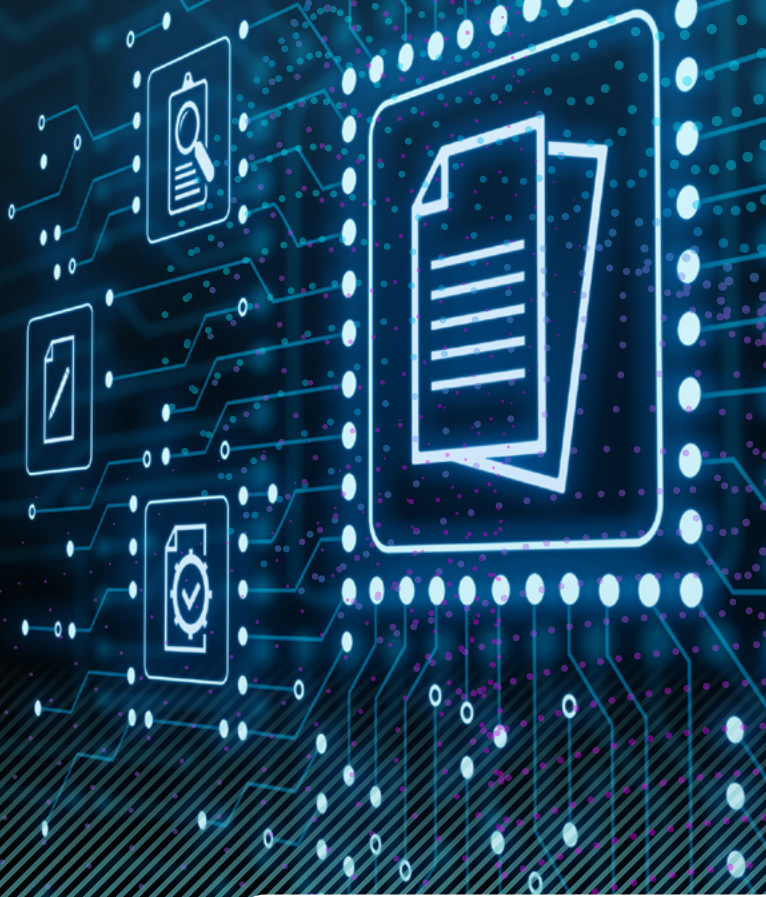
SSON | THE WORLD'S LARGEST SHARED SERVICES & OUTSOURCING NETWORK

HOW TO BECOME TRULY DIGITALIZED THROUGH INTELLIGENT DOCUMENT PROCESSING



Sponsored by

SINGULARITY SYSTEMS

Introduction

The age of digitalization offers tremendous opportunities for enterprises determined to leverage it for their advantage. Despite the fact that the past decade has shown automation technologies to be highly effective in improving process efficiency, it has also spotlighted the high cost of unstructured data.

Indeed, most of the data embedded in enterprise operations is still largely unstructured, and thus limits true progress towards digitalization. The goal, ultimately, is to flow data seamlessly across a digital ecosystem, leveraging algorithms for operating decisions. To this extent, the human-in-the-loop reduces operational efficiency.

In order to determine the extent to which organizations have recognized the challenges facing them and are taking steps to address these by digitizing data, SSON conducted a market survey in Q1 of 2021.

The results of the survey are shared in this report.

Editorial

Barbara Hodge

Global Editor and Principal Analyst, Shared Services and Outsourcing Network,
barbara.hodge@ssonetwork.com

"Digitalized" as the way FORWARD

Operational inefficiency has long been a nemesis for enterprises striving to stay ahead of the competition by reducing the cost of delivery. At the same time, increasingly discerning customers are putting pressure on organizations to minimize cycle times and prioritize experience. Both trends highlight the fact that too many enterprise resources are currently devoted to inefficient processes that fail to deliver value – and that the solution lies in leveraging "digital" for accuracy, speed and quality of service. Indeed, digital transformation is now on every board's agenda.

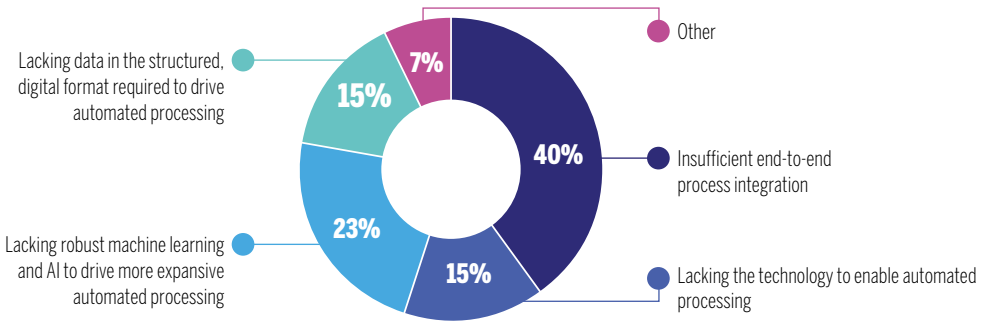
One of the fallouts of the pandemic was that it shone a stronger spotlight on operational inefficiencies. Respondents to the survey listed *lack of end-to-end process integration* as the greatest pain point resulting from last year's business disruption. Solving this is a priority for shared services leaders right now and any failures are closely linked to the inability to process unstructured data. Digitized data, in other words, supports the E2E solution shared services leaders are seeking.

Over the past decade, one of the preferred *digital* solutions has been robotic process automation (RPA) as a means of automating standard, rules-based processes. And RPA has, indeed, been effective as a task-based solution – not just in providing reliable transactional services but also in integrating applications with legacy systems to support process flow.

