

# Deep-Dive: Key Factors for Document Automation Proof of Concept Success

*Here we focus on these key areas:*

1

What **automation** using intelligent capture technology really means and why it is different

2

How to properly **scope** a PoC with a core framework (activities, deliverables, and considerations)

3

What is **needed** prior to the PoC in terms of resource commitments, inputs, goals, scope and success criteria

4

How to identify the right **participants** and what to require from participating vendors

5

How to compare outcomes to success criteria and **evaluate** PoC results

# Table of Contents

**03** Four Proof of Concept (PoC) Steps

**05** Look For Labor Reduction

**09** Document Automation Technology Without Jargon

**11** Clarify the Problems You Want to Solve

**14** Clarify What You Want the “PoC” to Prove

**16** Detailed Preparation

**19** Evaluate the Results

**21** How to Improve Your Accuracy Results

**23** Accuracy is Only Half of the Equation

**25** Not All “Machine Learning” Is real

**04** General Preparation

**07** Clarify Your General Future State Capture Approach

**10** Document Automation Should Do 4 Things

**12** Select the Vendors and Tools to Evaluate

**15** Select the Use Cases You Want to Evaluate

**18** Testing

**20** Example: Test Results Summary and Error Analysis

**22** Intelligent Capture Is Not CRM

**24** Solutions Expected to Deliver Precision Require Precise PoCs

### Four Proof of Concept (PoC) Steps

In the last few years, we have seen a few powerful, closely related trends. First there's been an explosion of innovation in Intelligent Process Automation, which includes the technologies for document automation, RPA, artificial intelligence, machine learning, computer vision and natural language processing.

Most of our clients—particularly large financial services firms, insurance companies and utilities—are rethinking their enterprise capture strategies in light of digital transformation (digital first!) replacing paper, and their dissatisfaction with their incumbent enterprise capture platforms and outsourcing capture vendors. Most of our clients are not just kicking the tires of the new document automation tools. They are trying to do POCs or bakeoffs between a few of these tools and typically their incumbent capture platforms or other tools. We think this definitely the way to go. We strongly recommend that you do a POC of these tools in your environment if you are thinking about your capture strategy. However, conducting a successful PoC is not easy.

**This eBook focuses on how to properly scope a PoC: what you need to in terms of preparation, how to identify the right participants, what to do with the results, and what next steps you should take.**

It's useful to break the PoC process for document automation into four big steps, each of which consists of several smaller steps. First there's general preparation, where you plan what you want from the PoC and make some of the big decisions such as what technology to test and why. Second, detailed preparation is where you get everything ready for testing. Third is the actual evaluation where you test the software tools using your own documents in your own environment. And then finally, there is the fourth step where you evaluate your results and decide your next steps.

**01** General Preparation

**02** Detailed Preparation

**03** Testing

**04** Results Evaluation & Next Steps

### General Preparation

General preparation has six steps. The first is to **assess your own capture** to determine what kinds of processes involve paper, digital, document and data ingestion. Is it big or small, centralized or distributed, in-house or outsourced?

The second step is to have a general idea regarding **where you want to be** in 18 months, three years and five years or more. We're at that inflection point where most of us are doing some form of digital transformation, which means redesigning our processes so they are born-digital. Your organization is almost certainly not going 100% digital, but you have to decide how many of your chips you want to spend on making your current inefficient paper-based and manual processes more efficient, versus redesigning them completely so there is no paper at all.

The third step is to target and clarify **the problems you want to solve and the entailed relevant solution**. So you might want to address your work streams that are all straight through processing except where there's a handprint step or signature step. You should understand the general solutions and approaches that are relevant to the kinds of problems you are focusing on.

The fourth step is to **clarify what you want to prove** in the proof of concept. Do you want to see if they work, if they work well with your documents and environment, if they are cost-effective? Or, do you want to find out something else?

The fifth step is given all the above, select the vendors and tools you're going to evaluate.

Sixth, **select the specific use cases** to evaluate. Which document types in what work streams do you need to test?

#### 01 General Preparation

#### 02 Detailed Preparation

#### 03 General Preparation

#### 04 Results Evaluation & Next Steps

1  
Assess your capture operation.

2  
Clarify your general future state capture approach.

3  
Clarify the problems to solve and the relevant solutions.

4  
Clarify what you're trying to "prove."

5  
Select the vendors and tools to evaluate

6  
Select the use cases to evaluate.

### Look for Labor Reduction | Overview

Since the first step in General Preparation is to understand your capture operation, let's look at the three typical flavors of capture operations:

1. It may include a centralized operation using one of the capture platform vendors; or
2. It may include some significant part that is outsourced to a for-profit capture vendor that does it for you; and
3. It most certainly includes some decentralized ingestion that takes in documents and data from multiple locations and channels. These include paper, multi-function devices, email and FAX and smart smartphones and portals.

Many operations have been doing something like this since the late 1990s, and for many of them, it is a mess. There are opportunities to greatly improve efficiencies in all of these scenarios with the new document automation tools.

A real business case exists for using the new document automation tools to reduce the labor in the capture workflow. This table below illustrates some of the potential that is just sitting there.

