

August 2023

Intelligent Document Processing: Ingestion Powered by AI

Deb Zawisza



Table of Contents

Summary and Key Findings	2
Introduction	3
Methodology	3
The Market.....	4
Ecosystem Map	6
Functionality	7
Solution Types	9
Vendor Metrics	9
Prominent Providers.....	11
General-Purpose Solutions	12
Insurance-Vertical Solutions.....	13
Fit-for-Purpose Solutions.....	14
Conclusion.....	15

List of Figures

Figure 1: Datos Insights P/C Core Ecosystem Map.....	6
Figure 2: Key Functionality Trends	7

List of Tables

Table A: The Market.....	4
Table B: IDP Solution Providers	11

Summary and Key Findings

Applications of artificial intelligence (AI) in the insurance space are continually evolving with the release of new solutions and increasing efficiency. For claims and underwriting, this has led to an evolution from optical character recognition (OCR)-based text ingestion to broader processing of unstructured data. Intelligent document processing (IDP) now covers the ingestion of a broad range of structured, unstructured, and semi-structured documents, such as submissions and claims notes. IDP solutions have evolved into three broad categories: general-purpose solutions, insurance-vertical solutions, and fit-for-purpose solutions.

As insurers begin to evaluate and implement IDP solutions, it is essential that CIOs remain grounded, provide key considerations for evaluating individual solutions, and provide their teams with realistic expectations for the solution's performance and applicability.

The key findings of this report follow:

- Modern IDP solutions are capable of extracting data from structured, semi-structured, and unstructured documents.
- IDP solutions have become commonplace, and adoption is growing, but price points remain a barrier.
- In evaluating IDP solutions, carriers should consider whether to pursue general-purpose, insurance-vertical, or fit-for-purpose solutions.
- Key metrics for consideration include how broadly applicable the solution will be, how long it will take to achieve value, and how easy it will be to tie into the rest of the insurer's ecosystem.

Introduction

Insurance companies deal with an overwhelming volume of policy documents, claims forms, invoices, underwriting records, and other paperwork. They face formidable challenges in managing, processing, and extracting insights from the wealth of information available to them. Conventional approaches to document handling have been characterized by manual data entry, sorting, and analysis, which not only incurs significant costs but also introduces a considerable margin for human error, leading to inefficiencies, delays, and compromised customer experiences.

IDP solutions, however, present carriers with an opportunity to streamline their document-centric operations. IDP solutions have the potential to transform document-intensive operations, increasing the amount and accuracy of data available for analysis while reducing the time and effort required to unlock that data. By leveraging advanced OCR, natural language processing (NLP), and other AI/machine learning (ML) techniques, IDP solutions provide the potential for highly accurate and rapid extraction of structured data from what have historically been unstructured or poorly structured formats, such as claims submissions, medical information, documentation from attorneys, loss runs, handwritten documents, and many other types of documents.

Methodology

This report provides an overview of the vendor marketplace for IDP solutions. Vendors continue to enhance their offerings, and this report represents a point-in-time view of capabilities. The vendor profiles are based on publicly available information, selected vendor briefings, and responses to survey questions.

The Market

Despite significant investments in digitization, unstructured documents are a key part of an insurer’s operational processes. Insurance carriers use administrative staff or business process outsourcing operations to manually enter information from submission documents, risk assessments, and financial information for underwriting. Similar processes are in place to handle the variety of documents needed to process claims. The variability of form and format is a characteristic of most information that flows through these processes.

Solution providers have offered OCR capability to the industry for years; however, accuracy and flexibility have been challenging. AI, including deep learning and large language models, is reshaping solutions to ingest information from a variety of document types and sources. As a result, the market is rich with very large horizontal providers and newer entrants that have fully embraced AI in their solutions.

A subset of the large horizontal providers and smaller, newer entrants have developed models trained specifically on insurance documents and use cases. By pre-training models on data that more closely resembles their day-to-day use, these solution providers have decreased the model training required to implement an IDP solution and accelerated time to accuracy.

In addition to greater accuracy of information ingestion, additional data beyond minimal requirements can be captured to fuel data repositories for analytics. These insights can enable carriers to better understand their overall submission portfolio and analyze claim trends and characteristics. Moreover, IDP can enable insurers to reduce underwriting and claims adjudication cycle times.