RPA is an essential technology to automate manual administrative work and to help alleviate the clerical burdens imposed by the limitations of legacy systems, like manually re-keying data. But it is not sufficient to achieve a digital transformation. To change the nature of work in your organization, you must integrate and orchestrate a variety of cognitive technologies, including data capture, intelligent document processing, natural language understanding, machine learning, process discovery, etc.

The goal, ultimately, is to allow data to flow freely through a digital ecosystem that links your stakeholders, internal and external, and allows as many operating decisions as possible to be made by algorithms fuelled by live data rather than by humans supported by their own data analysis. Every step of a process that includes human data processing is a point of friction that reduces operational efficiency.

How has the pandemic highlighted pain points in your current automation initiatives?



46% of global shared services plan to invest in new technologies to manage workflow / execution of processes in 2021.

Source: SSON GBS and Shared Services Survey 2021

With the COVID-19 pandemic forcefully introducing remote working practices across the globe, enterprises have recognized the stark truth that *digitalized processes* are the only way forward. This newfound awareness focuses on straight-through processing as an opportunity to optimize services, but also embraces "end-to-end" as a means of significantly shifting beyond incremental improvements.

The key objective is to provide a frictionless experience to customers, starting with their first point of interaction with the business.

What is preventing this from happening is the *lack of data in the format required*, explains Michael Grant, who leads North American sales at Singularity Systems, an emerging player in the data digitization scene.

"The vast majority of data flowing into organizations is in unstructured formats that cannot be consumed by machines. This leads to bottlenecks in processing capability," Grant says. "The goal of any digital transformation is to improve the flow of data through an enterprise, so transactions can be processed at machine speed**.** Data trapped in unstructured formats, like documents, forms, emails, etc., requires manual processing, which interrupts process flow and has a knock-on effect on upstream activity. In fact, SSON's survey confirms that more than 75% of organizations are struggling with significant levels of unstructured data, posing a significant hurdle to process continuity."

While automation unquestionably plays a key role in process inefficiency it can only succeed and scale, Grant confirms, if it is fed by data that has been cleaned and formatted for digital consumption: "This means *structured* and *digitized* so the automation software or other downstream systems can access, understand, and act on it, as appropriate," he adds.

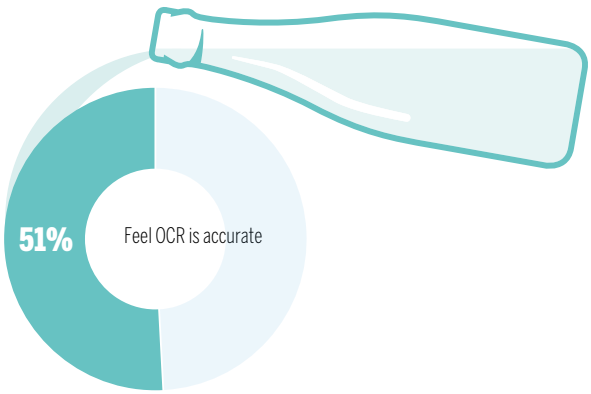
The **OCR** bottleneck

Optical character recognition (OCR) has long been the go-to solution for scanning and translating non-structured data into digital form, but to do so it relies on rules-based templates linked to common document standards. That is an approach increasingly out of touch with the reality of document variety organizations face in a digital ecosystem, says Dr. Tianhao Wu, Chief Technology Officer at Singularity Systems.

"OCR has proven quite *ineffective* when applied to the vast range of documents organizations face today," he explains. "OCR is primarily dependent on templates and zones to know where to look for data in a document. This works best when documents are consistent. But for something as simple as an accounts payable process, each vendor will have a different format for presenting the same information. Each new format might require its own template, which can inhibit scale. For something completely unstructured, like a contract, there is no template to tell the OCR where to look, so manual processing is inevitable."

Indeed, the majority of survey respondents confirm their current OCR solutions to be, on average, only 50% accurate, and only 20% of respondents are achieving accuracy levels above 80%. The impact this has on straight-through processing is clearly significant.

If you currently use an OCR solution, how accurate is it?



Source: SSON GBS and Shared Services Survey 2021

For organizations that have invested in RPA solutions but are now recognizing their programs hinge on the ability to *read* data, this is troublesome.

"RPA is, of course, not designed to *read* data but rather to *act on* data," Dr. Wu explains. "OCR has been the dominant solution to convert scanned documents into machine-readable text, but it only works when documents share a common structure, which leaves many use cases that involve complex documents out of scope for automation. This is one of several reasons why efforts to automate often are unable to achieve scale," Wu continues. "The amount and variety of data flowing into the enterprise is growing exponentially, which exacerbates the limitations of rules-based approaches to document digitization. The only way to harness enterprise data and extract value from it is to apply machine learning to the task of digitization and extraction from documents and images."

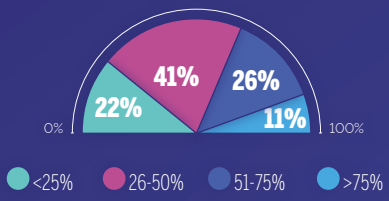
"Data is critical to process continuity, so data digitization should be a priority for all enterprises this year," says Dr. Wu. "What's limiting the success of digitally-driven transformation is simply the absence of structured and digitized data flowing through the enterprise."

"What's limiting the success of digitally-driven transformation is simply the absence of structured and digitized data flowing through the enterprise."

