

# The 6 Essential Elements of Smart Document Capture Technology

*A guide to discovering more about document capture, classification and data extraction.*

# TABLE OF CONTENTS

<b>Trapped Data</b>	<b>1</b>
<b>Smart Capture Workflow Overview</b>	<b>2</b>
Ingestion	
Image Processing	
Classification	
Extraction	
Validation	
Export and Deliver	
<b>Is Smart Document Capture right for you?</b>	<b>3</b>
<b>Types of Document Capture</b>	<b>5</b>
Manual Process	
Basic Document Capture	
Smart Document Capture: On-Premise	
Smart Document Capture: Hybrid	
Smart Document Capture: Cloud	
<b>Discover Shamrock's Intelligent Data Extraction</b>	<b>7</b>
<b>Advanced Capture Benefits</b>	<b>8</b>
Cost Reduction	
Process Improvements	
Productivity Benefits	
Security, Connectivity and Deployment	
Shamrock's Intelligent Data Extraction Differentiators	
<b>Building Your Business Case</b>	<b>10</b>
<b>Summary</b>	<b>11</b>

# Trapped Data

In today's world, where instant data access, business intelligence, security and efficiency are critical to success, many companies are realizing that valuable data is trapped in their documents. These documents could be paper, email or standard electronic office documents. The data stuck in these documents must be manually read, tracked, routed, processed and reported upon. In fact, more than 80% of information is trapped in unstructured content. This means that only 20% of data is structured and can easily be searched and retrieved from relational databases.

Document capture technology is not new, but the industry has advanced with innovative tools and functionality that allow businesses to do much more than simply scan documents. Now, technology enables businesses to classify, learn and extract meaning from their documents. Through automation, we can leverage and organize all data, both structured and unstructured.

The bottom line is that smart document capture technology is important not only for gaining efficiencies and reducing operating costs, but through classification and data extraction, it can lead to better business processes.



# Smart Capture Workflow Overview

There are 6 main steps of the workflow, although **only two require human touch points**:

## 1. Ingestion

There are multiple ways to capture data: scanners, multi-function peripherals (MFPs), UNC folders (network folders), fax, email, content services or document repositories, mobile devices or through an outsourced business process organization (BPO).

## 2. Image Processing

Documents and images are normalized, cleaned up and rotated in preparation for classification. The system applies despeckle and deskew filters to improve image quality. The resulting document can then be identified, and the data can be easily extracted.

## 3. Classification

This is where the system determines what type of document it ingested through Optical Character Recognition (OCR), Intelligent Character Recognition (ICR) and/or Optical Mark Recognition (OMR). This step will determine if a document is, for example, an invoice, patient record, loan file or tax record. An advanced document capture system only needs one or two samples, so it can “learn” to classify the documents; Shamrock Solutions accomplishes this via patented supervised machine learning algorithms. The system uses a variety of techniques to classify the data: search content, images, bar codes and one document merging. If the system has low confidence in any document it attempts to classify, the process can call upon a human operator for confirmation.

## 4. Extraction

This is the process of identifying metadata within the documents. Metadata is a set of data that describes and gives information about other data. In the case of documents, metadata can be used to organize, find and/or feed documents into another type of business system. The system is set up to extract the data based on business rules and information that a company needs through database lookups and fuzzy logic.