Service	Unit	Ex. Peer	Industry	For-profit	In-house
Sort & Prep	Per hour	\$ 35.50	\$ 35.28	\$ 34.15	\$ 36.40
Scanning	Per image	\$ 0.0546	\$ 0.0398	\$ 0.0298	\$ 0.0497
<b>Data Entry (onshore)</b>					
Tier 1 (<10 keystrokes)	Per index	\$ 0.0650	\$ 0.0891	\$ 0.0665	\$ 0.1117
Tier 2 (11-30 Keystrokes)	Per index	\$ 0.0950	\$ 0.2264	\$ 0.1630	\$ 0.2898
Tier 3 (ave. = 50)	Per index	\$ 0.3750	\$ 0.4589	\$ 0.3460	\$ 0.5717
<b>Retrieval / Lookup / Research</b>					
Research (Tier 1)	Per hour	\$ 47.75	\$ 49.70	\$ 49.00	\$ 50.40
Retrieval (Tier 2)	Per hour	\$ 63.75	\$ 58.50	\$ 59.40	\$ 57.60
Automated Recognition	Per image	\$ 0.0225	\$ 0.0206	\$ 0.0132	\$ 0.0280

### Look for Labor Reduction | Details

First, here are some typical costs from in-house and outsource capture operations. At Doculabs, we collected them from 25 years of capture benchmarking research. Your costs are likely similar to the PEER column. The INDUSTRY column combines the average costs of outsource (FOR PROFIT) and IN-HOUSE.

Second, the rows are the various tasks that are performed in a capture workflow. Note **automated recognition** at the bottom – that is cheap because it has very low labor costs.

Third, you can get significant improvement by moving the data entry labor to the automated recognition row. But that's not all.

You can also significantly reduce the **sort and prep labor and the exception handling labor** (research activities). **General QA, error detection and error correction** can also be reduced or avoided.

Fourth, we have found that many organizations just look at how they can use document automation for reducing data entry rather than the other labor activities in the capture workflow. They are leaving savings on the table.

Service	Unit	Ex. Peer	Industry	For-profit	In-house
Sort & Prep	Per hour	\$ 35.50	\$ 35.28	\$ 34.15	\$ 36.40
Scanning	Per image	\$ 0.0546	\$ 0.0398	\$ 0.0298	\$ 0.0497
<b>Data Entry (onshore)</b>					
Tier 1 (<10 keystrokes)	Per index	\$ 0.0650	\$ 0.0891	\$ 0.0665	\$ 0.1117
Tier 2 (11-30 Keystrokes)	Per index	\$ 0.0950	\$ 0.2264	\$ 0.1630	\$ 0.2898
Tier 3 (ave. = 50)	Per index	\$ 0.3750	\$ 0.4589	\$ 0.3460	\$ 0.5717
<b>Retrieval / Lookup / Research</b>					
Research (Tier 1)	Per hour	\$ 47.75	\$ 49.70	\$ 49.00	\$ 50.40
Retrieval (Tier 2)	Per hour	\$ 63.75	\$ 58.50	\$ 59.40	\$ 57.60
Automated Recognition	Per image	\$ 0.0225	\$ 0.0206	\$ 0.0132	\$ 0.0280

## Clarify Your Future State Capture Approach | Overview

To clarify your future state capture approach, below is a picture of the capabilities you should focus on based on the approach you take.

**H** stands for High, **M** for Medium and **L** for Low, using a metaphor of having a stack of chips or dollars and deciding how much to put into improving your current processes versus leap-frogging them and redesigning them as all digital. Most enterprises do **moderate transformation for capture**.

You should focus most on the dark blue boxes: multi-channel and multi-format capture, the entire middle row – classification, extraction, validation, and conversion. And then, focus on task automation with RPAs.

Legend

H	M	L
---	---	---

**Moderate Document Automation** dramatically improves the efficiency of current processes, most of which involve paper:

Multi-channel Capture	Multi-format Capture	E-forms	E-signature
Classification	Extraction / Enrichment	Validation / QA	Conversion / Export
Content Analytics	Process Management / Automation	Task Automation (RPA/RDA)	Administration, Monitoring, Reporting

**Advanced Document Automation** redesigns the processes to leapfrog paper-based challenges:

Multi-channel Capture	Multi-format Capture	E-forms	E-signature
Classification	Extraction / Enrichment	Validation / QA	Conversion / Export
Content Analytics	Process Management / Automation	Task Automation (RPA/RDA)	Administration, Monitoring, Reporting

**H** stands for High, **M** for Medium and **L** for Low, using a metaphor of having a stack of chips or dollars and deciding how much to put into improving your current processes versus leap-frogging them and redesigning them as all digital.

## Clarify Your Future State Capture Approach | Details

When focusing on the dark blue middle boxes, classification is automatically identifying what pages, what documents and what document packages the system receives. Extraction recognizes the data on the pages, and possibly enriches it (e.g., by judging customers’ sentiment given the letters they sent you). Validation and QA checks for completeness and correctness of the information and document sets. Finally, conversion/export transforms the data into the standard format the downstream systems and people need (e.g., converting customer name into ALL CAPS with NO SPACES or PUNCTUATION).

Advanced transformation for capture is all digital. So, you don’t care so much about capture or about the various kinds of document processing in the middle row. Instead you focus on e-forms, e-signature, process and task automation (like case management, BPM and RPAs), content analytics for mining that rich trove of digital data and making smart decisions about it, and then lots of monitoring and reporting. That whole middle row represents a **technical revolution in document processing**, along with RPAs in the bottom row. Next, we will clarify some facts about these technologies.