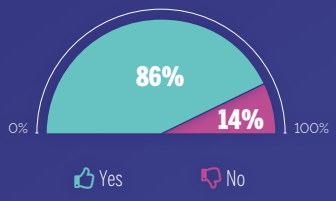
What about **data strategy?**

The challenge in processing documents is primarily one of *accessing the data in those documents*. And yet, despite the fact that 86% of survey respondents acknowledge the problem of unstructured data in their enterprise, half of them don't have a strategy in place to address that problem.

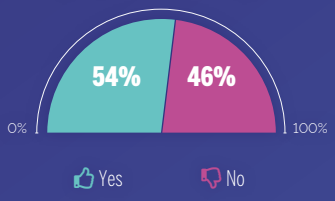
How much of your enterprise data do you estimate is currently unstructured?



Do you believe the problem of "unstructured data" is impacting your operational productivity / business effectiveness?

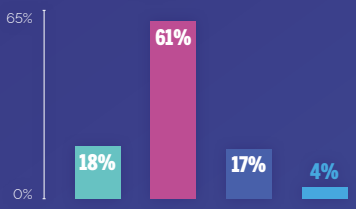


Do you have a strategy in place to address this data problem (i.e., unstructured)?



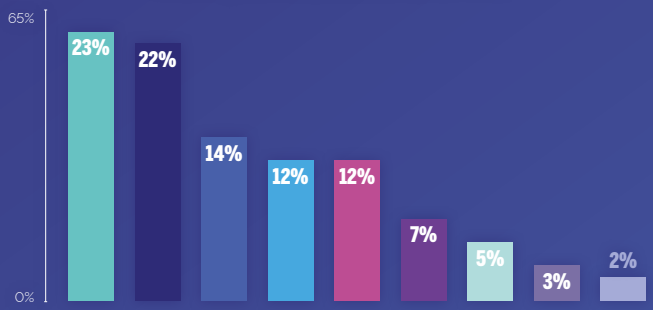
This trend is confirmed by SSON data* revealing organizations are staggeringly unprepared for the demand for digitized data. Sixty percent of organizations describe the extent to which they are *data-ready* as "basic" at best. Perhaps even more worrying: Less than 20% of organizations believe they are well prepared as far as their data is concerned.

How data-ready is your organization to support self-service and real-time analytics?



- We are data-ready (Staff are trained in data literacy, standardized rules exist about processing of data from disparate sources and external networks)
- We are still addressing requirements (Data is siloed, lack of measures to standardize quality of data across systems)
- We have yet to take steps towards this / we are not performing well
- N/A

If your organization is not currently data-ready, which option best accounts for this?

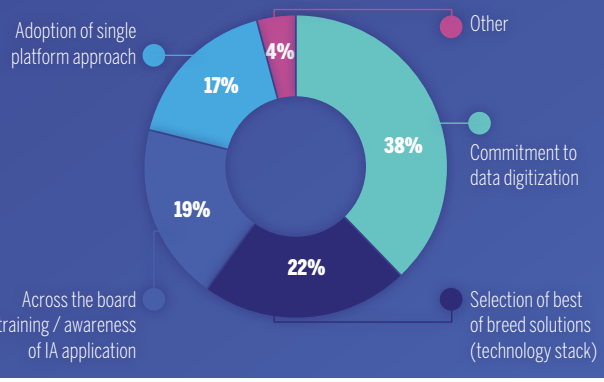


- Silo mentality in the organization limiting access/sharing
- Data not sufficiently organized/structured/digitized
- Lack of skills / talent with data analytics abilities
- We are data-ready
- Poor data reliability (Errors, incomplete data, lack of standardization)
- Inadequate IT infrastructure
- It is not a strategic priority
- Customer information privacy rule (e.g. GDPR)
- Other

"The lack of data digitization is not just a problem but a *roadblock* for enterprises that want to seamlessly manage the multitude of documents that initiate and drive their processes," explains Dr. Wu. "It's an unavoidable issue that needs to be solved today."

This urgency is recognized by the market: SSON's annual shared services industry survey showed enterprises listing "data digitization" as *the leading* priority to ensure the success of their intelligent automation initiatives this year*.

What will drive / enable future IA-based success?



FOOTNOTE: * SSON GBS and Shared Services State of the Industry 2021, SSON Analytics

Challenges in tackling data

Machine learning has the power to overcome the limitations of rules-based approaches to data extraction, but these projects come with their own challenges. Typically, machine learning projects require a large amount of data that has been properly labeled before you can even start training a model. This requires a team of data scientists, data engineers and business SMEs to work on data labeling, a process that is difficult to scale across multiple parallel projects without a large team of expensive data scientists.

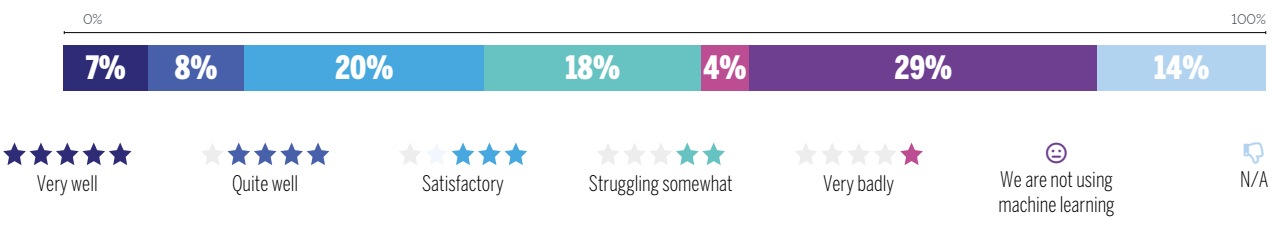
After the machine learning models are finally trained and deployed in production environments, they become less accurate over time. This is because business data is dynamic. The machine learning models and predictions become out of synch with the latest trends of your data and become less accurate as the data continues to evolve. Ultimately, the team of data scientists, engineers and SMEs must re-assemble periodically to retrain the models with yet again a large amount of labeled data.