Table A summarizes trends in the IDP market and their implications.

Table A: The Market

Trends	Implications
Expansion of AI capabilities	<ul style="list-style-type: none"> • Enables processing different document types and formats • Provides carriers with a wider set of use cases across business functions

Trends	Implications
Data extraction into third-party services (e.g., risk models, underwriting workbenches, and policy administration systems)	<ul style="list-style-type: none">• Enables the use of advanced analytics to assess submission portfolios, claims trends, and performance metrics• Enables additional opportunities for automated calls to external data sources and straight-through processing of more complex business
Well-trained models with insurance artifacts	<ul style="list-style-type: none">• Enables low time to accuracy and time to value via pre-training• Provides carriers with proven use cases, requiring relatively little model training to implement

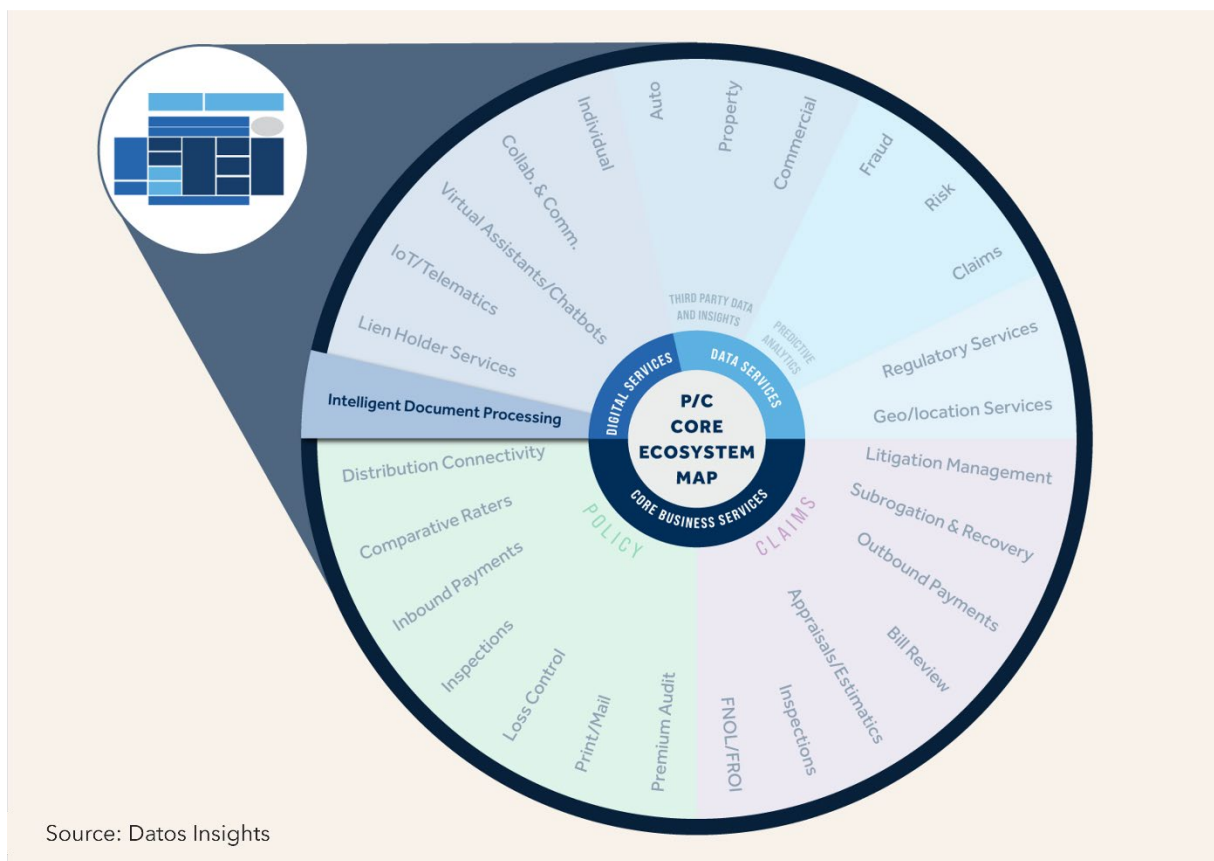
Source: Datos Insights

Ecosystem Map

Solutions in this space fall under Digital Services in Datos Insights' P/C Core Ecosystem Map, specifically the Intelligent Document Processing category (Figure 1).