Legend

H	M	L
---	---	---

**Moderate Document Automation** dramatically improves the efficiency of current processes, most of which involve paper:

Multi-channel Capture	Multi-format Capture	E-forms	E-signature
Classification	Extraction / Enrichment	Validation / QA	Conversion / Export
Content Analytics	Process Management / Automation	Task Automation (RPA/RDA)	Administration, Monitoring, Reporting

**Advanced Document Automation** redesigns the processes to leapfrog paper-based challenges:

Multi-channel Capture	Multi-format Capture	E-forms	E-signature
Classification	Extraction / Enrichment	Validation / QA	Conversion / Export
Content Analytics	Process Management / Automation	Task Automation (RPA/RDA)	Administration, Monitoring, Reporting

H stands for High, M for Medium and L for Low, using a metaphor of having a stack of chips or dollars and deciding how much to put into improving your current processes versus leap-frogging them and redesigning them as all digital.



### Document Automation Technology Without Jargon or Magic

There is a lot of confusion about the new Intelligent Process Automation (IPA) tools, the new messy technology category that includes document processing, RPA, AI, machine learning, computer vision, natural language processing and generation, cognitive computing as well as a few other exotic-sounding technologies. In a nutshell *without jargon or magic*, here is a first cut for understanding and communicating the most important aspects of the new document automation tools. Improved recognition and learning are separate capabilities although they are typically combined in the new products. The results are much better than what you got from the older 1990s OCR tools:

First, the new tools are better in **how they recognize the text** – in how they classify documents, paginate pages and extract documents. In a sense, they have changed from how **young children** read into how **adults read**. Young children focus first on letters and then words in a painful, slow process. Adults use more context, including existing knowledge, the overall document and page layout and look, what the document's probably about, etc. The new tools may use computer vision, pattern analysis and other methods to take advantage of more context.

Second, the new tools are better in **how they improve that recognition** – in how they continuously improve the accuracy of their opinion in light of evidence. The learning capabilities range from not-so-intelligent to brilliant. They range from the case where: 1) the human supervisor must do everything (e.g., drawing bigger zones in the template); 2) the human must provide them with positive and negative feedback (“these are correct answers” or “these are wrong answers”); 3) it's a mix of assisted and automated; to where 4) the system can brilliantly improve its recognition automatically 100% by itself.

#### *Document automation tools have improved in recognition and learning:*



##### *How the systems recognize the text:*

- To extract data, classify documents, or paginate/separate pages and documents
- Movement from how young children read to how adults read – using more context



##### *How they improve that recognition:*

- How they improve the accuracy of their “opinion” in light of evidence
- Whether the human supervisor must do everything (e.g., drawing bigger zones in the template), or must provide them with positive and negative feedback (“these are correct answers” or “these are wrong answers”), whether it is a mix of assisted and automated, or whether the system can improve its recognition automatically %100 by itself

# Document Automation Should Do 4 Things

So the new tools are better at recognition and better at learning but what makes a good tool for your real world production environment? Generally, any adequate document automation solutions should do four things:

1. It must work well enough to adequately **reduce or eliminate human intervention**.
2. It must either:
  - a) have enough **functional breadth to be standalone**, *or*
  - b) **fit easily into a larger environment** that provides the functional depth (your capture platform).
3. It **must fit into the larger process environment** beyond the recognition stages (your downstream BPM or case management environment).
4. Also, it must provide an **adequate user experience** for any humans who are assisting the automation (with QA, error correction, etc.)

### 01

It must work well enough to adequately reduce or eliminate human intervention.

---

### 02

It must either (a) have enough functional breadth to be standalone, or (b) fit easily into a larger environment that provides the functional depth (your capture platform).

---

### 03

It must fit into the larger process environment beyond the recognition stages (your downstream BPM or case management environment).

---

### 04

It must provide an adequate user experience for humans who are assisting the automation.

***We have recently empirically evaluated some solutions, and some fulfill these requirements more favorably than we expected.***

### Clarify the Problems You Want to Solve

An important part of this is to figure out what kinds of problems you should solve in your capture operation. Specifically, there are problems that these tools can address that the older generation of tools could not. Below is a table that shows easy, moderate and difficult problems in the three primary document automation processes: classification, extraction and validation.

As a first cut, you can probably solve the easy problems with the solutions that you have today. Your biggest opportunity in the next couple years will likely be to focus on the **Moderate** problems and some of the **Difficult** problems. And then, you can consider tackling the more complex **Difficult** tasks after you have that set in happy production.

For example, if you can do a good job at document classification you don't have to do much document preparation. These new tools can automate or near-automate this task better than previous tools. Once you tackle that—in a few years—you can take on package classification.

Regarding extraction, you can crank up accuracy on machine text well beyond 99%. Also, you can address handprint and handwriting, first in structured English and then later in unstructured other languages.

For validation, you can check for completeness and correctness in entire documents – what's called NIGO processing (Not in Good Order). Later on, you can do NIGO processing for entire packages of docs, as in a loan application. Or, check correctness in addition to signature presence.