This puts a premium on leveraging the latest advancements in deep learning, transfer learning and AutoML to make machine learning practical for enterprises to tackle the data challenge. Organizations can hire a data science team to tackle several important challenges, but for a machine learning solution to the unstructured data challenge to scale, it must be easy enough to use for the business analysts and process subject matter experts who live with these documents every day to achieve success without the data science team. The survey highlights just how big a challenge organizations are facing in this area.

Business users must be able to label data and train machine learning models on their own, if the business expects to digitize data at scale. Data scientists are too expensive and too scarce to require their involvement in every project. The survey tells us that three-quarters of those who are currently using machine learning are not succeeding with labeling efforts – their initiatives in this area are satisfactory at best. Nearly 30% of the market is currently not using machine learning at all.

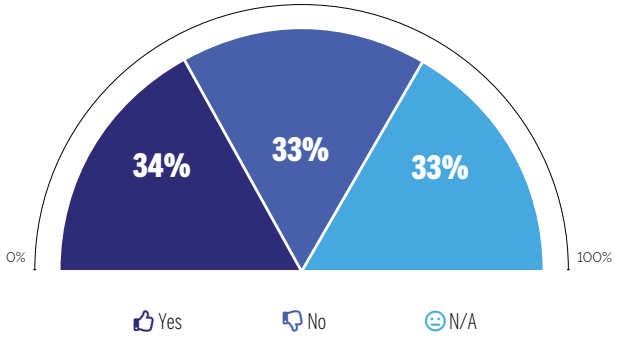
> "Despite the often significant investments made in RPA, many opportunities to optimize that investment are being hindered by inadequate data inputs. It's emerging as the efficiency gap across all industries."
>
> **Michael Grant, Singularity Systems**

How well are you currently managing data labeling for the purpose of supporting machine learning?



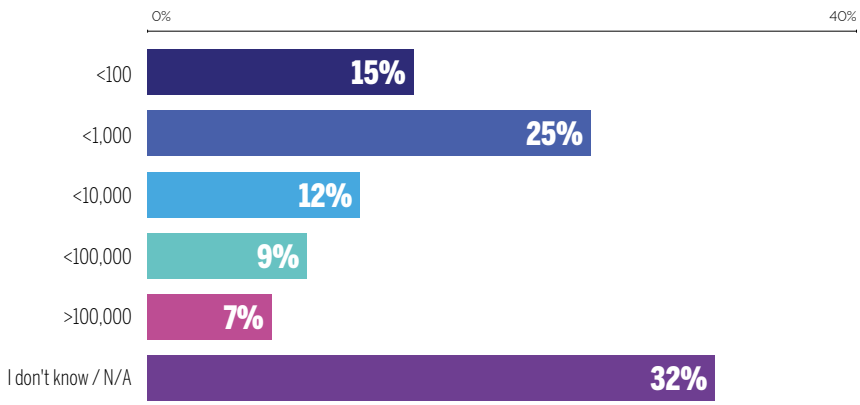
Another concern is that the current approach to data labeling requires skilled data scientists – a scarce and expensive resource that many organizations are struggling to access. Amongst organizations already using machine learning, there is a 50-50 split between those who *do* currently have access to data engineers to support data labeling, and those who *don't*. Data scientists need labeled data sets for the purpose of training a supervised machine learning algorithm but they do not want to be the ones doing the labeling. The solution is to turn your business line subject matter experts into data labelers, so these labeled datasets become by-products of ongoing business operations.

Are your data labeling efforts hampered because you don't have access to skilled resources (e.g., data scientists)?



Another problem is that many organizations require large volumes of documents to train their machine learning algorithms. The consensus amongst technology experts like Dr. Wu is that 500-1000 documents marks the sweet spot to train a robust model to read complex, unstructured documents, and the survey data confirms this. However, only 40% of enterprises are currently managing at this level, with many requiring far greater volumes of documents. Advanced techniques like transfer learning from large natural language and image backbone models are essential for reducing the amount of labeled data required to fine-tune accurate document models.

How many document samples do you require to train a production-ready solution?



Straight-through processing: THE HOLY GRAIL

Just as end-to-end (E2E) integration and process optimization lead to the best outputs, straight-through processing is a key objective for organizations focused on efficiency. The ability to funnel workflow "straight-through" without clunky handovers or human error is the Holy Grail for any process owner.

To date, however, *straight-through* is still relatively rare and, similar to end-to-end, those areas that are achieving the highest levels of success are characterized by highly structured data. Procure-to-pay, as well as order-to-cash and record-to-report do well in this regard. All other processes listed in the table below involve large quantities of unstructured documents, whether these are onboarding forms for HR; driver's licenses for claims; or financial reports for financial spreading.

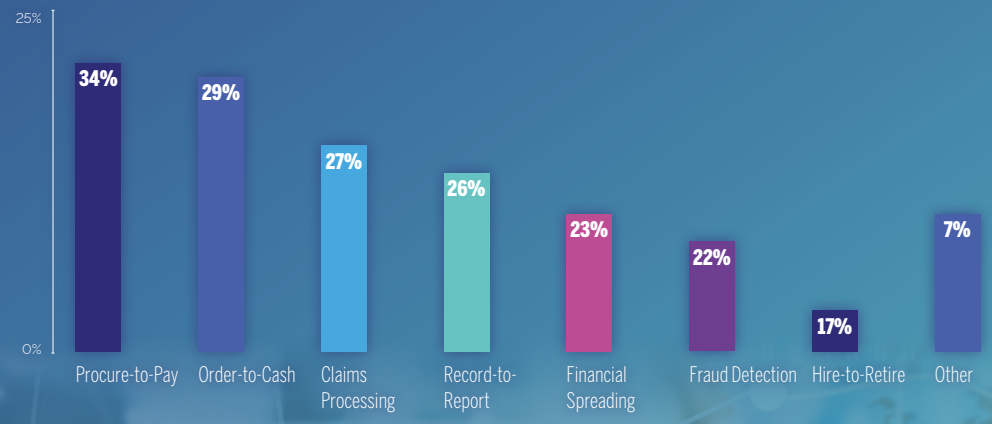
"We see a lack of digital format as the challenge for customers."