## 5. Validation

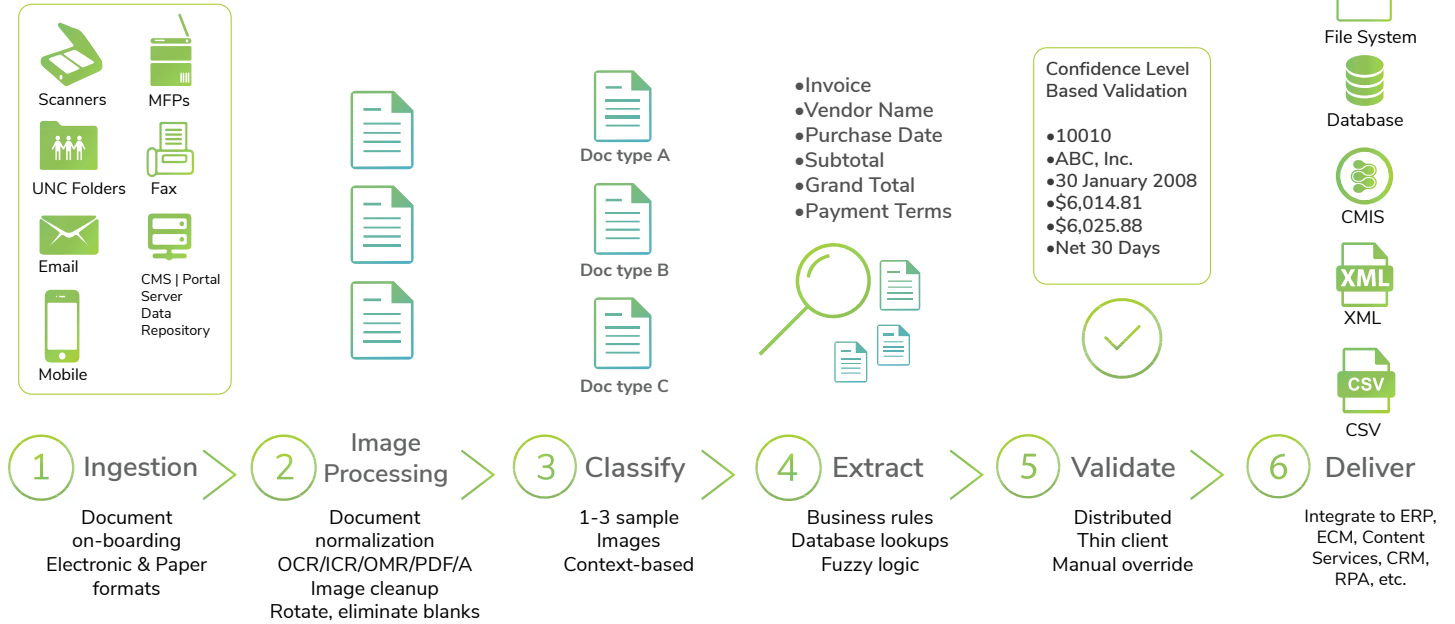
If there are any documents that fall below pre-set tolerance levels, they are highlighted for human review. For example, this can happen when there are smudges, spills, blurry characters or possibly missing fields. The system alerts you to these documents for manual verification and correction.

## 6. Export and Deliver

Once all documents have been validated, the documents and data are moved to a repository or other line of business system. The exported documents and data can be stored on a local server or cloud-based storage, like Alfresco, Box or SAP.

In summary, although the workflow consists of 6 steps, there are only two that require human touch points – Step 1, Ingestion and Step 5, Validation. Most of the work is done through the software.

## Advanced Document Capture Process Automation



## Is Smart Document Capture right for you?

Organizations that have large paper volumes as well as critical document-driven processes will benefit most from intelligent content capture technology. For instance, mortgage and title companies often have hundreds of pages of contracts and loan forms, healthcare facilities that keep handwritten patient records, lab reports and forms, government agencies that are notorious for storing forms, contracts and records, banks or financial institutions that hold account information and loan data, and any company that has a mailroom, are just several examples of organizations that create and manage a high volume of paper and can reap the benefits of advanced capture. Other high-volume verticals include the public sector, education and the intelligence and security communities, among others.



Companies with heavy document volumes often face similar challenges:

- Slow, manual processes
- High costs of ink and paper for printing
- Need for physical storage space for paper files
- Time consuming efforts to manually sort and file documents
- Manually applying barcodes, document preparation and other labor-intensive processes
- Ability to scan files but can't search for specific data in documents
- Invoice data that can't be easily matched to purchase orders or receiving documents
- Difficulty predicting trends or analyzing customer data
- Prone to human errors

If your company faces any of these hurdles, advanced document capture technology is available to transform your data, maximize efficiency and cut costs.

***“This solution has been a sea of change for us. We’re moving toward paperless workflow and we’ve opened doors for scanning on the front end. It’s a very positive change.”***

---

– Susan Hartsock,  
Information Technology Supervisor at Mountain West Financial, Inc.



# Types of Document Capture

There are five approaches to document capture that you may consider to address your requirements. These range from simple to sophisticated. Depending on your business model and document volume, most companies, large or small, will gravitate to a system that provides the best mix of functionality as well as value.



## Manual Process (on-premise)

Users scan documents into a PDF or TIF, then email the file or save it locally (file on computer, shared drive, etc.). The information is digital, but there are numerous drawbacks. For example, there could be documents that may not be easily usable or searchable, and that content could be exposed to unauthorized users or accidentally deleted. Typically, this is the best solution for smaller companies with infrequent document transactions, low overall volume and documents with handwritten comments, especially documents that contain cursive writing.



