

Intelligent Document Processing (IDP)

White Paper

Introduction: A Solution for Paperwork and Data Overload

If there's a universal challenge across industries, it's the buildup of paperwork. According to **Forbes**, as much as 2.5 quintillion bytes of data are created every day. That's a lot of data in need of manual processing! Even though the world is rapidly digitalizing, document processing has lagged behind. Businesses often handle documents in the same way we did 70 years ago. From human resources (HR) forms and financial invoices to contracts and healthcare insurance documents, businesses across the globe spend countless hours reading, categorizing, extracting information from them, and making sure the correct action is taken in response. It's a daily routine, forcing employees to focus on manual, time-consuming, and mundane work, instead of higher-value tasks that achieve business objectives and drive company growth.

Fortunately, the digital age offers a solution that minimizes the manual, routine work of data processing and gets employees back to higher-value tasks. Advances in artificial intelligence (AI) have led to the development of software robots that can read documents of any difficulty, pull desired information, and pass it to where it needs to go with a high level of accuracy and reliability. The automation of various intensive manual business chores through robotic process automation (RPA) can substantially reduce the time and cost associated with routine document processing tasks. In fact, intelligent document processing (IDP), also referred to as document understanding, is one of the fastest-growing segments of the intelligent automation (IA) industry.

Gartner predicts that RPA can save finance departments 25,000 hours of avoidable paperwork annually. A **PR Newswire report** predicts the global intelligent document processing (IDP) market is expected to grow from \$860 million in 2021 to \$4.15 billion in 2026 at a CAGR of 37%.



The rising need for enterprises to process large volumes of semi-structured and unstructured documents with greater accuracy and speed, increasing investments in digital transformation, and the rising adoption of cloud-based document processing solutions are among the major factors driving the IDP market's growth. Among the industries best positioned to benefit from IDP software adoption are banking and financial services, insurance, healthcare, pharmaceuticals, government and public sector, fast-growth tech, telecommunications, retail, and travel and hospitality.

IDP augments the recognition of unstructured data through data science tools like computer vision (CV), optical character recognition (OCR), machine learning (ML), and natural language processing (NLP) in each stage of document data integration. Automated solutions are typically non-invasive and easily integrated with internal applications, systems, and other automation platforms.

The key benefits of implementing IDP affect these bottom-line priorities common to virtually all industry verticals:

- **Accelerated operational efficiency**

Spend less time and money on high-volume, varied document processing and more on real business investments.

- **Simplified compliance**

Reduce the resources dedicated to sorting out documents, key-data extraction, data validation, and other repetitive tasks that eat up time and money. Through IDP, these activities can be automated, which will, in turn, simplify compliance. That is possible because IDP leaves a digital trail that one can use for auditing and ensuring compliance with complex regulations.

- **Improved customer experience**

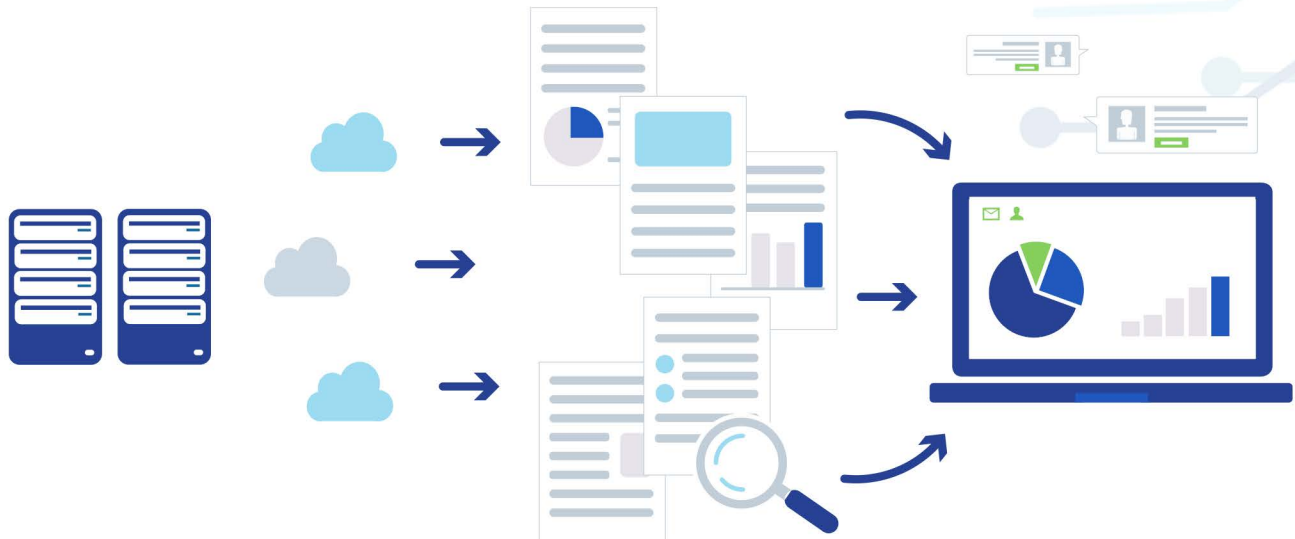
Accelerate the speed and accuracy of responses to customers to win their respect with an exceptional customer experience.

