?

Document Scanners

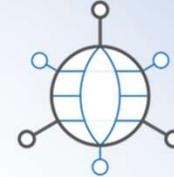
MFPs

Both document scanners & MFPs

None of the above



Components of an Intelligent Capture Workflow



Best in Class Scanning

Ricoh Document Scanners combine excellent paper handling and stellar image quality to ensure that documents are quickly and reliably digitized

Intelligent Image Processing

Automated document processing removes image artifacts and background noise to improve recognition accuracy & enhance the presentation of images

Advanced File Separation

Versatile multi-page document separation capabilities maximize scanner ADF capacity & overall throughput by reducing manual processing requirements.

Automated Data Extraction

Utilize a host of different methods to classify & index documents to extract information and accelerate retrieval in content management systems.

Effective Quality Assurance

Streamline QA workflow with an easy-to-use visual UI assisted by innovative features to highlight & correct difficult images.

Multi-Station Architecture

Scale the size, speed, and capacity of the capture solution by distributing the scanning, indexing, and QA workflow over multiple stations.

Seamless Integration

Multiple integrations with content management and cloud service solutions allow for fast and easy transfer of documents & metadata.



The Right Tool for the Right Job!



Paper
Handling



Image
Quality



Data
Integrity



Reliability



Easy
to Use



Automatic

Image Enhancement Produces Clean Data



	Original Document	Ricoh fi Series
Reverse type		
Halftone removal		
Watermark removal		
Stamp removal		

Original Document

I AM APPLYING FOR A (check any th
 Learner Permit ID card

IDENTIFICATION INFORMATION Do
Driver license? Yes No
Learner permit? Yes No
Non-driver ID Card? Yes No

FULL LAST NAME
John
FULL FIRST NAME
Adam
FULL MIDDLE NAME
Smith

Low Volime Scanner fi-5650C
First In Warranty Program Incl
NBO/2PM/Consumable/
Partes/Travel/Labor
Cover : 1 Year

will also the Fund is used for organ
and transplant rese
t below) educational projects p
organ and tissue donation.

nsidered to have decided not to registe
like the Board of Elections to be notified
dy Registered/I do not want to notify.

Ricoh fi Series

I AM APPLYING FOR A (check any th
 Learner Permit ID card

IDENTIFICATION INFORMATION Do
Driver license? Yes No
Learner permit? Yes No
Non-driver ID Card? Yes No

FULL LAST NAME
John
FULL FIRST NAME
Adam
FULL MIDDLE NAME
Smith

Low Volime Scanner fi-5650C
First In Warranty Program Incl
NBO/2PM/Consumable/
Partes/Travel/Labor
Cover : 1 Year

will also the Fund is used for organ
and transplant rese
t below) educational projects p
organ and tissue donation.

nsidered to have decided not to registe
like the Board of Elections to be notified
dy Registered/I do not want to notify.

AUTOMOBILE CLAIM REPORT
Agent name: _____
Date: _____
Vehicle: _____
Insured: _____
City: _____
State: _____
Zip: _____
PFI # _____



1

184756382624 300 Billing Copy
Payee Name: _____
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____



2

HEALTHCARE PROVIDER
Patient Name: _____
DOB: _____
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____

3

City Council
Name: _____
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____

4

184756382624 300 Billing Copy
Payee Name: _____
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____



5

GLORTEX
BILL OF LADING
Part of Oakland
Juff Barlow 540 555 0122
Oakland
Juff Barlow 540 555 0122
PFI # _____



6

XYZ Insurance 12345
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____

7

PROCESSED HERE
John Doe
PFI # _____

8

Warranty Registration Card
Name: John Mung
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____

9

BUSINESS REPLY MAIL
MAIL STOP 0-0
PFI AMERICA INC
808 E ANCHER AVENUE
SUNNYVALE CA 94089-8000
PFI # _____

10

AUTOMOBILE CLAIM REPORT
Agent name: _____
Date: _____
Vehicle: _____
Insured: _____
City: _____
State: _____
Zip: _____
PFI # _____



