

Intelligent Document Processing

Strategic Insights
for Choosing
the Right Solution

Sasidharan T

Delivery Manager, WNS-Vuram

Narendran Thillaisthanam

Chief Technology Officer, WNS-Vuram

WNS VU**am**

Imagine cutting down hours of tedious document processing, dramatically improving data accuracy and saving significant costs for your business – all while giving your team more time to focus on strategic initiatives. **Intelligent Document Processing (IDP)** is making this vision a reality for organizations worldwide. By extracting actionable insights from unstructured and structured data, IDP is revolutionizing how businesses handle documents.

The global IDP market was estimated at USD 2.3 Billion in 2024 and is expected to grow at an impressive compounded annual rate of 33.1 percent over 2025-2030.¹ This rising demand highlights the increasing recognition of IDP's benefits, including enhanced accuracy and substantial cost savings. However, businesses often face challenges in selecting the right IDP solution. The abundance of available tools, varying levels of accuracy, and complex implementation requirements can create confusion, resulting in poor tool performance, overestimated Return on Investment (ROI), and integration roadblocks.

In this paper, we delve into essential considerations for selecting the ideal IDP solution, breaking down core features such as Optical Character Recognition (OCR) precision, Artificial Intelligence (AI)-driven

data extraction, integration capabilities, and implementation methods. We further examine the importance of aligning IDP tools with business objectives, whether to drive automation efficiency, process improvement or scalability.

We also explore the financial considerations of implementing IDP, offering insights to help businesses evaluate ROI, ownership costs, and solution scalability. Lastly, we compare free and enterprise-grade IDP tools, exploring how the risks of free tools – such as limited support, reduced accuracy guarantees and a lack of customization options – can outweigh the perceived benefits.

By evaluating theoretical and practical considerations, we provide a comprehensive framework for selecting the most effective IDP solution. Whether considering cloud-based, on-premises or hybrid deployment models, organizations benefit from clearly understanding the key success metrics. With the right IDP solution, businesses can unlock unprecedented efficiencies, improve data accuracy, and achieve significant cost savings while positioning themselves for sustained growth in an increasingly digital world.



¹Intelligent Document Processing Market Size Report, 2030

Addressing Ambiguities in Selecting the Right Data Extraction Tool

Given the variety of technology tools available across sectors, finding the one best fit for purpose can be perplexing. Four key factors affect tool selection:



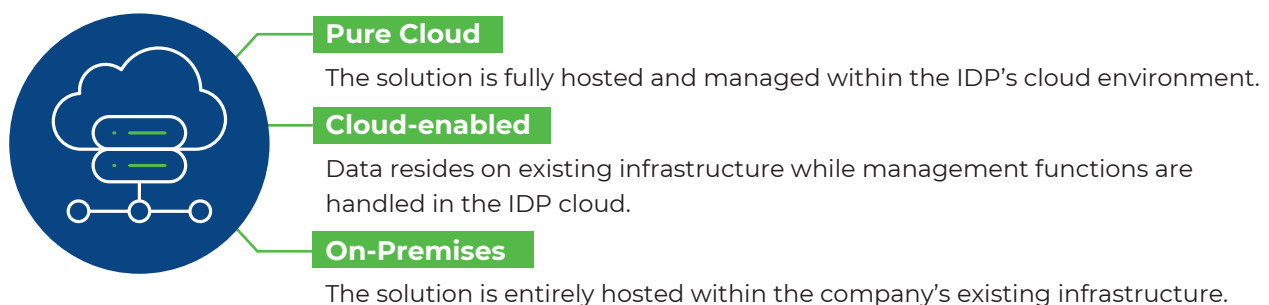
Consider the following scenario:



Given these conflicting signals, how does one determine which tool will **provide the most significant ROI**?

Next, consider the **contributions** of each tool **at every stage of deployment**. How does the selected tool generate value during implementation? Does it provide parallel support? And does it continue to deliver benefits after it is fully live? These assessments ensure the solution aligns with long-term business goals and delivers measurable returns.

IDP solutions are deployed in one of three ways:



Each method entails distinct considerations – such as the tool's features, security requirements, scalability, scheduling flexibility, and, of course, cost – that need to be evaluated to align with organization-specific needs.