Survey respondent – Intelligent Document Processing Readiness Survey, March 2021

"Straight-through-processing from data ingestion to extraction to processing will require special treatments through Artificial Intelligence (solving for the constraints of pre-processing), Machine Learning (for exceptions management), and eventually maturing into Deep Learning (for expansive automation)."

Survey respondent – Intelligent Document Processing Readiness Survey, March 2021

What rates of straight-through processing **over 60%** are you currently achieving for each of the processes listed?



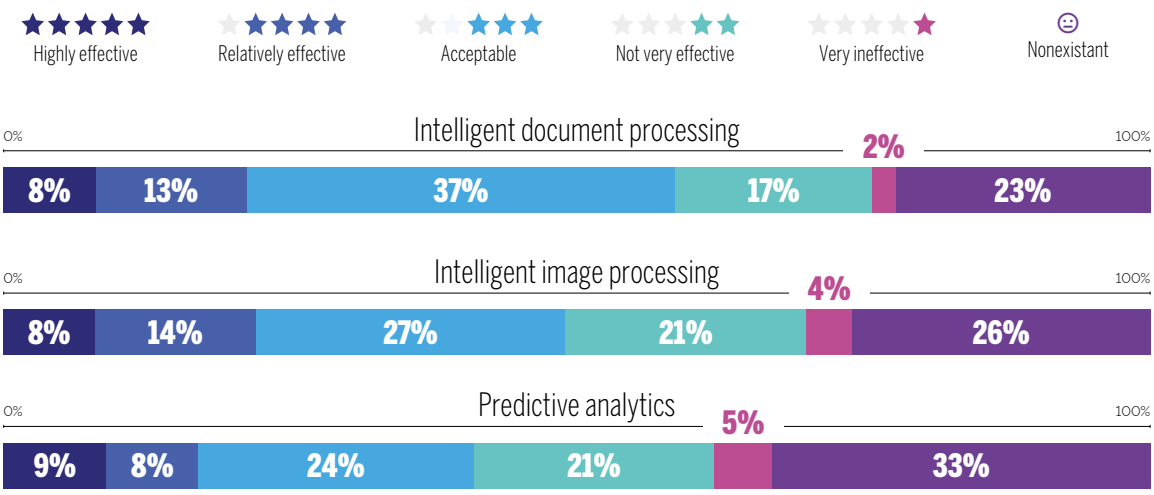
85% of enterprises not managing documents effectively

The fundamental problem most enterprises face is that current document management solutions are not effective enough in dealing with the vast range of modern-day document formats, starting with data digitization at point of entry. In a recent SSON survey, only two out of ten survey respondents described their current intelligent document processing / intelligent image processing solutions as effective. Eighty percent of enterprises, in other words, are working with ineffective solutions (or "acceptable," at best). At the same time, about a quarter of the market has not yet made a move towards integrating any document processing solution – a red flag for any operational excellence initiative.

On the process analytics front, results are even more discouraging: Only 17% of respondents believe their current predictive analytics capability is effective, and less than 25% are satisfied with their solutions.

> "Only two out of ten survey respondents described their current intelligent document processing / intelligent image processing solutions as effective."

Please rank your current solution effectiveness in terms of:



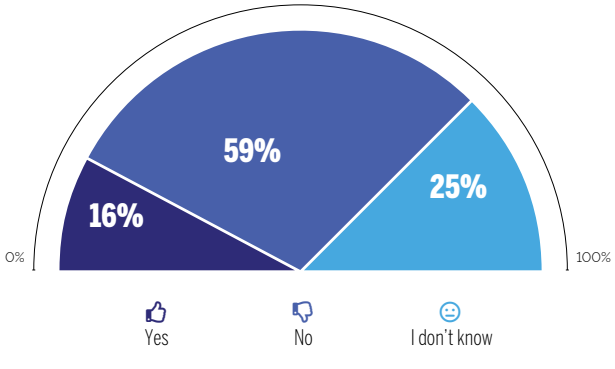
"Enterprises have not yet taken steps towards implementing and integrating machine learning solutions, which significantly improve the ability to extract and classify data from unstructured documents."

The main problem the survey highlights is that enterprises have not yet taken steps towards implementing and integrating machine learning solutions, which significantly improve the ability to extract and classify data from unstructured documents. This is re-enforced by the fact that 60% of respondents are not currently using GPU computing power in their automation programs (a prerequisite for efficient training of deep learning models), which implies a reliance on traditional OCR solutions.

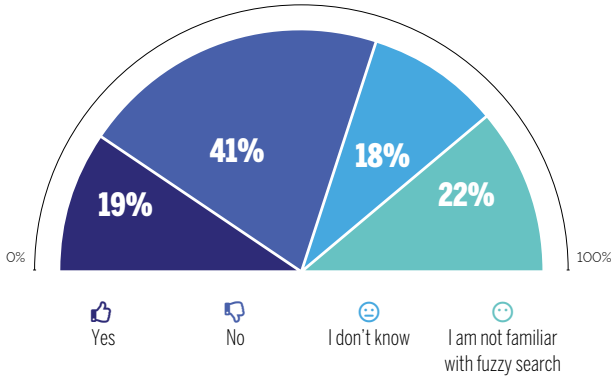
The stunning indictment: 85% of enterprises have not yet successfully addressed the challenge of managing the documents that drive their processes.