11

Packing List
Item # _____
Description: _____
Quantity: _____
PFI # _____

12

DentalPlan
Patient Name: _____
DOB: _____
Address: _____
City: _____
State: _____
Zip: _____
PFI # _____

13

IMPORTANT INFORMATION
When completing this form, please read the instructions carefully.
PFI # _____

14

REPORT OF TRAFFIC ACCIDENT
Occurring in California
PFI # _____

15

AUTOMOBILE CLAIM REPORT
Report Form
PFU

1

14975L382024 300 Billing Copy

From: John Doe, 123 Main St, San Francisco, CA 94102
To: John Doe, 123 Main St, San Francisco, CA 94102

2

HEALTHCARE PROVIDER

3

City Courier

4

14975L382024 300 Billing Copy

From: John Doe, 123 Main St, San Francisco, CA 94102
To: John Doe, 123 Main St, San Francisco, CA 94102

5

PFU BILL OF LADING

6

XYZ Insurance

7

ENDORSEMENT

8

Warranty Registration Card

9

BUSINESS REPLY MAIL

10

AUTOMOBILE CLAIM REPORT

11

Packing List

INVOICE #	QTY	DESCRIPTION
12345	1	Item 1
12345	2	Item 2
12345	1	Item 3

12

Denial Plan

13

DMW

14

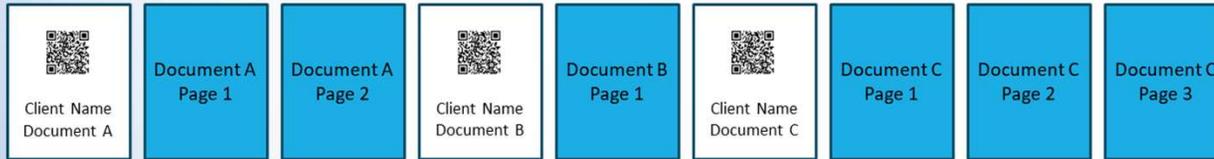
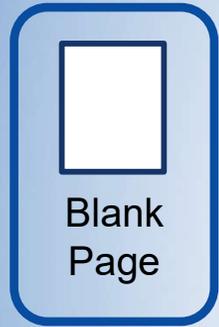
REPORT OF TRAFFIC ACCIDENT

15



After Scan Correction

Automatic Document Separation Saves Time & Accelerates Capture Workflow



Automatically separate and name multi-page files & storage file folders to accelerate capture workflow.





Form Separation Demo

Poll Question: How does your company or organization add metadata to your scanned documents?

Manually entering the data (e.g. Key from image)

No metadata, we're just creating searchable PDF files

Using automation (Zonal OCR, IDP, RPA, etc.)

We're not scanning documents





Key From Image



Advanced Zonal OCR



Handprint Field OCR



Choice List



Checkmark



1D/2D Barcode



JSON Parsing



Form ID



Database Autofill



System Data

Form A

Handprint Text: SARAH JOHNSON

Machine Text: Premier Drug Testing

Checkbox: YES NO

Barcode: 123456789

<https://www.pharmetk.com/patch-education/patch-removal/removal-bar-code-labels-and-chain-of-custody>

“Hand-printed forms are still a thing”





