THE FUTURE OF INTELLIGENT ENTERPRISE IS AI-POWERED INTELLIGENT DOCUMENT PROCESSING (IDP)

Hyland™

Documents are everywhere in our personal and business lives — for example, employee onboarding forms, insurance claims, electricity bills, know your customer (KYC) paperwork and identification documents. Organizations spend vast sums of money getting people to manually input, edit and process documents. However, this is mundane, time consuming and prone to human error.

Thankfully, advances in digital technology have provided the opportunity for organizations to make this process a lot more efficient and to derive new insight and value from the data contained within their vast treasure trove of documents.

Intelligent document processing (IDP) is a digital solution that is transforming and digitizing organizations' documents quickly and cost-effectively. It is an increasingly sophisticated intelligent automation technology that improves the precision and throughput of document processing beyond traditional optical character recognition (OCR) and robotic process automation (RPA) technologies.

IDP uses artificial intelligence (AI) information extraction capabilities to read, recognize and understand the text and structures within an organization's multitude of differently formatted templates, forms and documents. Machine learning (ML) can then be used to "teach" the IDP platform how to make sense of documents and classify the data within.

THE OPPORTUNITY TO ACHIEVE MORE AND BETTER WITH IDP

Most of an organization's data is in a semistructured or unstructured format — for example, manuals, image PDFs or faxes. Invoices alone make up an estimated yearly volume of 550 billion documents, and other invoice-like documents add a further 5–15 times as many.[i]

Wouldn't it be great if technology could capture, extract, interpret and process data within those documents?

Documents come in a huge variety of formats and layouts: Some contain typed text, some handwritten; others contain checkboxes, tables or handwritten signatures. Some are easier to read than others.

An invoice, for example, will usually contain the supplier name, the amount due, the invoice date, the invoice number and a purchase order number — but different suppliers will put this information in different locations and formats on the invoice and may label it with different names or no name at all.

A human accounts payable clerk can identify the relevant numbers and names, despite the variation, and type it into a computer system, but this is time consuming, tedious and error prone.

Traditional OCR software wouldn't be very much help: It could recognize the letters and numbers, but it would just digitize them into an unstructured jumble of text. Extracting the relevant portions and correctly mapping them to fields like "Supplier" and "Date" in an accounting program would be as much work for a human as just typing in the information in the first place.

IDP offers a solution to extract and interpret content and context from an organization's documents using AI and ML models that keep learning over time to become even more accurate and precise.

Forbes suggests that 80% of all business data is trapped in unstructured formats such as documents, emails, images and PDFs.[ii]

It takes a few months to train an IDP system using supervised learning. This training approach involves giving it correct examples to follow, consisting of a set of documents and their expected output.

But once trained, IDP offers firms a tremendous opportunity to gain new data-driven insight and a competitive advantage from the tremendous volume of data contained within documents. More complex documents, or those in a language the platform doesn't support, will need to be processed manually.[ii]

Once data is extracted in a structured format, IDP can then be combined with RPA to provide businesses with a simple yet effective way to automate business processes end to end.

The combination of IDP with RPA bridges the gap between unstructured data, human verification processes and structured data that is required to automate processes.

For example, data extracted by IDP can be input programmatically into back-end corporate systems using RPA to achieve a greater level of automation or can be stored in a repository to augment strategic decision-making.

Effectively extracting data and structuring information is a gateway toward automating the lion's share of business processes that, today, rely on manual inputs and human intervention.

KEY USE CASES FOR IDP

In any industry and any department, you will find a multitude of different document types and vast swaths of people struggling to meticulously process them with precision and speed. This means that there are a lot of potential use cases for IDP.

$ Finance and accounting
- Invoice processing
- Bank statements
- Collections
- Receipts
- Rebates or returns
- Tax forms
- Expense receipts
- Bank statements

Financial services
- KYC
- Insurance claims
- Account opening and closing
- Mortgage processing
- Income validation
- Identity checks

Human resources
- Employee onboarding
- Resume screening
- Identity documents
- Application processing
- Benefits management
- HR records
- References

Healthcare
- Patient registration
- Patient onboarding
- Patient records
- Processing claim-related documents
- Invoices

BFSI industry
- KYC compliance
- Bank statements
- Check processing
- Fraud detection
- Mortgage documents
- Claim management
- Policy administration

Government
- School or university applications
- Passport or vehicle license processing
- Immigration services
- Customs documents

Supply chain
- Customs declarations
- Bills of lading
- Proof of delivery receipts
- Order scheduling and tracking
- Insurance documentation

Procurement
- Purchase order processing
- Customer onboarding
- Vendor onboarding
- Contract administration
- Customer contracts
- Tenders

Manufacturing
- Sales order processing
- Data sheets
- Labels and packaging
- Customs (import/export) documents
- Rebates and refunds
- Compliance documents
- LEED letters
- Quality assurance records
- Regulatory documents

CASE STUDY

A manufacturing company used IDP to automate the processing of over 400,000 invoices per year from 500 different suppliers. The technology incorporated OCR to identify the characters in a scanned image of an invoice, natural language processing to understand the words and ML to recognize patterns in the structure and layout of invoices and improve this understanding over time.