- **Risk mitigation**

Eliminate the risk of human error to save you and your customers from related losses.

- **Better employee experience**

Save employees from mundane tasks and let them focus on higher-value work and objectives.



How IDP Differs from Traditional Automation

Traditional automation follows rules to automate work that does not vary. This approach is effective for repetitive tasks strung together to form a simple business process. However, scalability and ROI problems rapidly emerge when task variation is introduced. AI-driven intelligent automation is superior to rule-driven automation, as it integrates all RPA functionality and adds the adaptive capabilities of bots that learn from data in real-time. In IDP, machine learning, computer vision, text analytics, and NLP process unstructured data and automate tasks that require recognition, detection, and adaptation to constant change by producing structured outputs from unstructured data and facilitating end-to-end automations.

What Can Robots Learn About Documents?

RPA puts software robots to work as digital assistants to take on highly manual and routine tasks with an accelerated rate of efficiency. As we route paperwork to them, they ease workloads and make room for more inventive human pursuits. In IDP, bots are trained to process documents by ingesting information and using applications to achieve the same objectives as humans following time-consuming document-handling processes.

Using AI, software robots can classify, recognize, read, extract, interpret, and act upon data from documents. When working with fixed-structure data sources like forms, passports, or driver's licenses, rules are sometimes sufficient to guide bots through thousands of similar documents with no need for AI. At the same time, documents with varying layouts or with no fixed structure—like receipts, bills, invoices, or resumes—require advanced AI skills which can automatically determine the location of data even if the layout changes. ML models are continually improving robots' skills to make them fast and accurate at document processing.

How IDP Supports Digital Transformation

Businesses seeking to harness the benefits touted by technology vendors (e.g., 80%-plus cost savings in one year) often start their transformation with RPA. An RPA-optimized foundation speeds the digital transformation journey and improves short-term ROI potential. In many cases, however, these organizations quickly find their automation program running out of steam after about six months —as either the scope of the RPA solution scope did not consider fundamental data entry (i.e., how data gets digitized in the first place) or the extensibility of the solution is limited due to lack of scalability up or downstream.

Equally challenging are the significant recurring costs of storing and using unstructured data, which in the case of documents comes from paper storage; manual work required to classify and extract information; correction of mistakes made during that work; limitations in peak volume handling; extensive document processing times; and efficiency constraints. IDP used in conjunction with traditional automation removes these costly constraints to deliver complete, end-to-end automation with incremental long-term benefits.

What to Look for in an IDP Solution

Maximizing value in an IDP investment begins with the consideration of two principal decision factors. The first is identifying what functionalities are needed for accurate data extraction from any specific document. The type of data to be handled plays heavily into this needs assessment. IDP uses rule-based and model-based approaches to process data from a range of document types. In general, structured data requires less advanced technology, while unstructured or partially structured data necessitates more sophisticated, adaptive technology. Software robots can be taught to classify, recognize, read, analyze, and process structured, semi-structured, and unstructured documents.

