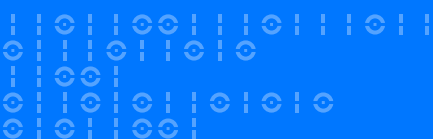**INDICO DATA**

Better data.
Better decisions.

Scaling IDP across the enterprise: The playbook for unlocking unstructured data

The AI revolution is starting, just not scaling

Automation has reached mass adoption. According to Deloitte's latest Automation with Intelligence Report,[1] nearly three quarters of respondents have implemented it into their business processes – a huge increase from just 58% the year before. What's more, IDC predicts that within the next two years, half of knowledge workers will regularly interact with their own AI-enhanced robot assistant, which will help identify and prioritize tasks, collect information, and automate repetitive work.[2]

While AI-driven automation touches most industries, Intelligent Document Processing (IDP). has particularly taken hold in sectors such as financial services, insurance, real estate, and healthcare, where employees have historically had to manually review complex documents review and/or rekey data into digital systems from forms.

IDP technologies automate these steps, from extraction and interpretation to classification and handling of digitized documents such as highly structured W-2 forms and mortgage applications or ungainly, unstructured KYC verifications, invoices, patient medical records or insurance claims.

The benefits of successfully scaling these sorts of automation initiatives can be enormous. Organizations can boost employee productivity, optimize operational efficiency and mitigate compliance risk. They can reduce costs and increase FTE capacity. Ultimately, they can drive revenue, increase customer satisfaction and improve employee experience.

	Mature Enterprises	Typical Enterprises
Return on Investment	**70%** achieved > 50% ROI **25%** achieved > 100% ROI	**24%** achieved > 50% ROI **41%** achieved 26-50% ROI
Cost Savings	**50%** generated US$10-50 million savings **38%** generated > US$50 million savings	**24%** generated US$10-50 million savings **68%** generated < US$10 million savings
FTE Capacity Created	**82%** created capacity worth 500-2,000 FTEs	**71%** created capacity < 500 FTEs
	Total hours saved through automation: **3.5x** for mature as compared to typical enterprises	
Key Areas with Maximum Improvement Over Pre-Automation Scenario	In addition to operational impact, mature enterprise's experienced improvement in • Business continuity planning • Employee experience	Typical enterprises saw improvement primarily on the operational metrics such as • Efficiency and turnaround time • Employee productivity
	Percentages indicates respondents in each category that selected that option	

Source: Everest Group[3]

> In 2018, only 4% of respondents said that they had achieved substantial scale with their automation initiatives (50+ robots). This number has been slowly increasing since then, and 13% of respondents were at scale in 2020.
>
> – Deloitte Automation with Intelligence 2020 Report

Unfortunately, the vast majority of companies have been unable to realize these transformational results. Many enterprises have trouble getting out of the blocks; recent studies show that only 11% of AI-enabled initiatives succeed in production. But it's not just the starting that holds organizations back; it's the scaling. Deloitte's Automation Intelligence Report reveals that just 13 percent of survey respondents are scaling (51+ automations), as opposed to piloting or implementing.

While AI represents a profound and exciting opportunity for solving Intelligent Document Processing, it requires a complex and well-calculated fuel mixture of powerful cognitive technologies, data availability and model training, cultural alignment and human-machine interaction to succeed. There's also the monumental impact gap created by the massive volumes of unstructured data that most organizations and their technologies are ill equipped to handle.

Featuring insights from Indico Data's automation experts, this playbook will:
- Briefly review some of the most common barriers we see to scaling IDP
- Provide actionable "plays" to accelerate, scale and unlock IDP
- Offer first-hand insights and experiences from experienced automation center of excellence (COE) leaders
- Explore an emerging solution that directly addresses and overcomes many of the barriers to IDP scalability

Why IDP fails to scale

What's holding enterprises back from scaling their automation and IDP initiatives? As you might expect, it's not one thing. Challenges abound, from the technologies to the people and processes involved.

Process design and fragmentation

Many automation COE leaders find that the introduction of an automation platform can reveal problems in existing business processes. By its nature, automation technology will break down and/or amplify biases, weaknesses or inconsistencies in a human-driven process.

"Unexplored process ambiguity is a huge challenge" asserts Chris Wells, former Chief Data Scientist at Chatham Financial and current VP of Solutions at Indico Data. "We see it all the time: subject matter experts label data and we find that the results come back poorly because the SME's teams have traditionally approached documents differently."