Complexity	Classification Capabilities	Extraction Capabilities	Validation/QA Capabilities
<b>"Easy"</b>	<ul style="list-style-type: none"> <li>• Page classification</li> </ul>	<ul style="list-style-type: none"> <li>• English machine text – structured layout</li> <li>• Mark and signature detection</li> <li>• Barcode recognition</li> </ul>	<ul style="list-style-type: none"> <li>• Doc type, page number and order, signature presence, date validation, field completeness</li> <li>• Data validation with source system, name/entity validation (against e.g. registry)</li> </ul>
<b>"Moderate"</b>	<ul style="list-style-type: none"> <li>• Document classification (single page type may belong to multiple documents)</li> </ul>	<ul style="list-style-type: none"> <li>• English machine text – unstructured</li> <li>• Multilingual machine text –structured</li> <li>• Multilingual machine text – unstructured</li> </ul>	<ul style="list-style-type: none"> <li>• Document field value consistency</li> </ul>
<b>"Difficult"</b>	<ul style="list-style-type: none"> <li>• Package classification</li> </ul>	<ul style="list-style-type: none"> <li>• English hand – structured</li> <li>• English hand – unstructured</li> <li>• Multilingual hand- structured</li> <li>• Multilingual hand – unstructured</li> <li>• Metadata enrichment (e.g. sentiment analysis)</li> </ul>	<ul style="list-style-type: none"> <li>• Signature correctness</li> <li>• Package completeness</li> <li>• Package field value consistency</li> </ul>

# Select the Vendors and Tools to Evaluate | Overview

Now it's time to **select the vendors and tools to evaluate** since you have a good picture of what the problem is and what you're going to be doing. So here's a useful graph of the document automation vendor territory. It's also a good way to think of a solution and how they fit into your situation. **Two dimensions** are most important for evaluating tools that are out there: deployment readiness and innovation.

**Deployment readiness** is how ready the technology is to be put into production in your environment – can you unbox it and plug it into your workflow? Or, is it engine parts that you need to assemble together to yield a custom one-off? Being a bucket of parts isn't bad if you have (or can buy) the expertise or if you need a specialized solution. However, you may want a more productized solution.

**Innovation** is the ability to provide good—accurate, high quality and usable—output. Innovation may be highly specialized: for example, focused on classifying very messy unstructured documents; extracting data from unstructured multi-lingual handprint; validating information across multiple documents in a set; or classifying packages of documents in a mortgage application.



*Tools differ primarily on deployment readiness versus innovation:*

*Category #1 –*  
Mature “platform” vendors

*Category #2 –*  
Mature “solution” vendors with innovative offerings

*Category #3 –*  
New “solution” vendors with innovative offerings

*Category #4 –*  
New “engine/tool” vendors with innovative offerings

### Select the Vendors and Tools to Evaluate | Details

It's useful to divide the vendors and tools into four categories. **Category #1** contains the mature platform vendors. You likely know who they are – they run your scanners and devices and provide the whole workflow from scanning to release.

**Category #2** are the mature solution vendors who are not the big platform vendors – and who typically specialize in the document automation processes, like OCR, other kinds of extraction, classification and validation. They have lots of mature, legacy capabilities, but they also have innovative automation capabilities with the best—as innovative as the newer vendors. Parascript fits in Category #2.

**Category #3** contains the new solutions vendors with innovative offerings. Most of these vendor have emerged in the last 5 years or so. They were born providing AI, ML, CV, NLP and other technologies. They are at least reasonably deployment ready.

**Category #4** are the new “engine or tool” vendors with innovative offerings. This group has the most innovative offerings, but they are definitely not deployment ready. The category includes many smaller startups as well as a few very large vendors.

The obvious vendor choice is: If you already have a platform or other relevant vendor, include them in the POC to give you a baseline and conservative option of sticking with your incumbents. If you want to deploy in the near term and can't use developers, select from the right side categories (1, 2, and some 3s). If you are doing more of a research project or have expertise at hand, include the vendors from category 1 and 2.



