reveal group

Rethinking Intelligent Document Processing: Making Today's Tools Work for You

Josh Michalko, Jian Chan & Jeff Brown

Despite the promise of a digital world, businesses are drowning in paperwork. Improvements to the speed, accuracy and cost of processing documents are achievable with today's automation tools. So why are so few companies shifting to automation rather than relying on onshore and offshore operations teams? Automating document-laden processes is hard. The automation landscape is oversaturated with tools claiming vast and varying capabilities and we're yet to see a one-size-fits-all automation tool that delivers on all its promises. On average, a process that transcribes data from paper documents requires two to three times the development effort of a process that pulls from a database.

So, how have a select-few companies leapfrogged their competition by successfully automating their document processes and reducing their operational footprint? They've taken a two-step approach best summarized as "simplifying and linking." Simplifying breaks down a process into interrogative components and identifies a suitable technology to help with each piece. Linking then enacts an underlying design that connects the tools and enables an end-to-end automated solution.

How have a select-few companies leapfrogged their competition by successfully automating their document processes and reducing their operational footprint?

Overcoming 'The Document Problem' with New Tools

Optical Character Recognition (OCR) software has proven adept at digitizing documents but relies on humans for data extraction. In lieu of capable technological alternatives, companies have historically turned to onshore and offshore operations teams to process documents and to transcribe data into digital records. Human error and latency during such manual tasks have been widely accepted as a cost of doing business. For high-risk documents, the model of employing two workers in a 'maker-checker' fashion has been commonplace, despite the duplicative costs.

Recently, a handful of technologies have achieved tremendous advancements in automating data extraction. The advancements have been so profound that they have born a new classification segment; Intelligent Document Processing (IDP). IDP merges traditional digitization (OCR) with cutting-edge decision making and machine learning models to enable automated extraction with accuracy as-good-as, and in some cases better than, humans.

So, why haven't all companies adopted these new technologies to realize swift improvements to transcription accuracy, processing times and reductions to operations staff?

Automating a document-laden process is hard. We often overlook the amount of processing power the human brain exerts when completing a seemingly menial task. Consider the act of reading the status of a checkbox that has inconsistent locations across documents due to differences in form versions and scan quality. To automate the action, a developer would need to build and train an IDP model capable of overcoming the variations in the visual cues to consistently locate and read the checkbox. When we expand the process requirements to include sending and retrieving documents to and from the IDP tool, shifting the extracted data to downstream systems, and updating a workflow management (WFM) tool, the automation complexity is raised significantly. Additional considerations, such as a document's upstream point of entry and downstream cataloguing introduce further intricacies.

To deliver an automation that meets these requirements, a developer must simplify the process into modular components, identify suitable tools for each piece and construct a solution that links the tools together.

Your Unique Challenge: Where to Simplify and How to Link

To better predict complexity and return on investment (ROI), all document processes up for automation consideration should be categorized and vetted before any delivery efforts are made. Key criteria to consider when reviewing a document process include: annual number of documents, associated pages, associated data points, language and the presence of any handwritten values. Ticking off such a criteria list enables a company to make educated decisions when prioritizing processes for automation and selecting vendors.

Key criteria to consider when reviewing a document process include: annual number of documents, associated pages, associated data points, language and the presence of any handwritten values.

While today's leading IDP vendors each possess their own strengths, they are all capable of delivering results suitable for automation. Rather than going down the path of overanalysing accuracy results across IDP tools, the lion's share of a first-time implementer's preparation should be spent reviewing their end-to-end process and simplifying it into modular components. "Document Receipt", "Sorting", "Routing", "Processing" and "Post-Processing" are common components that make up a simplified document process. Subsequently, each component of the process should be examined via a series of questions to enable design and tool considerations (see Figure 1.1).

From the simplification and granular review of the process, all necessary tools can be identified and a sound solution can be constructed that links the tools together. Accepting that a one-size-fits-all tool is yet to be fully realized, developers generally need to link three to four technologies (IDP, RPA, BPM/WFM) to deliver an end-to-end automation capable of processing documents and referring exceptions for human review (see Figure 1.2).

Way Forward

As technologies continue to converge and compete over the next few years, we expect that all automation vendors left standing will offer a comprehensive, or one-size-fits-all, toolset. However, contrary to the common vendor promise of a low-code interface, we anticipate skilled developers with tool familiarity will remain a necessity to design and build sound solutions that deliver measurable benefits.