Handprint Recognition Demo

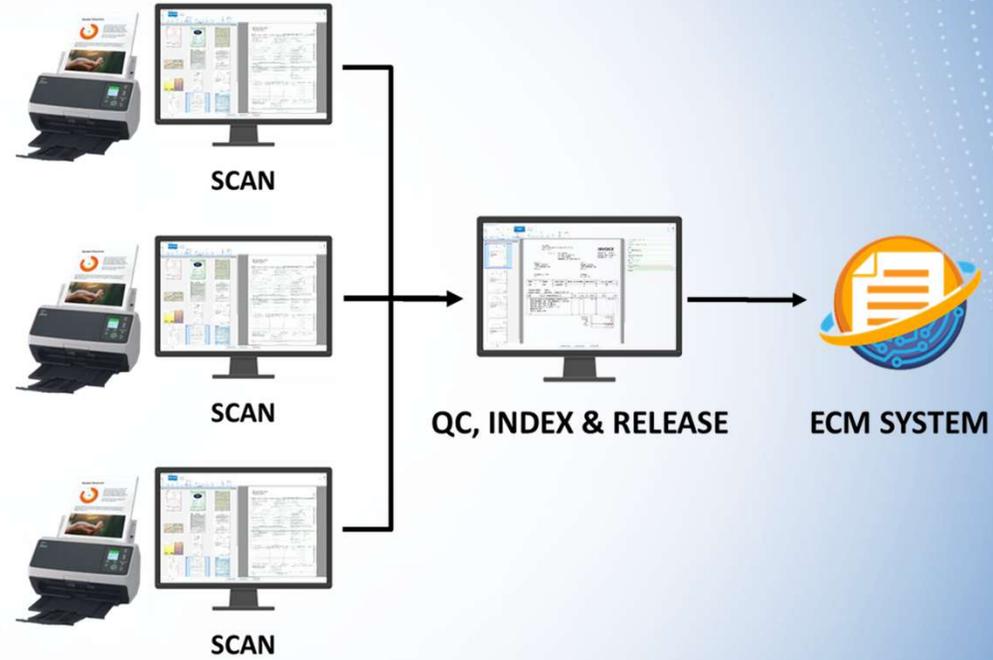
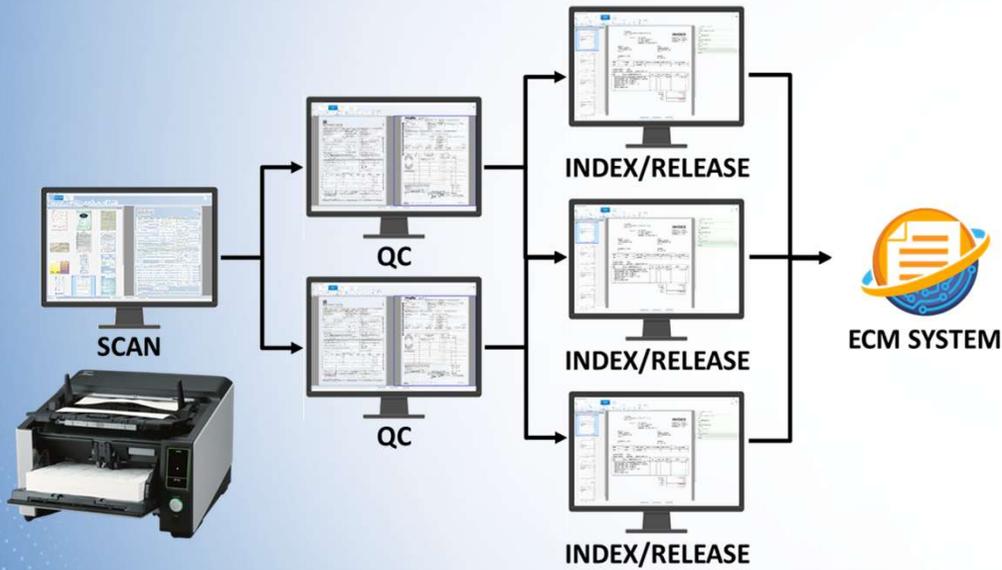


Multi-Form Processing

*“We need to
distribute our
capture
workflow”*



Multi-Station Processing Accelerates Capture Workflow!