Each section on the wheel represents a category of solutions with similar characteristics and objectives. Each section aligns with Datos Insights' core, digital, and data framework. The components within each section further define solutions and capabilities.

Figure 1: Datos Insights P/C Core Ecosystem Map

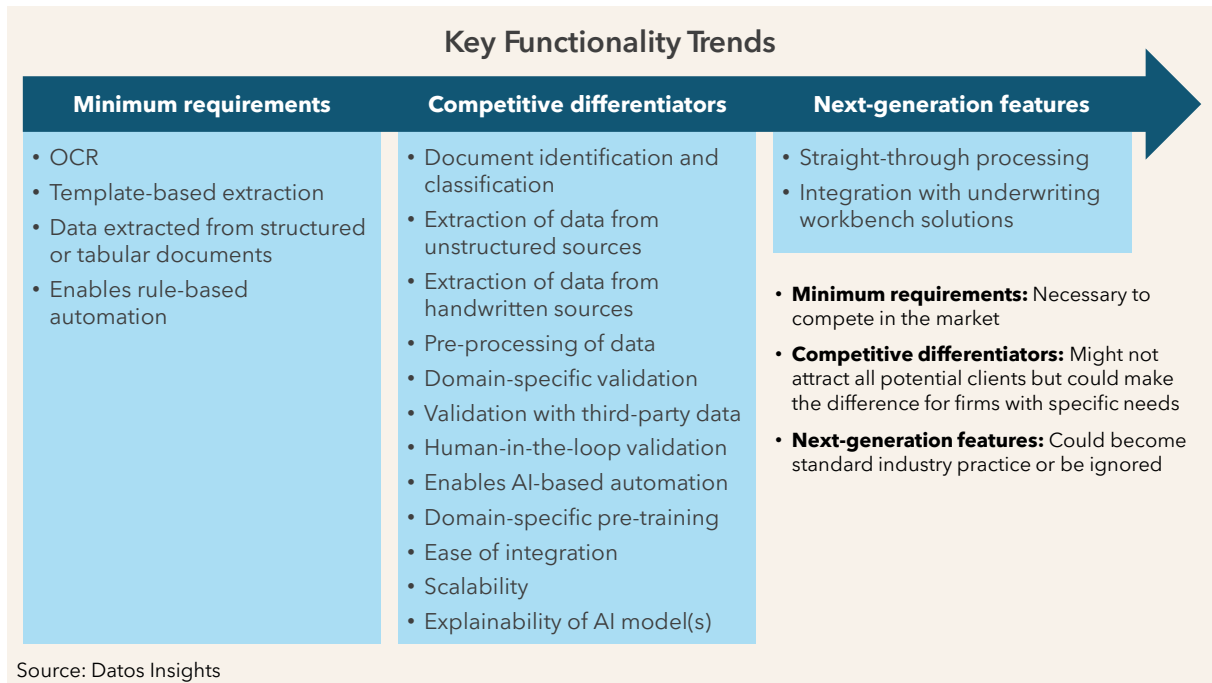


Source: Datos Insights

Functionality

Figure 2 shows the key functionality trends of IDP solutions.

Figure 2: Key Functionality Trends



Baseline IDP solution capabilities include the following:

- **OCR functionality:** This is the baseline ability to extract characters from images and store them in a digital format.
- **Template-based extraction:** Combined with OCR, this capability enables the classification of extracted data based on its location within a template.
- **Rule-based automation:** This capability includes triggering additional workflow based on the application of business rules to the extracted data.

Differentiating capabilities include the following:

- **Document identification and classification:** This capability includes automated determination of the document format and type to apply the appropriate logic.