8 Capabilities that Make IDP Worth the Investment

Capability	Business Benefits
1. AI-native platform	Improves accuracy, performance, agility, and flexibility with built-in AI platform capabilities
2. Auto-learning	Powerful self-learning capability helps users to accelerate time to production and significantly reduce ongoing system support and maintenance costs.
3. Varied Documents	IDP solution can be adjusted to any document structure and business requirements to convert data into automation required structure.
4. Complex-document handling	Effortless, automated extraction of documents previously requiring skilled staff; touchless processing
5. N-level classification	Automation of labor-intensive document categorization; enablement of digitization, improved process efficiency
6. Scalability	Easy incorporation of new use cases, document types, and extraction fields thanks to document type-specific process models. Also, can be scaled both vertically and horizontally when deployed to support high volume and multiple projects for variety of document processing scenarios.
7. Ease of management	Localized control by business users without need for IT support resources comes with both on-premise and cloud hosted options
8. Cutting Edge Tech	Through use of IDP, automation program capabilities are broadened significantly with new type of technology able to train own Machine Learning models.



For **structured** documents, a rule-based, or template-based approach is used to determine what data to read and collect. For example, a rule can be established letting the robot know that the first box of every patient's form will be the patient's name, and the numbers in the top right corner will be the date.

For **semi-structured documents**, the type of information to be extracted is the same in a particular document type but is arranged in different locations from one document variant to the next. One example is receipts collected for employee expenses. Semi-structured documents may require more customization than templates or rule-based models can accommodate. A pre-trained or customized ML model is better suited and more flexible for semi-structured document processing. Hybrid solutions that combine rules and ML models for extraction are available

For **unstructured documents**, data extraction is complicated not only by the absence of consistent fields ("labels") but by the potential for labels to vary in position from document to document. ML models can accurately process unstructured documents like customized emails by teaching robots to translate written text into data. This complex level of data extraction requires robust ML solutions trained specifically for the task. Specialized engines like ICR, OMR, and OBR are capable of extracting data from handwriting, signature, checkboxes, barcodes, and non-Latin alphabet characters.

The second principal decision factor is the range of formats of inbound data. In any industry, data may be transmitted through paper, faxes, email, and attachments, electronic formats such as PDF or Microsoft Word, Excel, PowerPoint, etc. Additionally, data arrives from multiple sources and locations and originates in disparate devices (i.e. desktops, laptops, smartphones).

Structured

Repetitive forms and documents with a structured template that can contain handwritten text, signatures, checkboxes, (forms, passports, licenses, time sheets, etc.)

Semi-Structured

Documents containing fixed and variable parts like tables where the look & feel of the document changes with different variants or layouts (invoices, receipts, purchase orders, medical bills, bank statements, utility bills, etc.)

Unstructured

Documents with no fixed format (various contracts, agreements, emails, disease descriptions, drug prescriptions, news, voice scripts, etc.)



How IDP Processes Work

IDP systems classify, recognize, and extract distilled information and then route it to required document workflows for review. The process model behind an IDP system includes five stages:

1. Classification

The first step is to classify documents by type and determine the beginning and end of each.

3. Verification

IDP platforms are unique in their ability to leverage external databases and configured lexicons to validate information. Validation rules indicate an exception when results don't match with configured business conditions or confidence levels. Exceptions are flagged for human review and corrections.*

5. Auto-Learning

For both classification and extraction, the system continuously re-trains its underlying advanced & complex ML models based on examples of cases that required human intervention.

2. Extraction

Once a document is classified, its data is extracted.

4. Release

Verified information is automatically exported into business processes and workflows. It then becomes available for immediate consumption and use by the organization to take quick actions and provide efficient service to customers.

**To ensure 100% accuracy and enable underlying ML models to learn with human corrections, a dedicated verification component enabling human-in-the-loop verification tasks is available.*



Achieving Success with IDP

The key to success with IDP automation projects is document data literacy. Before the platform is trained to identify and learn data, a significant amount of time must be invested in understanding the information available specific to IDP projects:

- Various document types and layouts
- Fields to extract
- Business rules
- Scan quality
- Extraction of handwritten data

The business outcomes related to that information are also vitally important. To achieve document data literacy, it's critical to consult subject matter experts who use the information to produce work. Their intimate understanding of both the business value and interpretation of the information on the documents they work on will ensure the right data is extracted and what should be done with it.