Another critical factor in tool selection is the **accuracy of the data extracted**. The issue is not just about reliability but also how seamlessly this data can be integrated into the existing Enterprise Resource Planning (ERP) system without requiring extensive validation or manual intervention.

As the shift to automation replaces manual data extraction, **concerns about job security** often emerge as a barrier. In fact, automation does not eliminate jobs – it transforms them. By freeing employees from monotonous, repetitive tasks, it enables them to focus on strategic, higher-value work. This not only enhances productivity it boosts morale, empowering teams to contribute more meaningfully to organizational goals.

How to Choose the Best IDP Tool?

Selecting the right IDP tool demands a balance of theoretical knowledge and practical evaluation. A structured approach – using a scorecard, matrix or checklist – can simplify decision-making. Here is a breakdown of key aspects to consider:

Infrastructure, Implementation, and Integration

- **Integrated Intelligent Automation Platform:** Does the tool integrate seamlessly with existing automation systems?
- **Cloud-native Architecture:** Does it offer flexible deployment models tailored to business needs?
- **On-premises Solution:** Can it be deployed and managed within the enterprise infrastructure if needed

Monitoring and Scalability

- **Market Recognition and Skills Availability:** Is the tool well known and are there skilled professionals who can implement it?
- **Business-friendly Setup:** Does it allow non-technical users to onboard documents with ease?
- **Dashboards and Analytics:** Does it provide actionable insights, track metrics, and help visualize performance?
- **Online Marketplace:** Does it offer access to re-usable assets, pre-built automation, and digital tools that accelerate implementation?

Product Features

- **User-friendly Interface:** Is it intuitive enough for technical and non-technical users?
- **Business Application Integration:** Does it support essential platforms like SAP, SharePoint, and other enterprise tools?
- **Pre-configured Use Cases:** Does it include industry and function specific ready-to-use templates?
- **Self-learning Capabilities:** Does it adapt to new document designs without additional templates?
- **Automated Grouping and Templatization:** Can it automatically group documents and create templates?
- **Auto-language Detection:** Can it identify and process documents in different languages?
- **Flexible OCR Engines:** Does it support multiple OCR engines for document reading?
- **Custom String Operations:** Does it provide customization for specific data extraction needs?
- **Human-in-the-Loop Validation:** Does it allow human oversight for quality assurance?
- **Minimal Template Training:** Does it reduce reliance on manual template creation and maintenance?

Should Free IDP Tools Be Considered?

Choosing between a free and enterprise-grade IDP tool is much like deciding whether to park your car in a free lot or a paid one. While both serve the same fundamental purpose, the risks and rewards differ significantly:



Paid Parking

Offers better security, fewer risks, and more reliable service, albeit at a cost



Free Parking

While free, it carries higher risks, including potential damage and lack of reliability



The analogy applies to IDP tools. Here is a quick comparison:

FEATURE	Paid Tool	Free Tool
Accuracy Guarantee	Yes	No
Support Availability	24/7 Support	Limited or None
Business Customization	High	Low
Advanced Security Features	Yes	No
Integration with Software	Seamless	Difficult
Implementation Support	Yes	No

While free tools may appear cost-effective at first glance, their limitations -- particularly around accuracy, security, and scalability -- often outweigh the initial savings benefits. For businesses managing sensitive data or seeking high precision, enterprise-grade solutions remain the smarter investment.

To make an informed decision for the long term evaluate the tool's ROI by considering:



Net Returns and Total Cost

Ensure that the benefits (accuracy, efficiency, security, support) justify the total costs (licensing, implementation, training)



Post Go-live Metrics

Once the tool is live, track Key Performance Indicators (KPI) such as:

- Straight-through Processing (STP) rates
- Document failure rates
- Total pages processed
- License expiration reminders

Ensuring High-Quality IDP: Checklists, Cautions, and a Human-in-the-Loop

To ensure the successful implementation of the IDP solution, assumptions in the following areas need to be checked and rectified:

- › Source of the input file and business rule
- › Total number of fields to be extracted, with mapping details
- › Multi-page elements (e.g. tables must have header information on every page)
- › Total number of unique documents recommended for training the chosen IDP tool
- › Reminders on clearing the files that need manual validation
- › Solution for storing the extracted values

Several factors can negatively influence IDP accuracy:

- › Blurry images
- › Watermarked images
- › Handwritten or cursive text
- › Artistic or irregular font styles
- › Small text size
- › Complex backgrounds, shadows, glare or perspective distortion
- › Low contrast or overly dark images
- › Missing or oversized capital letters
- › Subscript, superscript or strikethrough text
- › Images with resolutions lower than 300 DPI



Assessing Pre-processing Needs

Most IDP tools have advanced pre-processing capabilities to prepare documents before data extraction using OCR tools. For instance, if the input content is a tilted – or even upside-down – document scan, top-tier IDP tools will automatically correct the orientation before extracting data.