*Tools differ primarily on deployment readiness versus innovation:*

**Category #1** –  
Mature “platform” vendors

**Category #2** –  
Mature “solution” vendors with innovative offerings

**Category #3** –  
New “solution” vendors with innovative offerings

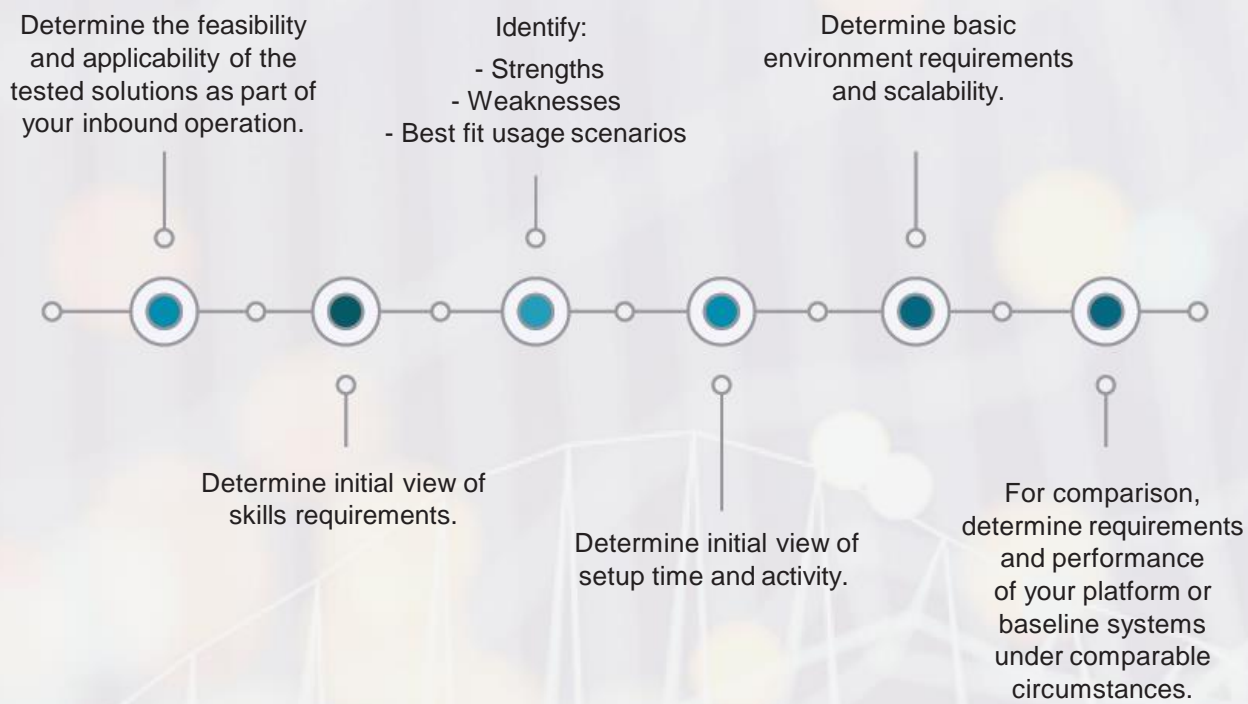
**Category #4** –  
New “engine/tool” vendors with innovative offerings

### Clarify What You Want the PoC to Prove

The next step is to figure out what your Proof of Concept should prove. Here are the most relevant items to prove.

1. **Determine the feasibility and applicability** of the tested solutions as part of your inbound operation.
2. **Identify the strengths, weaknesses and best fit usage scenarios.** If they work on handprint is it all handprint or just English, in a structure document format?
3. **Determine basic environment requirements and scalability.** Do they require substantial on-premise infrastructure? Does any change in the workflow require a labor intensive configuration in a black box or is it easy and transparent?
4. **Determine initial view of skills requirements.** Can you use your existing capture and industry-standard skills or does it require exotic scripting? Can capture supervisors and staff make necessary changes to the workflow or does it require IT expertise?
5. **Determine initial view of setup time and activity.**
6. For comparison, **determine requirements and performance of your platform or baseline systems** under comparable circumstances. Often, it isn't worth adding another vendor into the mix with all additional costs. It may be best to get more aggressive with your existing vendors and systems to crank up the automation.

Some of these criteria can be quantitative, such as the percentage of correct data versus incorrect data. However, many of these criteria are qualitative and require interviews and judgments of the participants and evaluators.



# Select the Use Cases You Want to Evaluate

Now the last step of General Preparation is to specify the specific use cases to evaluate:

- These include use cases such as **change of address forms** with English, French and Spanish machine print and handprint, multiple checkboxes.
- Or **customer survey forms** with 5 free-form comment boxes. You might say: we want to separate and paginate the surveys, and extract the comments. (We can also use the tools to do the simpler extraction tasks like barcode or machine text reading or use our platform capabilities for those tasks)
- Or **invoices in our AP process**, which come from a wide variety of smaller suppliers. We want to separate the invoices.

The table below illustrates the level of complexity involved in different types of use cases from easy, moderate to difficult use cases.

Complexity	Classification Capabilities	Extraction Capabilities	Validation/QA Capabilities
“Easy”	<ul style="list-style-type: none"> <li>• Page classification</li> </ul>	<ul style="list-style-type: none"> <li>• English machine text – structured layout</li> <li>• Mark and signature detection</li> <li>• Barcode recognition</li> </ul>	<ul style="list-style-type: none"> <li>• Doc type, page number and order, signature presence, date validation, field completeness</li> <li>• Data validation with source system, name/entity validation (against e.g., registry)</li> </ul>
“Moderate”	<ul style="list-style-type: none"> <li>• Document classification (single page type may belong to multiple documents)</li> </ul>	<ul style="list-style-type: none"> <li>• English machine text – unstructured</li> <li>• Multilingual machine text –structured</li> <li>• Multilingual machine text – unstructured</li> </ul>	<ul style="list-style-type: none"> <li>• Document field value consistency</li> </ul>
“Difficult”	<ul style="list-style-type: none"> <li>• Package classification</li> </ul>	<ul style="list-style-type: none"> <li>• English hand – structured</li> <li>• English hand – unstructured</li> <li>• Multilingual hand- structured</li> <li>• Multilingual hand – unstructured</li> <li>• Metadata enrichment (e.g., sentiment analysis)</li> </ul>	<ul style="list-style-type: none"> <li>• Signature correctness</li> <li>• Package completeness</li> <li>• Package field value consistency</li> </ul>

## Detailed Preparation | Steps 1-4 of Seven

Now we move to Detailed Preparation, which consists of 7 steps. The primary requirement for a useful test is to have plenty of Plan Bs. Many things are likely to go sideways – you might not get enough good documents to test, you might not get timely security and software access rights and many more possible issues. Try to head those off before they happen, but also be able to work with partial results. Here are steps 1-4 of 7 Detailed Prep steps:

1. **Select the document sets to evaluate**, typically 3-5 sets of e.g., 20, 50 or more – whatever you can do. The sets are for prep, for multiple passes in testing and for backup if things go wrong.
2. **Define the roles and participants**. What you will do, what the humans doing ground truth, keying and QA will do, what the vendor will do, and what the evaluators will do.
3. **Install and configure software in test environment**. Some vendors are completely hands-off in the cloud. You send them your scanned images, they do their magic and return the results. Others are off-prem, but part of your process and extended IT environment. Others are partly on-prem, partly off-site, and others may be 100% on-prem. Getting security and access clearance can be the most difficult task in the entire process so get rolling early on this.
4. **Prepare the document sets**. Make sure they have representative data – e.g., handprint. If you have to fake real docs by filling them out yourselves with no sensitive customer data like PII or PHI, try to fake it realistically.

01 General Preparation

02 Detailed Preparation

03 General Preparation

04 Results Evaluation & Next Steps



01

Select the document sets to evaluate.

02

Define the roles and participants.

03

Install and configure software in test environment.

04

Prepare the document sets.

05

Define outputs and "Ground Truth."

- Use auto and/or human.  
- Define whether "raw" or "normalized."

06

Create templates for each doc type (if using them).

07

Perform dry run with practice forms.



# Detailed Preparation | Steps 5-7 of Seven

Here are steps 5-7 of Detailed Preparation, which consists of 7 steps. The primary requirement for a useful test is to have plenty of Plan Bs:

- 5. Define outputs and “Ground Truth” data.** This can be much more complex than one would think. It’s deciding on the answer key. It’s typically not obvious “what’s there on the page.” You have to make decisions, and it must be useful. Be prepared to make decisions about spaces, random punctuation, upper or lower case, abbreviations, misspellings and other variances. For example, my home address can be written at least 27 different ways depending on how I capitalize, abbreviate “Street”, how I use punctuation, etc. So for those 27 examples, I could have 27 errors if they are different from my pristine definition of Ground Truth data, or they could all be correct, or some could be correct and some could be too far from the core simple case. It all depends on how you define Ground Truth. Make it reasonable and useful for the real world cases. Some other decisions:
  - Use auto and/or human. Often the process is to use human keyers to separately key in the GT and then compare. Many tests also use the tool to also get a vote.
  - Define whether “raw” or “normalized” – e.g., convert to all caps, no spaces or punctuation; or simple mark recognition in check box versus what the collection of 0-multiple checks actually means (e.g., \$100-150K income, likes apples, pears and bananas).
- 6. Create templates for each document type** (as needed). The vendor typically does this but you should watch and understand what they are doing and how scalable the process is.
- 7. Perform dry run with practice forms.** Always a good idea!

01 General Preparation

02 Detailed Preparation

03 General Preparation

04 Results Evaluation & Next Steps

01  
Select the document sets to evaluate.

02  
Define the roles and participants.

03  
Install and configure software in test environment.

04  
Prepare the document sets.

05  
Define outputs and “Ground Truth.”  
*- Use auto and/or human.  
- Define whether “raw” or “normalized.”*

06  
Create templates for each doc type (if using them).

07  
Perform dry run with practice forms.

# Testing

Testing has several steps:

1. **Run using auto software** (no human intervention).
2. **Compare extracted data to the “Ground Truth.”** This gives you raw baseline of how much correction is likely needed.
3. **Send all items (or items with low confidence) to humans.** Humans can QA all documents or you can use the self-assessment of the tool to determine what to send for QA. Note that some of these products are overconfident – they think.
4. **The humans do the QA and any keying to correct errors.**
5. **Compare corrected output to the “Ground Truth.”** It may be 99%+ or there still may be errors. Many of these are correctable – e.g., the humans didn’t get a clear set of rules in judgement cases.
6. **Vendor – or you – can tweak or train (optional).** Tweaking may be simple. However, training the AI usually takes many more samples and repetitions than you can provide in a POC.
7. **Run test again** (optional).