## Basic Document Capture (on-premise)

In this type of system, there may be a scanning room or employees dedicated to scanning and document preparation. There are a variety of vendors who offer software to scan and capture data. However, there are often limitations with the number of documents processed, ease of setup, configurability, along with other missing functionality such as auto-classification, web services APIs, a browser-based deployment model and the ability to run on platforms other than Microsoft Windows. These types of software don't understand content. For companies that already have simple, structured documents with a few data fields, this level of legacy capture software could be sufficient.





## Smart Document Capture (on-premise)

This solution can extract data utilizing technologies like OCR, ICR, free-form extraction, fuzzy database matching and machine learning algorithms. Supervised machine learning systems can offer classification accuracy near 100% and data extraction accuracy in the high 90% range. This leads to a decreased need for an employee to review exceptions with as much as 90% straight through processing. The best fit for this type of solution are enterprises with complex data, high volume and the need or desire to host the solution locally, on premise.



## Smart Document Capture (hybrid, a combination of on-premise and cloud)

Companies that have occasional peak or volatile processing needs can now use an on-premise solution for regular activities with the luxury of offloading heavy processing to the cloud for maximum performance and speed. Once the high-volume processing step is completed in the cloud, the data is sent back to the on-premise system. The hybrid model is flexible, scalable and based on demand.



## Smart Document Capture (cloud)

Industry pundits report that companies across all industries are rapidly moving their line of business applications over to cloud-based solutions as they are quicker to provision, cost effective, scalable and flexible. This solution has the same functionality as the on-premise version, but cloud-based systems use a SaaS model with rapid software deployment.

Cloud services can be beneficial to companies deciding between their operational budget vs. a capital expense. If an operational or departmental budget is being considered a SaaS model can make it more cost effective by spreading the costs out over time so there is a lower upfront cost. It is a hosted solution, so customers are not required to have an on-premise server or additional staff to conduct upgrades and maintenance. These factors may make the SaaS option more cost effective and flexible than on-premise solutions and there is a quicker time frame to achieve ROI. In addition, customers can select a shared, multitenant environment or a private cloud environment. The ideal customer for cloud-based solutions are organizations that are taking a modern approach and moving all or many of their other applications to the cloud.





# Discover Shamrock's Intelligent Data Extraction

Shamrock's Intelligent Data Extraction is an easy to use, Smart Capture® platform that can run either on-premise or in the cloud. Its patented, supervised machine learning algorithms are one of many ways that Shamrock's Intelligent Data Extraction is leading and innovating in the document and content capture space. What differentiates Shamrock's Intelligent Data Extraction from any other product is that the application was designed for and optimized for a cloud service environment. Shamrock's Intelligent Data Extraction also employs multiple classification methodologies to better identify documents so customers can extract actionable data and make better business decisions. Additionally, the platform has intuitive dashboards, a color-coded system to easily detect exceptions, a multitude of options to customize data extraction for the best results, and both Administrator and Operator screens and reporting tools.

Structurally, Shamrock's Intelligent Data Extraction has a complete browser-based interface for easy remote access; the application can even be extended to mobile devices with its SnapDoc app for mobile document capture. Shamrock's Intelligent Data Extraction is a Java based application. This provides the flexibility to run on Microsoft Windows or Linux platforms.

From a technical perspective, Shamrock's Intelligent Data Extraction's Web Services OpenAPIs allow the system to easily connect to and integrate with other business applications. Shamrock's Intelligent Data Extraction's Web Services OpenAPIs use a Representational State Transfer (REST) design which allows for loose coupling (fewer constraints on platforms, languages, and operating systems), lower bandwidth and is faster and more secure than SOAP technology (Simple Object Access Protocol). The advanced APIs allow the customer to incorporate advanced document classification and extraction capabilities within their existing business applications, providing a fully customized capture solution that most users are not even aware that they are using.



# Smart Capture<sup>®</sup> Benefits

Smart document capture solutions provide numerous benefits for customers worldwide.