Pre-processing addresses the following areas:



Image processing

Re-scaling | Binarization |
Noise Removal | Dilation / Erosion |
Rotation / De-skewing | Borders |
Tables Recognition



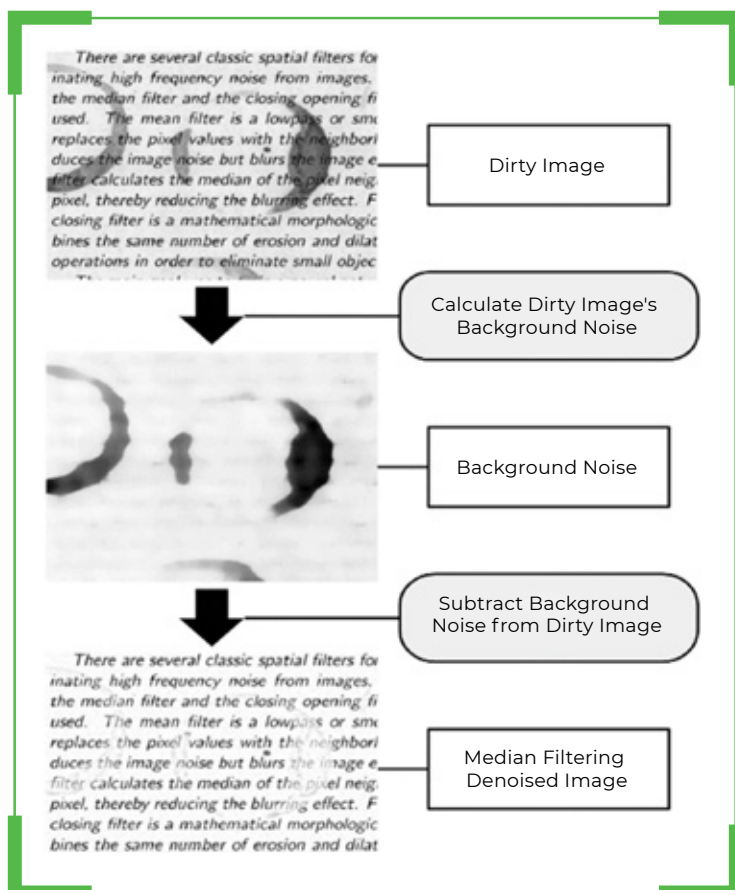
Page segmentation



Dictionaries, word lists, and patterns

Figure 1

An example of noise removal in image processing

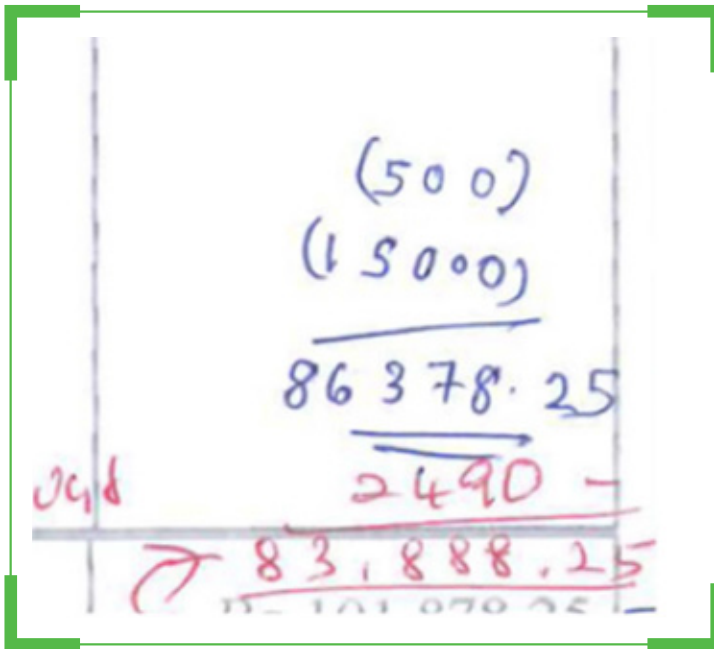


When selecting an IDP tool, the volume of documents to be processed is a key consideration. For low volumes in relatively clear formats, companies can leverage the free OCR tools usually bundled with leading Robotic Process Automation (RPA) suites. For higher volumes requiring pre-processing, tools such as Tessdoc can be employed to prepare the documents for OCR extraction.

Handwritten texts present an added level of complexity to OCR usage. While technology has improved significantly, no OCR tool gives reliable results when dealing with handwriting. For instance, a tool might read a handwritten bracket as 'C', '5' or 'S', as illustrated below.

Figure 2

Different OCR tools deliver different outputs for the same handwritten document



< Original document

OCR Output >

Tool A (Licensed)

C5001 0S0.0 863825 24q0

Tool B (Licensed)

(S00) (1S000) 86378.25 2490-

Tool C (Free)

c:%#)0 0S0^o-86878*;S

The ideal approach for improving accuracy in such cases is to **add a human-in-the-loop**. The system can employ a confidence score on the extracted data to trigger manual reviews. This approach is particularly useful when the document contains complex elements such as non-digital checkboxes or text embedded in images. For digital checkboxes and similar fields (such as radio buttons), leading IDP tools employ sophisticated Boolean logic to accurately identify and process the data, ensuring they are correctly interpreted without manual validation.

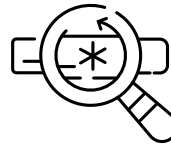
When working with IDP tools, a combination of pre-processing, advanced OCR, and human validation workflows is essential to ensure high-quality, accurate data extraction – especially for complex and handwritten content.

Assessing IDP Precision with a Matrix and Recommendation Meter



When there are concerns about accuracy, a great way to address these doubts is by using an IDP accuracy matrix. This tool helps visualize how various factors affect the accuracy of the solution. One approach is to apply the 80-20 rule, focusing on the most common or impactful document templates. By collecting sample documents with unique templates, we can train the IDP tool more effectively. Once these samples are processed, their accuracy can be evaluated and recommendations can be validated with the Center of Excellence (CoE) team. This process allows for a data-driven, transparent recommendation that instills customer confidence and provides clarity on the expected outcomes.