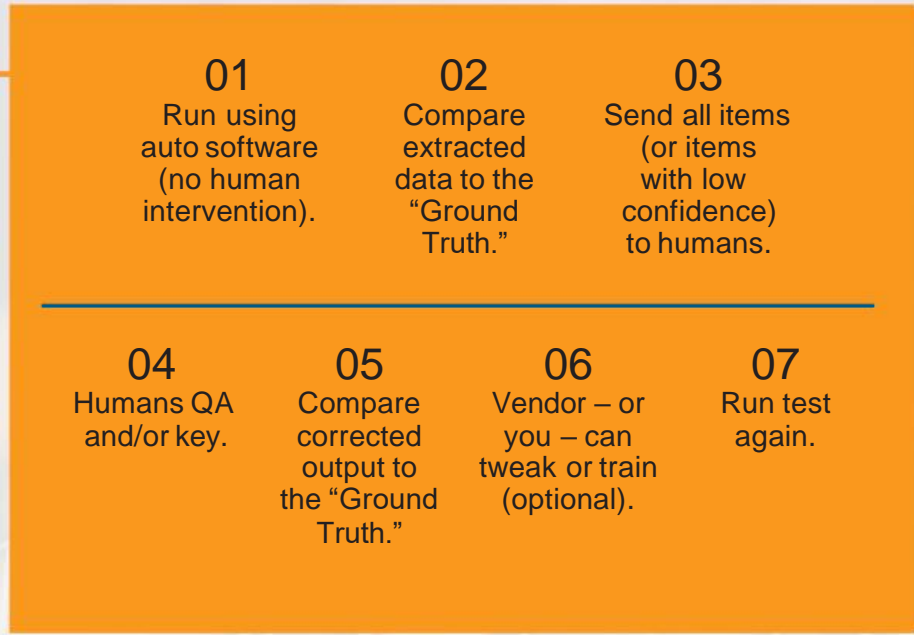


01 General Preparation

02 Detailed Preparation

03 Testing

04 Results Evaluation & Next Steps



# Evaluate The Results

Now it's time to evaluate the results and figure out what to do with what you learned. There are 6 typical steps:

1. **Identify strengths, weaknesses, best fit usage scenarios.** Machine print, docs in good order or sloppy ugly docs, hand print in English only?
2. **Determine basic environment requirements and scalability.**
3. **Determine initial view of skills requirements.** Interview the staff doing keying, the vendors, your IT and other capture staff.
4. **Determine initial view of setup time and activity.** This is quantifiable.
5. **Compare with the requirements and performance of your platform or baseline systems under comparable circumstances.** This may be an apples to oranges comparison, as the incumbents will already be in production. But they likely won't even be able to try some of the tests the new tools can tackle – like handprint.
6. **Determine your next steps.** This may be to:
  - It may be to run more docs and do more process tweaks to see what LEARNING and process tweaks can optimally yield you.
  - It may be to do a more real world test, plugging the tool into your enterprise capture operation.
  - Do a detailed business case using the new vendor's costs and performance data.



01 General Preparation

02 Detailed Preparation

03 General Preparation

04 Results Evaluation & Next Steps

01  
Identify strengths, weaknesses, best fit usage scenarios.

02  
Determine basic environment requirements and scalability.

03  
Determine initial view of skills requirements

04  
Determine initial view of setup time and activity.

05  
Compare with the requirements and performance of your platform or baseline systems under comparable

06  
Vendor (or you) can tweak or train (this is optional).

07  
Determine your next steps.

# Test Results Summary and Error Analysis Example

Here's an example from a real test:

- There were 9255 fields across all the documents in this test.
- The system by itself delivered 91.5% accuracy.
- When we added humans to QA and correct, we achieved 97.0% accuracy – note that this isn't perfect. However, some of the errors were eliminable with training and other tweaks. The baseline platform only gave 65.5% accuracy – and we didn't even test handwriting.
- If you look at the errors, you see that many of them could be reduced by normalizing the output (which smooths over differences in spaces, or formatting, or capitalization, or punctuation). Or by reaching agreement with the humans regarding what's acceptable or not.

## Example

Result Type	Fields	Extraction Accuracy	Comments
Auto System	9255	91.5%	No Human Intervention
Auto w/ Assistance	9255	97.0%	Auto System Plus Keying
Baseline Platform	9255	65.5% (Typed only)	No handwriting

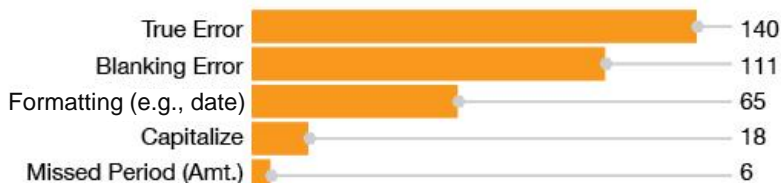
### Machine Error Analysis



= 1%  
 7% Discrepancies  
 (340)



### Discrepancy Analysis



# How to Improve Your Accuracy Results

Let's get a little more specific on where accuracy can be improved versus where it likely cannot:

1. You need higher volumes and repetitions to get significant improvement in machine learning.
2. You can get immediate improvement by talking to the keyers and getting consistency (and by normalizing output, both as ground truth criteria and as final output transformation step).
3. You can do 1990s-type constraints on the fields to constrain identification – e.g., SSNs only have numbers, state fields are just 1 of 50, etc.
4. Capture processes often yield terrible quality images. It's tempting (and sometimes necessary) to deal with this using smart recognition – e.g., solutions using computer vision are often best at identifying skewed and mangled images and letting the extraction engine get a chance at an accurate answer. However, first or at least in addition, try to ensure/improve the image quality – using the cleanup capabilities all capture programs have, and providing guidance to the staff who are doing the capture – even if they are customers and not employees.
5. The other problem is goofy writing, either because it's illegible or everywhere on the page except the actual box that's designed for it. Again, some of the really innovative tools can do fairly well on some of these items. However, the best approach is to provide guidance to agents and customers on how to write on forms.

## Opportunities to improve the results:

**01** **To maximize machine learning:**  
requires significantly more volume to be run.

**02** **Inconsistent keying:**  
can easily be improved with process instructions for consistency (e.g. rules for periods, spaces, interpretation).

**03** **Auto extraction mistakes:**  
some of these misidentification mistakes can be addressed with constraints on the fields (e.g. alpha filters, simple lookups, complex lookups).