The reverse side of the coin is the lack of organizations embracing fuzzy search capabilities. As Grant explains: "Rules-based approaches to database searches, like a keyword search, can be too rigid. You can easily miss critical results if your search is not flexible enough to include results that are misspelled or contain synonyms of your key terms. Fuzzy search is more effective for searching archived data because it will include results that are not exact matches to keywords or meta tags but are still highly relevant. The fact that 80% of enterprises are still restricting themselves to rules-based approaches to source data and manage content is a serious concern."

Is your organization using Deep Learning and GPU computing power in its current automation initiatives?



Have you implemented 'fuzzy search' capabilities in your enterprise content-based search engines?



Other limitations emerging from the survey include the fact that three-quarters of the market is not taking advantage of *mobile* – and thus missing the opportunity to connect with 100% of their employees that carry mobiles.

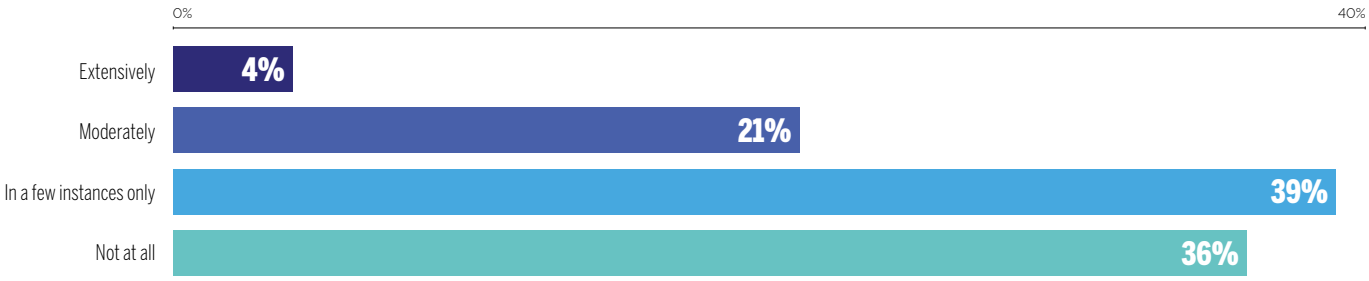
"As an input channel, mobile is second to none – just think of the accessibility!" explains Dr. Wu. "It should be taking the pole position but is as of yet untapped and underutilized, according to the survey data."

The implications for interacting with customers are also enormous, he adds: "It's a real opportunity missed. Less than five percent of respondents are *extensively* leveraging mobile apps to access data. So, a whopping 95% of organizations are missing out on the opportunity presented by a tool all of their employees and all of their customers carry in their pockets!"

Dr. Wu is referring to the ability to incorporate scanning apps in mobile applications that drive straight-through processing across document workflows. "Most of us have had the experience of depositing a check by taking a picture of it with our mobile phones. That is great for the customer, who does not have to go to the bank or find an ATM. But think of it from the bank's perspective: those mobile deposits represent millions of checks that don't have to be touched by a human. The customer digitizes the check for them with their mobile app and the check's data can be automatically exported to core systems without any human processing."

Dr. Wu continues: "One of our insurance customers has taken this approach with claims processing. When customers have car accidents, they submit mobile phone images of key documents, like driver's licenses and registration cards. They also submit images of the crash damage. We classify those images, extract the key information from documents, evaluate the crash damage to declare a total loss or generate a repair estimate. For this one customer, we process 400,000 images every day without any human touch. Mobile document capture will become a dominant input channel in future."

To what extent are you currently using mobile apps to scan and mine data?



"Humans-in-the-loop" *limiting volume*

The problem, of course, is not just the fact that data is unstructured, but that the sheer volume of documents enterprises are dealing with is growing constantly. The numbers are made worse as mobile engagements on the front end encourage even greater volumes, and in more and more different formats.

"The inability to process the volume of work that underpins enterprise operations – a.k.a. documents – is the biggest hurdle facing enterprise growth plans right now," warns Grant. "And it's a problem that threatens to overwhelm. The volume just cannot be handled by humans."

The survey confirms, in fact, that *humans* are what stand in the way of organizations successfully managing volume growth. Whether measured in terms of the overall cost of human resources, their limited availability, or the inaccuracy of human-based activity requiring rework … It adds up to a third of respondents listing the *human-in-the-loop* as preventing them from taking on more volume.

With most of the processing currently still done by humans, the ability to support growth is, today, severely limited, while the lack of sophisticated digitization and cognitive solutions is exacerbating the problem. Survey respondents acknowledge that the answer lies in more sophisticated learning modules that are able to manage a greater variety of documents, and an improved data digitization solution.

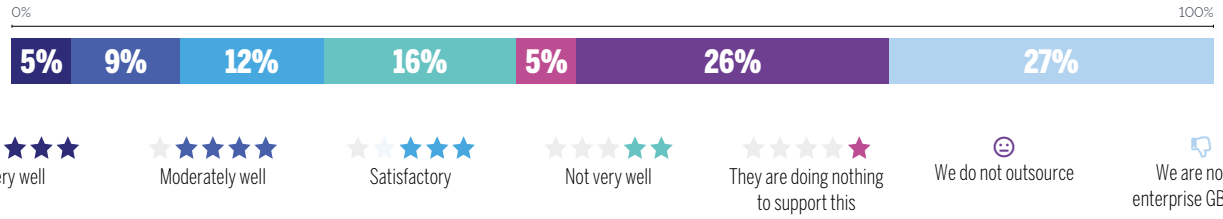
Are BPOs the solution?