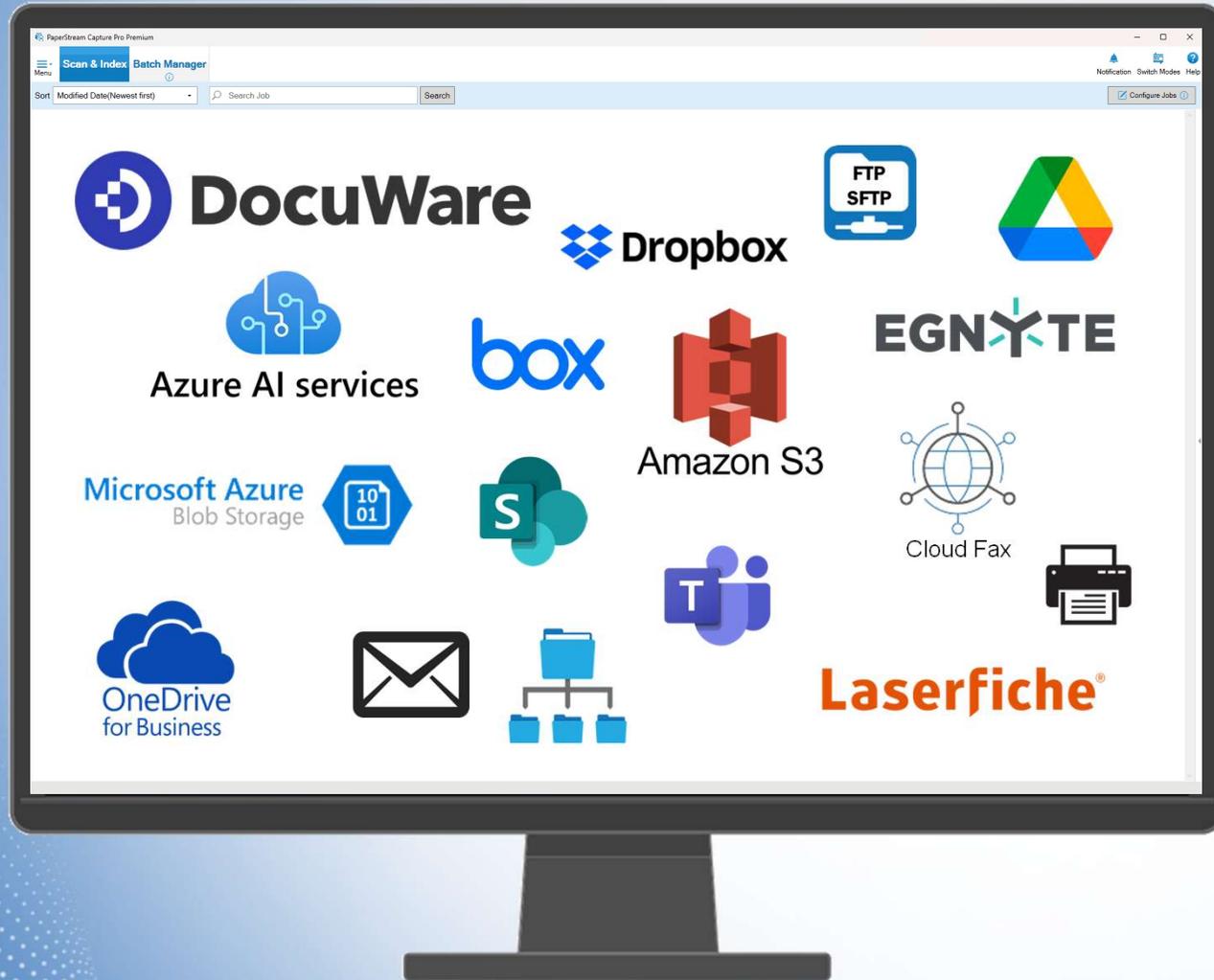


*“We need to
protect
confidential
information”*





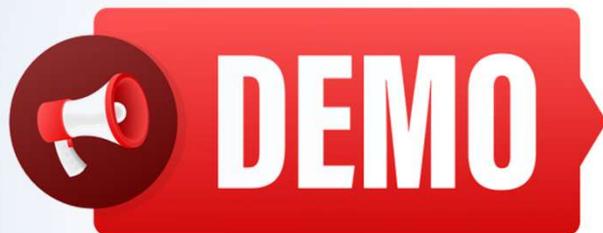
Field Redaction



**PaperStream
Capture Pro**



**Direct Application
Integrations Streamline
Capture Workflow &
Reduce Manual Processing**



Application Integration

Contact us to learn more!



**Come by our booth to see our scanning
& software solutions in action**



AI+IM
GLOBAL SUMMIT





PS PaperStream
Capture **Pro**

Thank you!