Deloitte's research loudly echoes Wells. "Immature and fragmented processes" have been cited in Deloitte's past three automation surveys as the most significant barrier to delivering intelligent automation at scale. When business-critical processes are not managed in a unified workflow, handoffs between departments and teams open automation to higher risks of error.

Top barriers to scaling intelligent automation
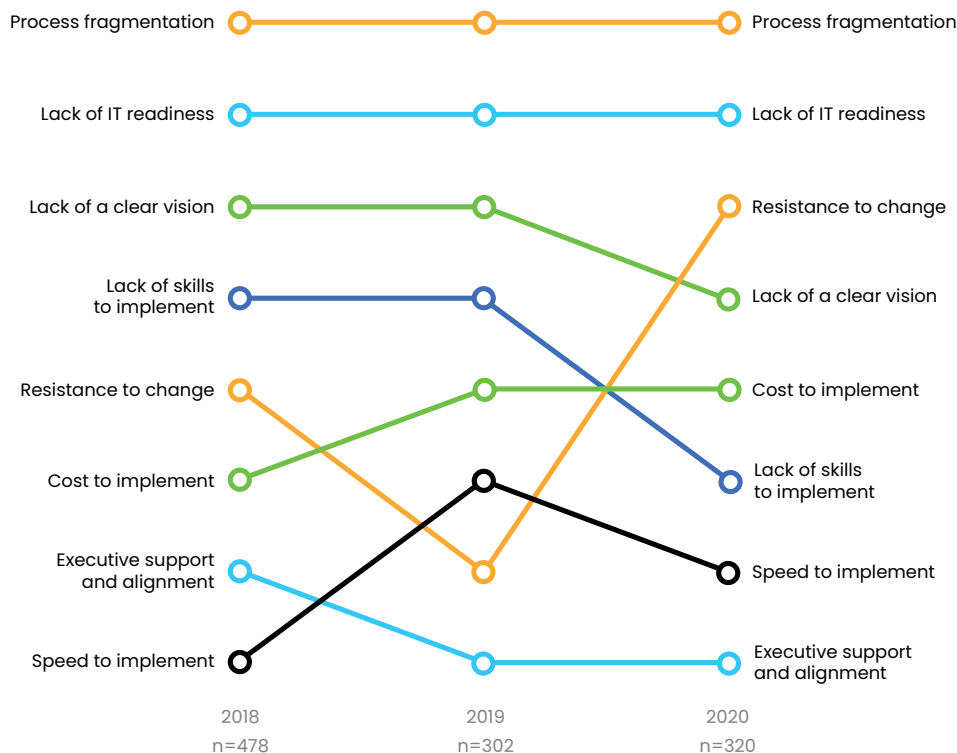
Source: Deloitte

"With a human workforce, organizations can paper over the cracks, because humans can adapt and react to change with a degree of flexibility that bots don't have," states Vishesh Bhatia, a process automation expert with Cognizant. "The bot continues to perform the old steps. It will either terminate or perform an undesired activity that might have a negative impact, kind of like a train going off the track."

Chris Well explains the challenge vividly with a baking simile: "It's like making bread. A subject matter expert looks at a document with five data points like they were ingredients and knows how to turn them into bread. But a data model doesn't inherently capture that decision, it just captures the eggs, milk and flour. It's not enough to know what you want to get to. You need to understand all the pieces of information you need to get to it."

Unready, uncoordinated or unwilling

Even the most powerful automation technologies fundamentally rely upon people (a point this playbook will more deeply address in a bit). But most enterprises do not recognize the wide-reaching change management required to broadly scale.

"A lack of realistic expectations and a thorough change management process doom too many automation projects," says Brandi Corbello, former VP of Transformation at corporate real estate leader Cushman & Wakefield and Indico Data's VP, Business Development. "IDP and automation can't succeed with a 'set it and forget it' mindset. A lot of organizations aren't planning for that."

Corbello believes that enterprises need to adjust their operational mindset. "Just as an industrial company needs to continually monitor its manufacturing operations, organizations need to maintain and optimize data automation on an ongoing basis. Factories have machine operations, and enterprise automation COEs need ML Ops – machine learning operations."

