

Whitepaper 2021

# Guide to OCR, Invoice Scanning & Data Capture





Most organizations say manual data entry and the high volumes of email and paper invoices they receive are a huge challenge

## The Invoice Capture Challenge

Third party reports from the field confirm the challenge organizations face when capturing invoice data. 80% to 90% of invoices received by both medium and large size companies is either paper or email.

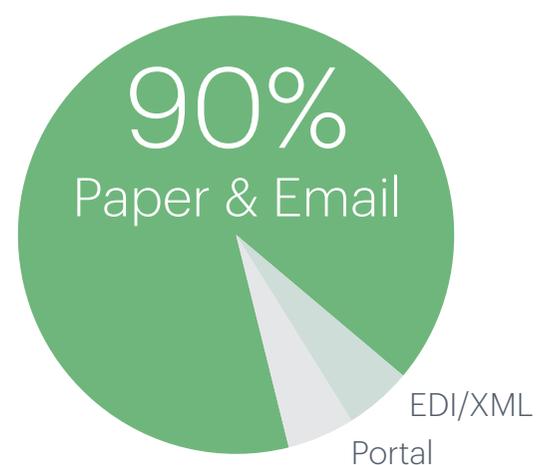
As you can see from the charts presented here, even with the advent of technologies such as EDI and network portals, the bulk of both invoice receipt and invoice data entry revolves around paper and email.

So how are organizations tackling invoice capture? Not very efficiently:

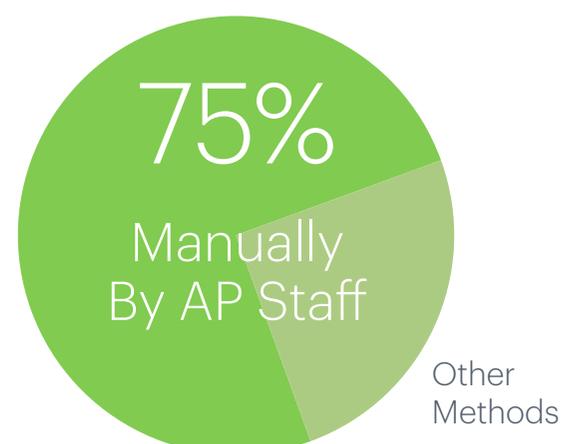
- Paper invoices are handled and entered into accounting systems manually
- Emailed invoices from suppliers are printed out and manually keyed into the ERP
- Invoice scanning is done with legacy OCR technology that has limited functionality, depth, and integration capabilities
- For growing companies, larger invoice volumes has meant adding to their headcount
- Most ERP solutions do not have the functionality to automate these areas

## Paper & Email Dominates in North America

How Companies Receive Invoices



How Companies Enter Invoice Data





## Capturing Paper & Image Data with OCR

### Formats Suited to OCR

- Paper Invoices
- Visual Digital Format Invoices
  - ▶ JPG, PNG, GIF (picture formats)
  - ▶ TIF (from scanning software)
  - ▶ PDF containing a scanned image

Because paper invoices and invoices sent as an image have no data layer, each invoice has to be scanned by Optical Character Recognition, or OCR. The collected data is then input to the ERP or accounting system.

### Resource Implications

Manually capturing essential data fields from emailed or paper invoices is time consuming and resource intensive. AP departments handle mailed or emailed invoices in various ways, with most of them keying in the data manually, a time consuming and error-prone process at best.

Conventional or legacy OCR systems often require manual input in order to function because of the time and effort for both field mapping and the collecting of data after it has been captured.

The latest advances in OCR technology reduce the amount of time involved in legacy OCR, introducing automation and intelligent data collection and retrieval.

**2020**

**95% accuracy rates are achievable with intelligent OCR technology**

## The Evolution of OCR

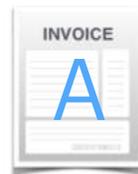
### Traditional OCR - 2000s



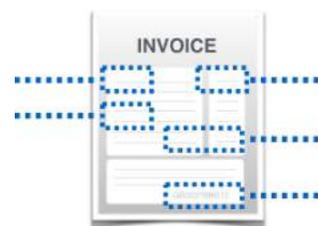
OCR is used to process either paper or emailed invoices, and the invoice layouts are manually field-mapped. There are often issues with integration and a lack of workflow.

### Advanced OCR - 2014

OCR is able to detect the invoice type based on its layout and sorts it accordingly.



### Intelligent OCR - 2020



Intelligent OCR teaches itself to identify supplier invoices based on their layouts and then automatically stores that information for future reference.

Intelligent OCR also intuitively maps the data fields in each invoice using intelligent tags, called "intellitags", as identifiers, and stores them in a database.

Because of the use of automation and smarter processing, 95% accuracy rates are achievable with intelligent OCR technology.



## Capturing Electronic Format Invoices

### Formats Suited to Digital Capture

- Digital/Electronic Invoice
  - ▶ Spreadsheet (Excel)
  - ▶ Word (.doc)
  - ▶ HTML
  - ▶ XML
  - ▶ Data PDF
  - ▶ EDI (EDIFACT)
  - ▶ CSV

A digital or electronic invoice is a document that exists in a digital format (composed of data in the form of binary digits). It is output by a supplier's ERP system as either structured or unstructured.