String Operations and Regular Expressions



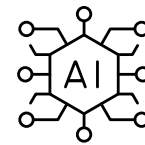
A fundamental part of preparing data for IDP involves string operations. These are essential for manipulating text data, including extracting substrings, removing unwanted characters, comparing values, converting cases, and determining string length.

Modern IDP solutions have built-in, customizable string operations that can be applied directly to extracted values. These operations are integrated into the document training process, eliminating the need for complex "if/else" conditions within RPA or other automation tools. This streamlined approach simplifies the workflow as string manipulation is handled within the IDP platform.

Beyond the default string manipulation functions available in RPA tools, more advanced techniques like Regular Expressions (Regex) can sometimes be useful. Regex are mathematical notations for locating, matching, and managing strings based on specific patterns. For digital documents with consistent variations, Regex can efficiently extract data without manual intervention.

Specifically, Regex is an optimal solution, as it can be customized to match specific patterns within digital documents, significantly reducing the need for manual adjustments. It is used to extract data at zero cost, whereas relying on a top-tier IDP tool can cost ~USD 4,500 per month for similar functions in our experience.

The Impact of Gen AI on IDP

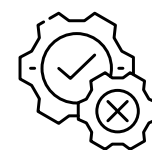


Currently, IDP requires sample documents to create unique templates for each model, with human intervention needed to train the model according to business requirements for data extraction. This manual effort – such as defining field mappings, positioning value regions, applying validation patterns or custom logic, and categorizing models – remains a significant challenge.

Gen AI is transforming this process by reducing the need for human involvement. As a broader category of AI capable of generating text, images, multimedia, and even code, Gen AI streamlines IDP by eliminating manual training wherever possible. Instead of relying on pre-defined rules, Gen AI can automatically detect patterns, interpret text meaning, and improve accuracy over time through machine learning.