- **Handwritten data sources:** This capability expands basic OCR functionality to more complex and varied text, such as natural human handwriting.
- **Pre-processing of data:** This capability includes de-skewing, denoising, cleaning, organizing, and transforming the raw data to match the quality required by the IDP solution's models.
- **Domain-specific validation:** This capability includes comparing extracted data to reasonable expectations of the data format and range based on an understanding of the document being processed.
- **Validation with third-party data:** This capability expands domain-specific validation to include accessing a third-party data source for validation (e.g., motor vehicle records).
- **Human-in-the-loop validation:** This capability involves referring low-confidence model results to a human for further assessment. In more sophisticated solutions, human-provided classifications/corrections are subsequently used to improve the model's accuracy.
- **AI-based automation:** This capability expands the basic rule-based automation to include more complex or variable conditions.
- **Explainability of AI models:** This capability includes the specific design of the solution such that the model's results can be audited to determine how a decision was made.

Next-generation capabilities may include the following:

- **Straight-through processing:** This capability includes the complete automation of the claims or underwriting process. This functionality is commonplace for applications or claims that are low-risk or easily processed. Next-generation straight-through processing functionality will expand to include the more complex cases currently handled by humans.
- **Integration with underwriting workbench solutions:** This capability includes integrating IDP solutions with advanced underwriting workbench solutions to improve turnaround times, increase underwriter productivity, and increase underwriting quality.

Solution Types

As with the evolution of any space, IDP solutions have started to specialize, filling distinct niches. That specialization has driven solutions into three broad categories—general purpose, insurance vertical, and fit for purpose—each of which has its own pros and cons:

- **General-purpose solutions:** Often those from big-tech companies such as Microsoft, AWS, and Google, these are the ones that are most broadly applicable. They are trained broadly to process documents of any sort from any industry. General-purpose solutions will be widely applicable across the enterprise. Still, they may take significantly longer to achieve value than other solution types.
- **Insurance-vertical solutions:** These are similar to general-purpose solutions but with additional training and/or design components specifically applicable to the insurance industry. These solutions are applicable across most functions of an insurer, from underwriting to claims, and require only moderate investment in training to achieve results in the short term.
- **Fit-for-purpose solutions:** These are targeted even more narrowly than insurance-vertical solutions (e.g., claims or underwriting for specific lines of business). Fit-for-purpose solutions will generally be able to achieve accuracy—and value—very quickly when applied to a purpose for which they were designed and trained; that accuracy and near-immediate return, however, can be fairly expensive. A subset of fit-for-purpose solutions do not come with pre-trained models. The solutions will require a substantial time investment to train them with the appropriate data and consequently take longer to reach their potentially high accuracy.

Vendor Metrics

No matter the type of solution pursued, all insurers should pay close attention to certain metrics when evaluating a potential solution provider:

- **Time to accuracy:** Vendors should be able to predict how long it will take for their solution to deliver highly accurate models for any potential use case and ideally show evidence of those timelines being achieved by other users. Time to value, a similar but distinct metric, will be closely related to time to accuracy but must include further consideration of the business value that the solution delivers.

- **Use cases and model data:** For all types of IDP solutions, the insurer and solution provider must understand the product lines and specific use cases being considered. There is some level of commonality across products and use cases, but the intricacies of each product and line are where it can become challenging to achieve—and maintain—high levels of accuracy. Hand-in-hand with understanding how the model will be applied, it is also critical that the data used by the model for training, retraining, and day-to-day use comes from a well-understood and well-characterized source. Undisclosed data sources are a cautionary sign that can be ignored at one’s own risk.
- **Pricing:** Pricing considerations, unsurprisingly, are also important. Specific pricing models range from platform fees and transaction pricing to broader submission fees. For every use case, carriers should model the costs and potential returns to ensure clients are getting actual value from the service into which they may be investing substantial time, energy, and resources.

Prominent Providers

The IDP market space, like most AI-enabled technologies, continues to change rapidly. The relatively low barrier to entry for new solution providers has pushed the space close to being fully commoditized, and that trend is unlikely to change. Nonetheless, several solution providers have been at the forefront of this change and are likely to remain in a prominent position.

Table B compares select IDP solution providers.