That said, most enterprises may not have a designated automation COE. Deloitte's research cites lack of IT readiness as its second greatest barrier to scaling, citing that only 37% of organizations reported that they have appropriate standards controlled by an intelligent automation center of excellence.

What's more, there's often ambiguity or even disagreement about who should own automation. Chris Wells shares, "Early in my work leading automation, I wish I had been more aware of the resistance to change we would have within the company, and not just from line of business. IT is often adjacent to automation and might be afraid of it. If it breaks, they think they'll be responsible."

How to Speak with LOB and IT: For strategies and recommendations for working with line of business, IT and other colleagues, check out Indico Data's **"Automation Conversation Guide: How COEs and LOBs Can Find a Common POV for Business Impact."**

The wrong tools for the wrong tasks

Maybe the biggest challenge: 85% of all data in large enterprises is unstructured. As its name implies, unstructured data by its nature does not follow an established format or model, making it challenging to search and analyze. It can be generated by humans or machines. It can be text based, or not. So, while some forward-thinking organizations have attempted to drive IDP initiatives with technologies like Robotic Process Automation (RPA), they've only been able to harness a meager 15% of what they have at their fingertips, because RPA solutions are not well equipped to handle unstructured data.

Brandi Corbello adds: "There are two ways I've seen traditional IDP initiatives fall down with unstructured data. First, a lot of COEs are trying to run IDP with strict optical character recognition software, which is inherently template based. This doesn't work for unstructured documents, which can't follow rules.

"The other is that some more mature COEs are embedding some ML/AI, but the tech is only capable of going so far with unstructured documents. Invoices are a great example. An automaton solution may be able to extract data, but the invoice is just lines. They still need to do lots of manual post-processing or build and connect additional RPA to handle post-processing. This requires even more application development, and then more time on support and maintenance because of brittleness in production. It's very hard to scale that way."

Big plays: several X's and O's for scaling IDP

There's no silver bullet for scaling any automation. But there are several strong bets to help break down the biggest barriers to accelerating, scaling and unlocking Intelligent Document Processing. Here are six plays you can make to immediately move the ball down field for your IDP initiatives. Then, read on to discover a new technology breakthrough from Indico Data that is tackling the greatest obstacles to scaling IDP head on – specifically, the historically unsurmountable unstructured data challenges.

1. Understand and optimize your workflow

Again, the most common barrier to scaling isn't the automation technology; it's the process. Begin by capturing the intelligence behind your processes in a machine driven workflow. Mine your process and map out the ideal state – and then determine the best ways to get there. From there, plan out your process in a way that makes it easy for the people training the data models and then the models themselves to label and process data effectively.

"Sometimes, a line of business brought us a 'really defined process' to automate," recalls Chris Wells. "Then, we would take the process into an automation tool and LOB would claim the tool didn't work [when the results weren't satisfactory]. The truth was, the AI model couldn't learn, because there was a hidden human variable. No two data labels were the same, and LOB teammates were treating data differently. Success all depended on who was doing the work. Only if you do your process due diligence can you understand if there are human failures or model failures holding back your automation."

By capturing, auditing and clearly mapping a process, you'll have the power to truly reinvent it and make it scalable. In turn, this delivers cascading benefits in the long run for your business processes and your people: less training, more resilience to employee turnover, a happier workforce, and ultimately more effective automation.

And remember, there's no one best answer for workflow design. "For any given process, unless dead simple, there are likely half a dozen ways to split it up," says Wells. "Really understanding the process enough to plug out the pieces and put into a workflow is a real art form. Be creative."

2. Commit to continuous improvement

Congratulations, you've successfully mapped, optimized and deployed your IDP initiative. But you're not done yet. In fact, the work is never over. Successfully scaling automation means ridding yourself and your organization of a finite mindset and committing to a culture of continuous improvement – a digital equivalent to the quality management and automated operations practices of industrial manufacturers.

"It's critical to have the right plan for your running operations," states Brandi Corbello. " If a bot breaks and a business-critical process stops running, what are you going to do? If you don't have that structure in place, who's gonna perform that task? You need to focus on IDP model operations so that you can manage your processes and make sure they're running soundly every single day."

Consider your machine learning operations procedures for continuous QM and improvement: monitoring, reporting, performance reviews, updates, and contingency/continuity planning. As Corbello says, "Putting IDP into production and just leaving it is like telling a five-year-old who just learned how to read to complete a five-hundred-page book. You can't just walk away."