### Structured Digital Format

XML, EDI, Data PDF, HTML, CSV, Spreadsheet, Word (.doc)

### Unstructured Digital Format

Invoice data cannot be automatically read from the document into accounting systems.

Today most ERPs can output a structured digital format, such as a PDF. The majority of invoices received by email are structured invoices in one of the formats listed above.

**2020**

Accuracy rates of **100%** thanks to the data layer inside each electronic format invoice

## The Evolution of Digital Invoicing

### Traditional EDI - mid 1960s



Companies have their own form of electronic data interchange until the establishment of the ANSI X12 EDI standard.

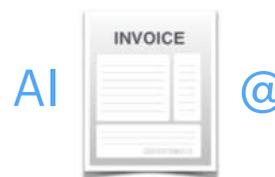
### Web-based Portals - 1990s

Online submission of individual invoices as well as EDI file uploads, including CSV, PDF, and XML formats.



### Intelligent Capture - 2020

ANY FORMAT



Automation

A digital format invoice is sent to a designated email address where capture automation and AI instantly reads the data and extracts it, forwarding the invoice data to the buyer's ERP.

Because the electronic file has a data layer that reflects exactly what was input, the information transfer is 100% accurate.

Invoice processing happens automatically and within minutes of receipt.



## Capture Methods Compared

Invoice Format	Speed to Set Up	Infrastructure Changes	Cost to Buyer	Cost to Supplier	Data Accuracy	On-Boarding Rates	Technology
PDF/ Electronic	Fast	None	¢	Free	100%	High	Intelligent Capture
Paper	Fast	None	¢	Postage	95%	High	Intelligent OCR
Portal	Fast	Some	\$	\$\$	Low	Low	Network
EDI/XML	Slow	Many	\$\$\$	\$\$\$	100%	Low	EDI

**For organizations that want to capture invoice data digitally, there are a few things to consider before choosing the best capture solution.**

The business case is based on considerations such as:

- the time it takes for set up
- infrastructure changes and I.T. resources
- the cost to the buyer
- the cost to the supplier
- data accuracy
- on-boarding rates

### Speed to Set Up

How long will it take to set up and implement the technology to capture invoice data efficiently? Intelligent Capture is fast because the supplier only has to send the email invoice to a designated email address. Intelligent OCR accepts paper, so the supplier can continue to send paper invoices.

2020

No costs incurred by suppliers

### Infrastructure Changes

What infrastructure is necessary on the supplier's end? Intelligent Capture has no infrastructure requirements on the suppliers end, because most ERPs can output an invoice in PDF or other electronic format.

### Cost to the Buyer

What is the cost to your organization? Things to consider are user licenses and software costs. The cost for Intelligent Capture/OCR is low because pricing is based on volume and allows for an unlimited number of users.

### Cost to the Supplier

Both Intelligent Capture and Intelligent OCR are free for suppliers.

### Data Accuracy

What data accuracy can you expect? For paper invoices, 95% accuracy rates. For electronic formats, 100% accuracy rates. Encouraging suppliers to send an electronic format instead of paper benefits both parties - lower costs, accuracy, and faster payment cycles.

### On-Boarding Rates

How many suppliers will on-board to your invoice capture solution? On-boarding rates for both Intelligent Capture and Intelligent OCR are high because there is no cost to the supplier, and all suppliers need for emailed invoices is a designated email address.

## Invoice Formats

- **Paper Invoice**
- **Digital/Electronic Invoice**

An electronic invoice is a form of digital invoice. A digital invoice is a document that exists in a digital format where digital is defined as "composed of data in the form of especially binary digits" (Merriam-Webster). There are 2 types of digital formats – visual and data.

- **Visual Digital Format Invoices**

JPG, PNG, GIF (picture formats)

- TIF (from scanning software)PDF containing a scanned image

- Data Digital Format Invoices

***Unstructured -***

The data cannot be automatically read from the document into accounting systems

***Structured -***

Spreadsheet (Excel)  
Word (.doc)

HTML

XML

Data PDF

EDI (EDIFACT)CSV

## Delivery Formats

Digital/electronic invoice formats are sent using the following means:

- **Email (most common)**

- **Data Transfer Protocol**

SFTP

HTTPS

AS2

Palette Software is a market-leading vendor of financial process automation for domestic and global corporations, including AP Automation and Purchase to Pay Automation.

Palette solutions automate the connecting and matching of purchase orders, invoices and contracts, on-premise or in the cloud.

Customers experience significant and measurable cost savings, productivity gains and operational excellence. Palette solutions are GDPR compliant and optimize financial management for more than 4,000 customers in 50+ countries.



96% Customer Satisfaction Rating

## 3,500 Customers Worldwide 250,000 Users



**Schedule an Appointment**  
to increase the productivity  
in your organization

[Schedule an Appointment](#)