This transformation reduced the workload of the 29-person data entry team by over 70% in four months, freeing them from the repetitive and boring work of keying in invoices. The business upskilled six of them to oversee the new process: Training the IDP system, checking the quality of its work and handling the remaining documents that were too complex for it. Two more of the team were put in charge of deploying similar digital transformations in other departments of the business. The remainder were reskilled to work on other more interesting and value-add tasks elsewhere in the business, such as business performance commentaries and analysis.

CRITICAL COMPONENTS OF ANY IDP PLATFORM

To deliver IDP capabilities, organizations need to capture documents and ingest, process, classify and validate them, verifying them either digitally or with a human in the loop. Finally, the data from the documents needs to be intelligently processed and embedded into systems and applications.

Below is a typical flow for an enterprise-grade IDP solution:

1. **Document ingestion:** Ingest a document in electronic or paper format.
2. **Image processing:** Process the document image to improve image quality.
3. **Indexing and classification:** Classify a document into the correct document type using text mining and/or ML algorithms.
4. **Extraction and validation:** Extract and validate data contained within the document. To do this, an IDP system needs to have been trained on samples of similar documents. Apply predefined validation rules to extract valid information and insights contained within the document and to categorize and organize the document itself.
5. **Human-in-the-loop verification:** IDP tools provide interfaces that permit human validation of the extracted data as needed to cut down on the use of "people" in close succession. People can check that the correct information and insight have been extracted and further improve the accuracy of the ML model by training it.
6. **Delivery of data to another system:** Accurate data that has been extracted from a document can then be sent to another computer application or used in a decision-making process (e.g., ERP, CRM Snowflake or a decision model).

HOW TO CHOOSE THE RIGHT IDP PLATFORM

There are a wide variety of IDP platforms to choose from in a growing market. According to Everest Group, the estimated global market for IDP in 2020 was $700–750 million, and in July 2021, it predicted that it would grow at a rate of 55–65% over the coming year.[iii]

But how should organizations best decide which IDP platform is most suited to their needs? The selection criteria below can be used to judge the suitability of an IDP platform:

1. The accuracy of the extraction, verification, processing and classification engine, including the ability to ingest handwriting, tabular data or bar codes and to understand semantic context.

2. The ability to automate the capture, extraction and verification of data from documents and images — of different file types, in different languages and across departments.

3. The ability to perform image cleanup functions like line straightening, removing lines and dot shading and character enhancement.

4. An intuitive user interface for ease of use, particularly if an organization wants its line of business users to own and train their ingestion models.

5. The ability to program and extend an IDP platform using low code. Organizations are rapidly moving toward highly configurable low-code platforms with intuitive tools like checkboxes, radio buttons and configurable menus that enable them to make changes at any time, without expense or time-consuming and difficult-to-maintain coding or scripting.

6. ML models with the ability to determine where the important information lies — e.g., in images, tables, documents or bar codes.

7. Automatic identification of a document or documents, ensuring the correct information is extracted from each page, regardless of the business process.

8. Out-of-the-box functionality and the ability to expand on that to meet increasingly complex business requirements.

9. Access to key metrics and performance indicators for your processes and metrics that highlight the accuracy and performance of the IDP platform.

10. IDP data logging for security risk assessment, compliance and auditing purposes. The ability to log and retain records, enforce structured data retention policies, support information governance and minimize legal risk is evermore important in the data age.

11. Built-in intelligence that continuously improves over time, further reducing manual touch points and accelerating document processing. This eliminates the burden of manual processing and data entry as well as accelerates downstream classification and data processing, which improves information accuracy across business systems.

12. Extensibility to integrate with, and directly deliver data to, other platforms, systems or technology solutions, such as enterprise resource planning (ERP), enterprise content management, data management systems, business process management or other business technology solutions without disruption. This ensures that teams are working with the most complete and accurate data needed as they make critical business and process decisions.

13. A portfolio of complementary platforms and products offered by the IDP vendor, such as case management, workflow, intelligent automation or RPA technology platforms.

CRITICAL SUCCESS FACTORS FOR MAXIMIZING BUSINESS IMPACT FROM AN IDP PLATFORM

Implementing any business technology platform takes careful planning. There are challenges to implementing an IDP platform that may limit its success, but done right, an IDP program can deliver benefits far in excess of its cost. To get the best results from an enterprise IDP solution, follow these five steps.