3. Think end to end

It's easy to keep your focus on the primary extraction step of the IDP process, where your machine learning models extract the relevant data so that you can leverage it for application or insights. But the ability to accelerate and scale doesn't rest in this phase of automation alone. There are very powerful opportunities for optimization and improvement by keeping pre-processing, classification and post-processing top of mind.

> "
> Automation COEs should be thinking about how they can create a resilient, repeatable process which, in turn, accelerates downstream processing as well. That's the real way to scale.
> "
>
> – Brandi Corbello, VP, Business Development, Indico Data

Pre-processing can be impacted by the amount of noise or skew in the document. It can also be seriously affected by dealing with documents in multiple languages and with handwritten elements. Handwriting is especially crucial in sectors like healthcare, where physicians take notes and provide recommendations on patient records with pen and ink. Consider the types of documents coming into your IDP platform; the more languages it can read, and the more flexible it is with machine- or handwritten text, the more scalable it will be across use cases and global organizations.

Classification is another opportunity for increasing speed and scale. Enterprise documents contain various types of information. They can be multi-paged. They can also be multiple documents bundled together – each with different forms and formats – which makes them even more difficult to classify. And not all IDP solutions are equipped to handle this well. Corbello shares that, from her experience, "The amount of modeling inside or outside the platform to take apart and unbundle complex PDFs could take months."

It's also extremely prudent to consider what will happen during your workflow's post-processing stage, where data may need further validation before routing to a source application. Very often this work is handled by either an RPA solution or manual review. Either can be very costly and time consuming.

"Even on lower complexity use cases, you could typically be looking at forty to eighty hours of development, and another twenty to forty hours of testing on end users on post-processing," says Corbello. "Then that RPA script has to be maintained. If the use case is complex, it may be brittle, too. And there are the costs of ongoing licenses. Automation COEs should be thinking about how they can create a resilient, repeatable process which, in turn, accelerates downstream processing as well. That's the real way to scale."

4. Keep humans in the loop

Too often, people may think of artificial intelligence as an android that's out to replace people at work. In truth, AI is typically more like a bionic arm that can make employees better and faster, whether they're data scientists, line of business experts, doctors, lawyers, care aides, or carpenters. And the ability to scale IDP largely depends on how people can help to train, use and improve the automation models that are augmenting human work. This is often referred to as Human-in-the-Loop (HITL) – and HITL is critical in both how you deploy and how you improve your IDP processes.

"Humans are still the center of all of this and your humans aren't going away," contends Chris Wells. "That's really important. Your process is going to be human driven, even with an AI assist. And I've never seen an unstructured data use case that can be fully automated without the input of a human."

Wells points specifically to training IDP models as an opportunity for scale and velocity: "When the subject matter expert is available to provide oversight and correction, the workflow achieves both acceleration and accuracy gains. Without an SME and an easy user interface, you'd need to train thousands of examples. Human-centered machine teaching lets you do it with hundreds in a day, instead of weeks. You have to put a human in the loop first."

Brandi Corbello agrees, and she points out the imperative for engaging employees on how they can and should engage with AI to be more effective. "You have to train SMEs appropriately. Upskill and reskill them as necessary. Get to uniformity on process by saying, 'This is how you've been doing it before, this is how we'd like you to label now.' That's how you'll get better results and more satisfied employees – people who don't keep getting paid to do mundane tasks and can create more value."

5. Don't be fooled by OCR

Optical Character Recognition (OCR) technology is core to IDP, often used to handle scanned documents. Images, of course, are not readily machine-readable, so computers can't immediately process what humans can clearly see as text in the document. OCR addresses that issue by identifying text in such documents and converting it to a digitized format that computers can manage and then automate processes with.

With OCR playing such a critical role to Intelligent Document Processing, it's unsurprising that many IDP providers will tout their OCR capabilities. But our experts warn that solution providers sometimes play fast and loose with their claims around what OCR can do from a processing standpoint, especially when it comes to unstructured data support. The vast majority of these are referring to basic template and rule-based approaches. When projects move from POC to production and encounter real-world variability, they fail...and fail hard.

"From my time as a consultant, I saw a lot of my clients blow up when they tried to scale with what their vendor sold them as flexible OCR," says Brandi Corbello. "That sounds great, but in reality it doesn't exist. An OCR vendor might say it could handle all of the company's different invoices, but we learned this wasn't the case. We ended up having to create templates for each individual vendor. And then, three months later, we'd redo it again for a new vendor. At a large company, that's thousands of vendors – and ultimately unscalable."