To achieve this, the second most important factor lies in the business rules. One must understand and set the business rules using the tool-provided options, tailored-designed scripts, and integration with an external data source like Master file or Master DB. Lastly, proper selection and management of the dataset play a key role as the core of any IDP solution is continuous improvement and retraining of the advanced ML models used to digitize the documents.

Gaining a system-wide understanding of what your data represents and how it is used in the upstream systems paves the way for automated workflows through intelligent automation and business process re-design (as required).

3 Steps to a Successful IDP Journey

1. Establish foundation for measuring value

- Build awareness among executive leadership, employees, and client teams on the existing physical management of documents and what the technology can bring.
- Define the business case and roadmap of critical activity required.
- Prioritize backlogged IDP opportunities within high-value customer journeys.
- Select IDP solution and run a best practices pilot.

2. Expand the implementation based on early outcomes.

- Scale the pilot.
- Implement additional learning capability.

3. Continuously enhance and grow IDP capability.

- Scale IDP implementations to deliver significant value
- Utilize digital pipeline to manage the automation program end-to-end toward greater consistency, success rates, and ROI.



IDP Technology Costs

Most IDP tool licensing models base pricing on the number of pages to be digitized. The all-in cost reflects factors such as cloud or on-premises hosting, complexity, document volumes, etc. Depending on the total solution scope, prices range from ~\$0.1 to \$1 per page.

Value by Industry

Every company in the world has physical or digital documents, and they create more of them every day. Automation provides an efficient way to deal with high volumes of documents. A finance department of 40 full-time accounting team members can prevent 25,000 hours of rework and save the company \$878,000 per year caused by human errors, according to a **Gartner** press release. Multiply that value potential by the number of document processing functions across additional departments like HR and marketing. Businesses in document-intensive industries like insurance, banking, manufacturing, healthcare, and the public sectors will realize the most large-scale cost savings.

Following is a breakdown of some of the most common document processing workflows in which IDP can improve efficiencies.

Banking & Financial Services

- Mortgage Applications
- Know Your Customer
- Bank Statements
- Insurance Claims
- Account Openings

Healthcare

- Physician Referrals
- Claim-related Documents
- Patient Records
- Patient Intake

Travel & Hospitality

- Custom Declarations
- Driver Logs
- Maintenance Logs

Retail

- Invoice Processing
- Purchase Orders

Manufacturing

- Quality Assurance records
- Change Requests
- Sales Orders

Supply Chain

- Order Scheduling & Tracking
- Bills of Lading
- Proof of Delivery

Government

- Immigration Applications
- Education Applications
- HR Records
- Employee Onboarding



Conclusion

Automation has been around for decades, but only recently has it started to become what we think of as intelligent. IA combines RPA, AI, and other related automation technologies. Intelligent automation grinds through data to carry out tasks and deliver reports. It listens, reads, interprets, speaks, and analyzes. The smartest technologies also learn.

Mundane, time-consuming, and error-prone—we face many challenges when processing documents manually. RPA is a step forward. But as organizations scale their RPA initiatives, they frequently encounter limitations in capturing and processing the information from simple to complex multi-varied documents. Traditional OCR technologies and those in RPA platforms can struggle with:



**Varied document
formats**



**Interpreting
natural language**



**The presence of
distortion**

Insufficient intelligence means lower accuracy and much lower straight-through processing rates, impeding productivity improvements. With IDP, organizations can transform business processes by integrating more intelligent technologies to automate document-driven use case targeted opportunities.

IDP is an automation accelerator that captures documents before extracting, classifying, and verifying their data for further processing. The solution uses an integrated technology set of OCR, human-in-the-loop (HITL), ML, and RPA integrators. The power of IDP lies in its ability to handle multiple document types, processing any type of structured or semi-structured documents, ranging from invoices to know-your-customer forms to patient records, and much more.