Enterprises will have to work their way through the problems highlighted above to make progress. And while in the early 2000s, business process outsourcing (BPO) was the solution of choice for organizations looking for a competitive advantage, we don't see the same knee-jerk tendency when it comes to plugging today's processing gaps or missing skills.

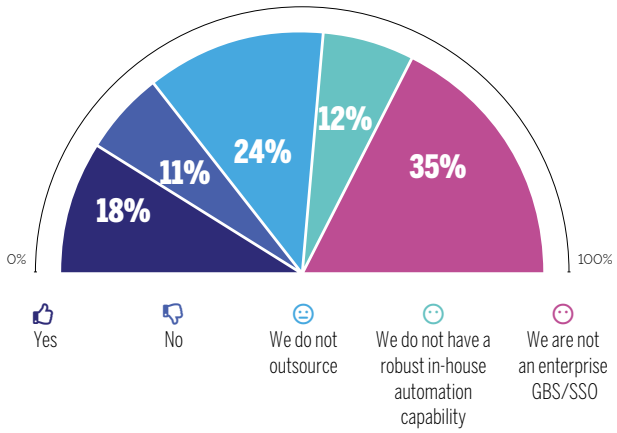
As the rising tide of automation has crept slowly but surely across business services, some enterprises are recognizing the opportunity to either push their BPO partners for more digital value-add, or bring work back in-house, now that it can be done by technology.

Results are not encouraging for BPOs. Of the organizations currently outsourcing work, roughly 70% are *at best* satisfied with their outsource providers' support of their digital transformation objectives. In fact, twice as many BPO customers feel their outsource provider is not supporting their transformation objectives as those who feel they are. A damning result, and one that is not sustainable given the continued prioritization of digital transformation on board's agendas. In addition, most of the enterprises currently working with BPOs but also building an in-house automation capability plan to bring some of this work back in-house once their automation is up to the job. Said differently: almost twice as many BPO customers are considering moving work back in-house, as plan to keep it outsourced.

If you are an enterprise GBS/SSO and work with an outsource provider (BPO), how well are they supporting your digital transformation objectives?



If you are an enterprise GBS/SSO working with a BPO, and IF you are ramping up your in-house automation capability, are you considering bringing some of the work back in-house?



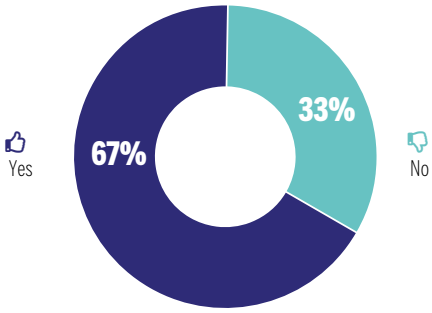
This message is being heard by BPOs. Two-thirds of them confirm they are under pressure from customers to offer digital transformation support.

Those BPOs that have already invested in a digital strategy will reap rewards. They will add to their current efficiency and productivity value-add the ability to harness *digital* to streamline operations and automate processing. Customers will benefit through improved speed, more straight-through processing, less waste, and more focus on revenue generating ideas, BPOs explain.

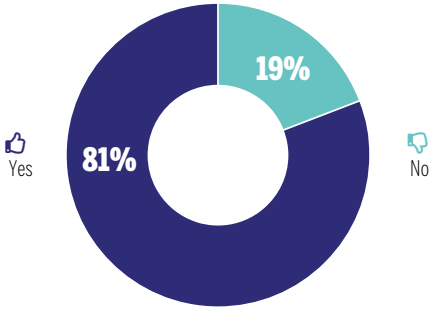
There is also, however, a business advantage for BPOs who are able to better mine customer data and therefore predict shifts in demand and meet emerging needs. They will also gain agility, given data-driven insights that support faster decisions and the ability to change course at any point.

BPOs responding to the survey confirm that the vast majority (80%) of their customers expect them to have a digital transformation practice of their own. The expectation is that outsource providers should be utilizing all their tools to drive down the total cost of operations and support customers in their digital agendas. Getting their own house in order should, therefore, be a priority for BPOs who want to remain relevant.

As a BPO provider, are you feeling pressure from customers to offer digital transformation support / services?



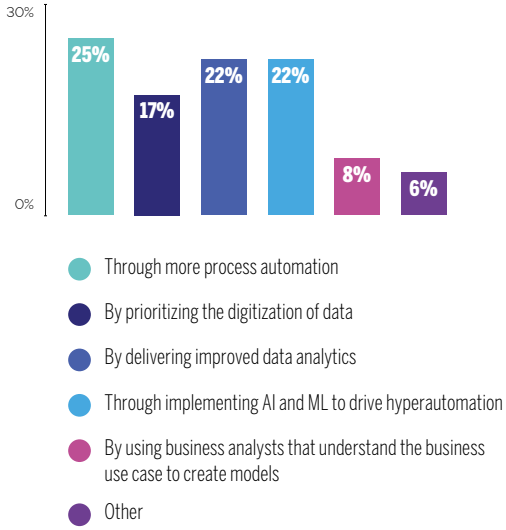
Do your customers expect you to have a digital transformation practice?