Modern IDP solutions are increasingly focusing on self-learning capabilities, leveraging human-in-the-loop validation to refine models dynamically and reduce repeated manual interventions. As IDP tools evolve, Gen AI integration is becoming a key differentiator, making it a crucial factor in evaluating product capabilities.

A Brief Note on Non-Functional Requirements



When designing an automation solution, functional requirements typically address process logic, workflows, rules, and input / output specifications. Non-functional requirements focus on the quality attributes of the automation, such as scalability, performance, security, and reliability. These aspects are essential for ensuring efficiency, sustainability, and long-term success.

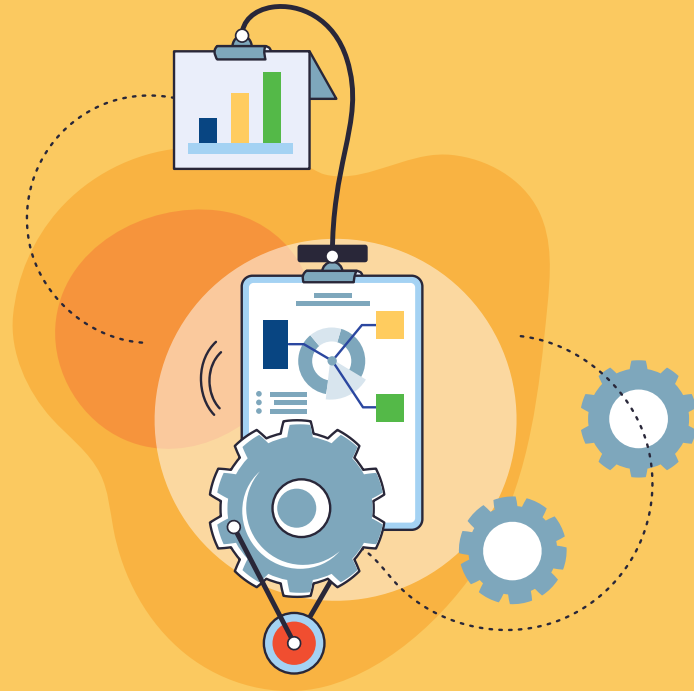
Conclusion:

The Future of Enterprise with IDP Tools

IDP tools have emerged as a critical solution for overcoming OCR automation challenges in the modern workplace. A thorough investigation into IDP technologies and the data they extract is essential for achieving successful implementation. Recent advances in IDP tools have driven significant improvements in accuracy, highlighting this technology's growing importance across industries.

Furthermore, a shift toward creativity and innovation is empowering developers to bring fresh perspectives, unlocking new possibilities for more effective automation solutions. As IDP tools continue to evolve, integrating AI and machine learning, they are set to revolutionize documentation automation and deliver transformative value across sectors.

To know more about selecting and implementing the right IDP solution for your enterprise, [talk to our experts today.](#)



About WNS-VURAM

WNS-Vuram is the global hyperautomation wing of WNS, with a specialization in low-code enterprise automation. Since its inception, the company has maintained 100% customer success and customer references. Powered by passionate people, WNS-Vuram is committed to driving digital transformation for organizations worldwide through an 'Empathy-First Technology-Next' approach. WNS-Vuram has received several prominent recognitions from Inc. 5000, HFS Research, ISG, Entrepreneur India, Fast Company, Business Brilliance Awards-Employer of the Year, and Great Place to Work, among others.

WNS-Vuram's hyperautomation stack – encompassing intelligent business process management (iBPM) / low-code application platform (LCAP), robotic process automation, intelligent document processing, process discovery, and analytics – drives maximum value and high-impact transformation in enterprises. With a strong presence across seven countries, WNS-Vuram offers end-to-end consulting, design, development, post-go-live support, and maintenance across all technologies via offshore, nearshore, and onshore delivery models. For more information, visit <https://www.vuram.com>.



**Inc.
5000**

*** ISG®**

WNS VUram

AUSTRALIA | CANADA | INDIA | MEXICO | NETHERLANDS | USA | UK
2025 Copyright WNS-VURAM™

WNEVERSTOP

Reimagining Possibilities to
Solve Complex Business Challenges