### ***Document automation may not effectively address:***

**04** Image quality issues: upstream issue with upstream recommendations; better image cleanup (deskewing, despeckling, optimized contrast and density for machine recognition) should reduce at least some errors (similar looking characters, characters overlapping lines, etc.).

**05** Goofy writing where an individual wrote outside the expected zones or wrote bizarre text; this will likely always be an issue, but you can get the overall incidence down with guidance from staff and customers.

# Intelligent Capture Is Not CRM

We engage with a lot of organizations directly and indirectly through our partner channel. We find that it's just a fact that many organizations—even those that have already incorporated what they call intelligent capture, some cases OCR—are not sure how to evaluate intelligent capture software in a way that really tests its strengths and weaknesses.

Advanced capture is not your typical enterprise software. We often compare it to “CRM” (Intelligent Capture is not CRM) because most people are familiar with customer relationship management software, but you can almost put anything else under the sun there where the intent or final intent isn't based on precision or accuracy. Unlike intelligent capture, CRM can be evaluated based on features and functionality.

However, intelligent capture is really about getting document-based tasks or processes automated. In order to do that, it's all based off of the precision of the system: how much data you can crunch in an accurate way. Certainly the underlying features and functionality support that ability, but there's a lot of stuff that goes into this and a lot of things that you have to be aware of to be able to really evaluate intelligent capture systems. Features are important, but they only get you, maybe 20% of the way there in terms of intelligent capture evaluation.



The purpose and value of intelligent capture is automating as many document tasks as possible at the highest attainable levels of accuracy.



Yet many evaluations and Proof of Concepts (PoCs) focus on features and functionality.



So, is understanding features really important if you cannot verify the true performance of the system? Its precision and accuracy are essential to ensuring you get quality data results.

# Accuracy Is Only Half of the Equation

When data extraction or document classification accuracy percentages are used, it is critical to know what these accuracy percentages are really referring to because it can be misleading. The full set of measurements that allow you to figure out how much of that data can be processed straight through requires a close examination.

Read rates are important. The read rate means the amount of data that can produce answers and of that will then produce a statistical analysis on the amount of that data that is 99% accurate or 95% accurate or whatever accuracy rate you need to meet your requirements. It's impossible to tell much from an accuracy number unless you know what it is referring to.



You can't tell unless you know the other important attribute – how many “predictions” (e.g., what document type or field value answers can be provided).

You need to understand the % of answers that can be identified by the software as accurate or inaccurate.

# Solutions Expected to Deliver Precision Require Precise PoCs

The measurements of actual production tests are most revealing when they are done at the **field level**. Over the last two or three years, there's the perspective or perception that intelligent capture should be measured at the page or document level in terms of success or failure. An example would be an invoice, where the success rates are measured as the percentage of invoices were all fields can be located and extracted correctly. Here the real results are very, very poor. That is there is going to be a very, very low percentage of overall pages or documents that are 100% for every single field.

Anytime you've got a data extraction task with more than one field, it is absolutely **not** advisable to measure success in a business process based off of the page or a page or a document level success rate. The increasing use of robotic process automation or RPA is kind of making people believe they can get that because RPA started off with automating tasks at a gross level where they're very simple, route and easy to define. These tasks had very few exceptions and few variances. You could have a high percentage of tasks and go straight through with 100%. It's important to evaluate intelligent capture at the field level.



Regardless of Machine Learning, you need reliable and representative sample sets.

Number of Samples	Amount of Automation
10 Samples	Not more than 5%
100 Samples	Not more than 50%
1000 Samples	Not more than 70%
2000 Samples	Not more than 85%
5000 Samples	>85%



Representative sample sets take into account both document types and document variance within these document types.



# Not all “Machine Learning” Is Real

Machine learning and machine learning types of applications usually need a lot of samples. Simply put, intelligent capture with machine learning can identify variances, even slight variances and within documents a lot better than humans can. Machine learning can analyze a lot more documents than people can and a lot more efficiently. However, you need that data. So one of the constant questions we get asked is, well, how many samples do we need? With machine learning, you need plenty of samples.

It is really interesting that rules-based types of models (e.g., if we find a keyword look to the right of that to find the value and everything like that) are very brittle even though they are a form of artificial intelligence. They require a lot of analysis and are hand-coded. With rules-based intelligent capture, analyzing a lot of samples is too laborious and tedious.

With machine learning, this becomes a simple compute-time operation. You need plenty of samples. Certainly the amount of success goes up a lot as you are able to add more samples because you're able to take into consideration, but unlike other systems, machine learning based systems improve over time and work very well in dynamic production environments..



You can't tell unless you know the other important attribute – how many “predictions” (e.g., what document type or field value answers can be provided).

You need to understand the % of answers that can be identified by the software as accurate or inaccurate.

# Thank You

## For more information:

### Doculabs

[www.doculabs.com](http://www.doculabs.com)  
312.433.7793  
[info@doculabs.com](mailto:info@doculabs.com)

### Parascript

[www.parascript.com](http://www.parascript.com)  
888.225.0169  
[info@parascript.com](mailto:info@parascript.com)

©2021 DocuLabs & Parascript Management, Inc. All Rights Reserved.

This eBook content is built from—and the direct result of—the **AIIM webinar**, “5 Key Factors for Document Automation Success,” hosted by AIIM with speakers: Rich Medina, Co-Founder and Principal Consultant at Doculabs; and Greg Council, VP of Marketing and Product Management at Parascript.