## Cost Reduction

---

- Significant data entry savings through reduced manual keying
- Integration costs reduced through speedy deployment; CMIS (Content Management Interoperability Services) export plugin allows for integration with any CMIS compliant repository or business system
- IT infrastructure and SaaS model savings due to easy roll out and thin client (browser-based)
- Less paper leads to a more environmentally sustainable business and lower consumable costs
- Reduce mailroom and resource costs by automating document workflow
- Overall processing costs decreased by an average of 80%
- Avoid compliance and privacy (ex. GDPR) breaches and fines

## Process Improvements

---

- Auditability: the capture process performs with proven, trackable results
- Compliance: results are measurable and visible for easy compliance reporting
- Reduced Error Rates: studies show the human error rate is 8% in classifying documents, due to lack of time off, illness and fatigue
- End User Acceptance: users embrace the graphical user interface (GUI) which leads to positive cultural change, higher performance levels and immediate increases in efficiency
- A magnitude gain in document processing throughput volume
- 90% straight-through processing for most customers
- Processing time reduced by approximately 67%

## Productivity Benefits

---

- More work completed in less time with the reduction of manual processes: documents can be processed 20X faster than manual processes
- Productivity yields faster ROI for the company
- Increased customer satisfaction as payments are processed faster
- Faster path to informed decisions and ROI with streamlined workflow
- Reduce cycle times for early payment discounts and other benefits
- Improved accuracy through multiple recognition engines: OCR, ICR, free-form extraction, fuzzy database matching and machine learning algorithms
- Highest industry raw data classification with superior pre-validation accuracy
- Productivity gains average 75%
- Easy to access and share data with a distributed or global workforce
- Training is easy, including available self-guided [YouTube videos](#)

## Security, Connectivity and Deployment

---

- Install on-premise or leverage the cloud for easy, scalable access on demand
- Open source-based platform for easy integration with other business systems
- 100% browser-based architecture
- Available on Microsoft Windows or Linux
- No per seat charges
- No limitations on the number of pages processed
- Roles based UI – users only see what they need to see to do their job

## Shamrock's Intelligent Data Extraction Differentiators

---

- Shamrock Solutions is the only company that offers 100% browser-based, capture solutions
- Shamrock's Intelligent Data Extraction is certified for Amazon Web Services and Microsoft Azure
- Shamrock Solutions is customer and partner focused: we pride ourselves on being the easiest company in the industry to work with
- 24/7 enterprise support available to assist mission critical and global deployments
- Java based, open architecture using industry standard tools to ensure customers are not locked into a proprietary stack
- Patented supervised machine learning technology enables your document capture to improve over time, increasing its value
- Multiple ways to classify documents: through content, bar code or layout analysis
- Web Services OpenAPIs allow easy, fast integration into most systems



# Building Your Business Case

Convinced that your company needs Smart Capture® technology? Here's a guide to help gain buy-in from your organization and key decision makers.

1. Gather data. Most vendors will need some estimates to help size the system. Typical information needed includes an estimate of the number of documents or pages that are currently stored and need conversion, along with what you estimate your day-forward capture needs are. It is also beneficial to understand your current processing throughput and costs.
2. At Shamrock Solutions, our Solution Engineers can help guide you through gathering all the data you need.
3. Research case studies of other companies using smart document capture.
4. Demo and test the product. Our Solution Engineers and Sales can also help you present to your executive team using your own documents.
5. Research vendors. This includes product functionality, company stability, connectivity, innovative R&D and customer support options, along with ensuring that you are treated as a partner for a long-lasting relationship. Remember, your data is often the key to business and customer insight.

## Here are typical questions to ask the vendor you are considering:

- Does the software need to be installed on every machine?
- Do you need separator sheets for scanning batches of documents?
- What type of classification methods do you use?
- What happens with exceptions?
- Will the system lock up if I reach a predetermined monthly scan volume?
- What type of reporting tools do you have?
- What types of training programs do you have?
- What does your customer support include?
- How long does it typically take to implement?
- Do you integrate with my other software/ERP/CRM/RPA/Content Services or other repositories?
- Do you offer a cloud-based, SaaS, multitenant solution?
- Will you provide a Service Level Agreement?
- Will this effort reduce overhead? What are the expected savings?
- What partners do you work with that might also benefit my company?

## Summary

Smart document capture platforms are business improvement tools that help both large and small enterprises gain efficiencies and reduce costs while extracting meaning and data from any type of document. While cloud and SaaS models are gaining traction, some customers prefer on-premise solutions or a combination of both, so multiple options are available to you. Shamrock Solutions' primary goal is to help our customers automate and streamline processes to ultimately help them gain insight, improve efficiency and increase profitability. To learn more, visit [www.shamrocksolutionsllc.com](http://www.shamrocksolutionsllc.com).

## Contact Us

Contact us today at [sales@shamrocksolutionsllc.com](mailto:sales@shamrocksolutionsllc.com) for a personalized demo or call 833-875-8259.