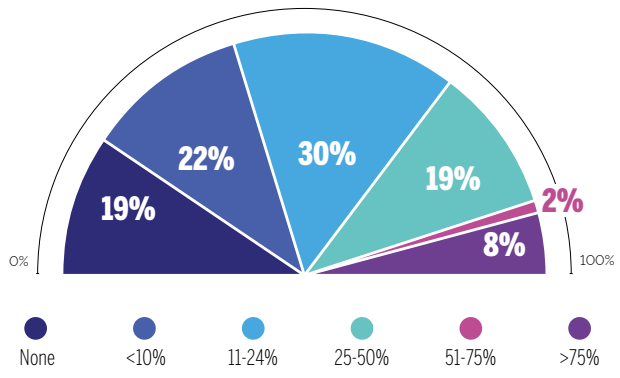
The survey highlights three core areas in which BPOs can support their customers' digital transformation objectives: process automation; improved data analytics; and leveraging AI and machine learning to drive hyperautomation. All three areas require a commitment to embracing digital strategies that is currently lacking, however.

Whether BPOs are meeting clients' expectations or not, customers are pushing down on margins. More than a third of BPOs say 50%+ of their customers are demanding price reductions in their contracts to reflect automation gains, and 80% of their customers have shared their plans to move work back in-house once they have a robust automation capability of their own. This is a challenge that BPOs will need to meet head on. Getting ahead of customers in terms of digital transformation is one way to retain their attention, and perhaps their business.

Where do you see the biggest opportunity for supporting customers' demands for digital transformation?



What percentage of customers do you see moving work back in-house once they have built a robust automation capability?



Intelligent document processing: the silver bullet for frictionless workflows

The survey data shared in this report points to a clear and present problem for global enterprise: an inability to rely on seamless, efficient document processing to drive business operations. It's a challenge that threatens to derail, if not destroy, organizations that do not heed the warning.

The solution lies in embracing intelligent document processing from the initial point of entry, via upstream processing, and through to the desired output. Modern digitalization solutions place this goal well within reach of modern enterprise. However, this requires a solid commitment to tackling the root cause of the problem (non-structured data) as it enters operations and throughout its lifecycle. The benefits are extensive, ranging from the plain and simple operational advantages of unencumbered processing, to improved customer experience through reduced cycle time, greater transparency and increased quality of service.

It's not just enterprises that recognize the writing on the wall. BPO providers that traditionally made their living from enterprise inefficiencies are having to wake up to the challenge of finding themselves unable to plug their customers' current and most pressing gaps.

Dr. Wu confirms: "Corporate leaders are urgently looking for modern ways to create customer intimacy, operate more efficiently, ensure compliance and reduce fraud. All of these objectives can be met by allowing data to move freely through your organization. When you optimize data flow you increase your customer satisfaction, and you decrease your cost of delivery. To remain competitive in the modern business environment, you must become a data-driven company."

Summary

Many organizations today still rely on human resources to keep their engines running. However, faced with the increasing variety and volume of document formats that underpin operations, this strategy is fast becoming unsustainable. The sheer cost of human resources, along with their propensity for error, is leading organizations to consider the benefits of a more *intelligent* document processing platform.

The challenge is that current data digitization strategies still rely largely on scanning solutions that are outdated. Such rigid, rules-based approaches are now giving way to AI-driven automated solutions, driven by machine learning.

Intelligent document processing leverages ML and AI to support and enable the entire process lifecycle. Building a machine learning model takes a matter of weeks and requires no more than 1000 documents. Even the data labeling hurdle can today be overcome through AI.

This new approach is perfectly designed for the digital ecosystem developing around us. Machine learning-based models know what data they need, where to find it, and how to process it. They also easily scale to manage increased volume. Intelligent data processing, therefore, does away with the human element that threatens to choke up operational performance.

The challenge is that in spite of the fact that most enterprises recognize the data problem, nearly half the respondents to the survey don't yet have a strategy in place to address this.

Time is fast running out.

SINGULARITY SYSTEMS INC.

Singularity Systems Inc. is a global provider of artificial intelligence solutions for intelligent document processing, intelligent image processing, and predictive analytics to help customers get value from their data. Our SinguAI platform combines NLP, Computer Vision, Deep Learning, and our proprietary OCR to offer a comprehensive solution for unlocking the value hidden in unstructured documents and images. The platform is self-training and adapting, providing real time interaction between the human and AI to empower business users with no data science background to build highly accurate models with small sample sizes, delivering value in days rather than months.

www.singularitysystems.com

ABOUT THE SHARED SERVICES & OUTSOURCING NETWORK (SSON)

The Shared Services & Outsourcing Network (SSON) is the largest and most established community of shared services and outsourcing professionals in the world, with over 170,000 members.

Established in 1999, SSON recognized the revolution in support services as it was happening, and realized that a forum was needed through which practitioners could connect with each other on a regional and global basis.

SSON is a one-stop shop for shared services professionals, offering industry-leading events, training, reports, surveys, interviews, white papers, videos, editorial, infographics, and more.

ABOUT SSON ANALYTICS

SSON Analytics is the global data analytics center of the Shared Services & Outsourcing Network (SSON), the world's largest community of shared services and outsourcing professionals. SSON Analytics offers valuable, visual data insights to support decision-making on the basis of data trends and benchmarks.

Thousands of global business services and outsourcing professionals use our data to understand the Shared Services landscape in their region through a variety of interactive data tools, analytics reports and customized data insights.

www.sson-analytics.com

Additional SSON Analytics Workbooks that may interest you:

- Global SSO & GBS Re-entry Strategies: Establishing New Rules in the Physical Office & the Future of Work
- Navigating The New Normal: APAC Shared Services Building Resilience with Automation & The Future of Work
- SSO Critical Survival Strategies in the Face of COVID-19