1. **Build a strong business case that articulates the benefits of IDP.**
   - Run a proof of concept to gauge the suitability of an IDP technology solution within your organization.
   - Demonstrate a technology platform solution to key stakeholders and seek their input, direction and feedback.
   - Deliver workshops to collect and prioritize a backlog of high-value IDP business opportunities.

2. **Seek executive approval and sponsorship**.
   - Get an executive sponsor to champion the IDP program at executive, board and company levels.
   - Present your business case, including total expected benefits, and your roadmap for implementation to the executive team to get approval for a program budget.
   - Set realistic expectations on implementation time and running costs.
   - Outline a clear governance structure to track business opportunities and benefits over time.

3. **Reach out to vendors and select an IDP platform and implementation partner.**
   - Create a vendor selection matrix and assess each IDP vendor and their platform.
   - Engage your chief information security officer (CISO) and your risk, finance and procurement teams to negotiate a win-win deal with your IDP vendor of choice.
   - Select a secure, enterprise-wide IDP platform.
   - Go to market and select an implementation partner if your vendor does not have an implementation team.

4. **Roll out an IDP solution.**
   - Start small but think big: Scale the IDP solution quickly to deliver significant value back to the business on a consistent basis.
   - Build a center of expertise to own and drive IDP adoption, best practice and use.
   - Develop a comprehensive training and change management program to gain and retain buy-in for your IDP program.
   - Train the IDP ML models using large volumes of use-case-specific documents and data.
   - Focus on data quality and integrity by employing data governance controls that are supported and enforced by data champions.
   - Continually monitor IDP metrics and performance indicators for your processes that highlight the accuracy and performance of the IDP platform.

5. **Scale your IDP solution across the enterprise.**
   - Rapidly scale your IDP solution. For example, start with invoices in a single country or documents from the top vendors within a geographic region, then rapidly scale across functions, regions and international geographies.
   - Utilize complementary technologies, such as intelligent automation or RPA. This will help accelerate the flow of data between applications, systems and people to ensure that your staff members are always working with the most accurate and recent information as they make critical business and process decisions.
   - Continually extol the benefits and impact of your IDP program to retain business interest and attract further funding and ideas for your IDP delivery program.
   - Govern the delivery of IDP to ensure value is delivered over time.

CONCLUSION

In an increasingly digital world, information is quickly becoming an organization's most valuable asset. But relying on staff to manually receive, process and input information into your core systems is expensive and leaves room for errors, slower processes and delayed and inaccurate decisions.

Intelligent document processing can accurately extract and speedily classify data held within organization document repositories. IDP platforms are now capable of accurately ingesting high volumes of documents at scale with outstanding precision and speed at a price point that suits most budgets. It is now possible to drive consistency, success and enhanced ROI from a successful IDP delivery.

There will always be documents and data fields that fall beyond an IDP system's capabilities. But thanks to AI and ML, IDP platforms are constantly becoming more powerful, accurate and quick to train.

IDP allows organizations to unlock hidden information and value from their vast document libraries and frees the intellectual capital of its workforce, who can now spend time on more interesting and valuable work. The data unlocked by IDP can enable better decision-making and thereby transform today's businesses into intelligent digital organizations.
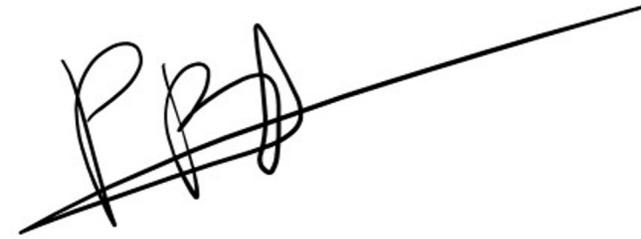
ABOUT PASCAL BORNET

Pascal Bornet is an expert, author, keynote speaker, influencer and thought leader in artificial intelligence and automation. He is a pioneer in the field of intelligent automation (IA) and the author of the best-seller book, "Intelligent Automation." He has received multiple awards, and he is regularly ranked as one of the top 10 global experts in the fields of AI and automation. Bornet has published articles in Forbes, Bloomberg Businessweek, McKinsey Quarterly and The Times. He is a member of the Forbes Technology Council and a senior advisor for several startups and charities. After 20+ years of experience in consulting and building the Intelligent Automation practices at Ernst & Young and McKinsey, Bornet is currently on the leadership team of Aera Technology, an innovative startup.

Learn more at **Hyland.com/RPA**

i. Billentis, *The e-invoicing journey 2019-2025*, 2019.

ii Forbes, *Machine Learning Can Set Your Document Data Free — Here's How*, 2021.

iii Everest Group, *Intelligent Document Processing (IDP) Adoption Swells as Enterprises Seek to Lower Costs Through Automation; IDP Market to Grow 55-65% in Next Year,* 2021.

Hyland™