"OCR gets you machine-processable text, but that's not the intelligence you need to solve document understanding," echoes Chris Wells. "It's critical to ask what a particular solution is doing and what it isn't doing. Very often – and often too late – COEs ultimately find most vendors only solve part of the problem."

6. Choose technologies that are flexible enough to scale

While subject matter experts may not believe in the concept of "flexible OCR," they strongly acknowledge that scalability requires implementing technologies that are inherently flexible enough to scale. Enabling immediate and long-term scalability means being able to handle multiple use cases, manage multiple document types, understand multiple languages, and make use of documents that may be low quality or include handwriting.

"If you don't have the right tool, the other things don't really matter," states Chris Wells. "Like we've said, if a technology is rules-based, it won't scale. If it's a rigid point solution, it won't scale. If it's a platform that requires significant compute power, it won't scale. COEs should make sure that their technologies check all the boxes to give them the flexibility to expand."

Make IDP go further, **faster**

Enterprises have long struggled with their unstructured data. Though effective with structured data challenges, RPA vendors and point solutions have fallen down or fallen short with traditional approaches to automation. But now, the tide is turning thanks to breakthrough technologies like **Indico Intelligent Intake.**

Through its innovative AI and ML software, the Indico Platform allows enterprises to ingest unstructured data at massive scale and add structure, enabling them to do what's been impossible with traditional automation and analytics tools: realize the unlimited potential of their unstructured data.

For the first time, Indico gives enterprises a single solution that allows them to ingest and structure a diverse range of unstructured formats – text, CSVs, videos, audio files, PDFs, contracts, emails, and much more – and gain rich insights, as well as maximize the value of their existing software investments, including RPA, CRM, ERP, analytics, and more.

Now, Indico takes customers a quantum leap forward in handling virtually any unstructured data use case, with powerful new features to dramatically accelerate, scale and unlock Intelligent Data Processing:

Automatic document unbundling

Intelligently split a file into documents using a point and click interface – no templates or rules required. Unbundling provides a rapid solution to unpack even the most complex document bundles so organizations can unlock automation across more processes.

Linked labels

The Linked Labels feature enables you to extract data and automatically capture the relationships between document elements. With a more plug-and-play approach, there's no need to write post-processing scripts to reassemble the line item, table, calculation input, and other relationships in extracted data for downstream systems – accelerating deployment times enterprise wide.

Staggered loop training

Using human-in-the-loop data, the proprietary Staggered Loop system delivers a highly transparent, push-button process that takes hours, not weeks. This approach accelerates continuous improvement and empowers ongoing, next-gen Machine Learning operations.

Universal document support

Indico features out-of-the-box support for handwritten and hybrid text/handwriting documents, and it supports 70+ languages natively, including Chinese, Japanese and Korean. What's more, you can reduce costs by preserving valuable human resources for where they're needed by automatically processing both high- and low-quality scans with Microsoft ReadAPI.

Workflow canvas

Plan, deploy, and audit workflows with unparalleled ease. Indico's Workflow Canvas tool, featuring an intuitive visual interface and click-and-drag canvas, allows users to build and review the steps of each automation process, allowing for significantly more complex workflows and simplified auditing.

Ready to go further, **faster?**

With Indico, your Intelligent Document Processing can go further with empowering line of business, increasing efficiency and reducing costs. And go faster than ever before. To learn more, **request a demo and no-cost consultation at IndicoData.ai**

About Indico Data

Indico is The Intelligent Intake Company, enabling enterprises of all sizes to automate the intake and understanding of unstructured documents, emails, images, videos and more; analyze unstructured data, extracting actionable business insights and intelligence; and apply this data to create new application experiences to transform manual and inefficient processes into powerful solutions to solve complex business challenges. Through the Indico Platform, enterprises can gain rich insights and maximize the value of their existing software investments, including RPA, CRM, ERP, analytics and more.

For more information, visit IndicoData.ai

Sources

1. Deloitte, Automation with Intelligence 2020 Report

2. IDC, "FutureScape: Worldwide Artificial Intelligence 2021 Predictions"

3. Everest Group, "Scaling Up Intelligent Automation: What Can We Learn from the Best?"