Skilled developers with tool familiarity will remain a necessity to design and build sound solutions that deliver measurable benefits.

Today, companies considering automating their document processes must decide between moving forward with a multi-technology approach or waiting for a vendor to emerge with a one-size-fits-all tool. Ready adopters will also need to consider the best delivery approach for their organizations, whether that is outsourcing development, developing internally, or enlisting the experience of a consultancy to help establish the capability. When successful, the first-mover advantage gained by companies choosing to move forward today will enable speed and cost savings that can ultimately be passed along to their customers.

reveal group

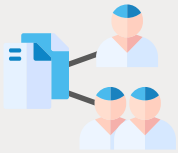
Figure 1.1: Questions to Simplify a Typical Document Process

Receive Documents

- How are documents received? Fax, Mail, Email
- If by mail, are documents digitized via a scanner?
- If received digitally, what is the format? (i.e., PDF, JPEG)
- Are all documents received to a centralized company address? Or are multiple addresses used?
- What level of structure do the documents have? Structured, Semi-Structured, Unstructured
- Do any documents include handwritten responses or signatures?

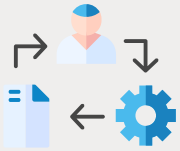
Sort Documents

- Is any type of OCR/Digitization technology currently utilized?
- How many different types of documents are received?
- Are there variations to any of the documents? If yes, how many?
- How many pages does each document contain?
- If documents are comprised of multiple pages, are the pages always received in order?
- Are documents ever received with missing pages?
- Do any of the documents contain sensitive information or PII which require a higher level of permissions to handle and review?
- Are document packages ever received containing pages related to more than one customer/case? If yes, how are the documents separated?
- Are any supplemental documents expected? If yes, are supplemental documents required to be organized and processed with related case documents?



Route Documents

- Post-sorting, are documents required to be routed to separate processing teams?
- If yes, how do downstream teams currently receive documents? Hard copy, Email, File Share, Other
- Have any teams been established to handle sensitive documents or employee-related documents?



Process Documents

- How are documents received? Fax, Mail, Email
- If by mail, are documents digitized via a scanner?
- If received digitally, what is the format? (i.e., PDF, JPEG)
- Are all documents received to a centralized company address? Or are multiple addresses used?
- What level of structure do the documents have? Structured, Semi-Structured, Unstructured
- Do any documents include handwritten responses or signatures?

Post-Processing

- Where are documents stored after processing?
- Is metadata stored and referenceable for each document? Received Date, Document Type, Completion Status, Unique ID / Case ID
- Are supplemental documents stored with links to parent documents?
- Are clients required to be notified post-processing? If yes, how are they notified? Phone call, Mail, Email
- In the event of a client inquiry, how is the client's case located?
- Is the client's original document referenceable from the same source as the client's case data? Or is a separate repository required to be accessed to recall client documents?

reveal group

Figure 1.2: Sample IDP Solution Linking Four (4) Technologies

Receive Documents	Sort & Extract Data	Retrieve & Enter Data	Referral Handling	Post-Processing

Email, Fax and Scanner

1.

Documents received via email and fax are manually added to a defined folder within a file share. Documents received via mail are first scanned and then added to the folder.

Robotic Process Automation (RPA) Platform

2.

RPA bots continuously monitor the shared folder for new document files and once identified, send document files to the IDP platform

4.

RPA bots retrieve extracted data from IDP and enter data into target system(s) as per business rules. Original and post-IDP document files are moved to a completed folder in the file share.

Intelligent Document Processing (IDP) Platform

3.

IDP initially sorts and categorizes documents. IDP then extracts defined data points. Supervision tasks are created for humans to confirm any values falling below the defined confidence threshold.

Workflow Management (WFM) Tool

5.

RPA bots add a unique ID per case to WFM along with links to the original and post-IDP document file. Cases which could not be fully completed by the bots are marked as “Business Exceptions” and are manually completed by the business team.

6.

In the event of an inquiry, the business team retrieves the client’s case by searching the WFM across several criteria, including client name, document received date and unique ID. Once located, the links are used to review the original and post-IDP versions of the document.