Table B: IDP Solution Providers

Vendor	Type	Common documents	Input structure
ABBYY	Insurance vertical	Claims, invoices, customer onboarding documents, purchase orders	Unstructured
AWS	General purpose	Claims, underwriting submissions, invoices, contracts, government-issued ID cards	Unstructured
Cinnamon AI	General purpose	Claims, invoices, receipts, financial statements	Unstructured
CogniSure	Fit for purpose	Loss runs	Unstructured
Google	General purpose	Claims, underwriting submissions, invoices, contracts, government-issued ID cards	Unstructured
Groundspeed	Fit for purpose	Commercial insurance submissions	Unstructured
Hyperscience	Insurance vertical	Claims, life insurance applications, medical exams	Unstructured
Infrd	Insurance vertical	Underwriting submissions, claims, invoices, receipts	Semi-structured
IntellectAI	Fit for purpose	Commercial underwriting submissions	Unstructured

Vendor	Type	Common documents	Input structure
Kofax	General purpose	Underwriting submissions, claims, applicant onboarding	Unstructured
mea	Fit for purpose	Underwriting submissions, claims	Unstructured
Microsoft	General purpose	Claims, underwriting submissions, invoices, contracts, government-issued ID cards	Unstructured
Pibit.ai	Fit for purpose	Commercial underwriting submissions	Unstructured
Send Smart Submission	Fit for purpose	Underwriting submissions	Unstructured

Source: Datos Insights

General-Purpose Solutions

The following represents an overview of select IDP solution providers with general-purpose solutions:

- AWS** offers Amazon Textract—an ML service that extracts text, handwriting, and other data from scanned documents—in combination with Amazon Comprehend, an NLP service that derives insights from text, including names, locations, dates, PII, and other entities. Additional human review of low-confidence data can be built into workflows via integration with Amazon Augmented AI (Amazon A2I). Publicly announced clients include Chisel AI and LexisNexis.
- Cinnamon AI** offers an AI Document Reader to automate data extraction from unstructured documents, including invoices, receipts, financial statements, and claim submissions. Cinnamon AI's clients include Dai-ichi Life Insurance Company, Nissay, and Tokio Marine.
- Google** offers Google Cloud Document AI, a product that combines OCR and NLP to extract information from structured and unstructured documents. Human review of ML predictions can be integrated via Google's Human-in-the-Loop AI feature. Custom models or uptraining of existing models can be used to achieve greater accuracy for data extraction from unstructured documents.

- **Kofax** offers TotalAgility, a solution for automating content-intensive workflows via the low-code design of IDP processes, augmented by Kofax RPA, where applicable. The solution can ingest structured, semi-structured, and unstructured documents, with or without tables. Additional capabilities of the solution include case management, rule management, and analytics. Publicly announced clients include Aviva and AIA Group.
- **Microsoft** offers AI Builder and Power Automate for building low-code/no-code end-to-end document processing workflows. Using as few as five sample documents, AI Builder can extract data from structured, semi-structured, and unstructured/free-form documents. Workflows can be extended to include human validation and export of data to storage systems. Microsoft's Form Recognizer tool provides a code-first solution to extract text, key/value pairs, and tables from documents. Publicly announced clients include Progressive Insurance, The Hanover Insurance Group, and Zurich.

Insurance-Vertical Solutions

The following represents an overview of select IDP solution providers with insurance-vertical solutions:

- **ABBYY** offers the ABBYY Vantage to enable straight-through processing by capturing, extracting, and processing data from structured, semi-structured, and unstructured documents. This includes invoices, claims, new account opening and customer onboarding documents, purchase orders, and contracts.
- **Hyperscience** offers a platform for the automation of claims processing, enrollment and underwriting, and account servicing. The Hyperscience Platform can extract and classify data from unstructured documents, including claims documents, life insurance applications, and medical exams, and then supplement that with third-party data. Publicly announced clients include Guardian, Mutual of Omaha, and Philadelphia Insurance Companies.
- **Infrd** offers a solution for augmented and automated processing of claims documents, including handwritten documents and receipts. Extracted data can then be used for straight-through processing. In cases with lower confidence, data can be directed to a human for review, otherwise freeing those roles for higher value-added tasks. A publicly announced client is State National.

Fit-for-Purpose Solutions

The following represents an overview of select IDP solution providers with fit-for-purpose solutions:

- **CogniSure** offers an AI platform that the company notes is purpose-built for the insurance industry. The company uses AI and ML to extract unstructured data from insurance documents and convert it to customer-specific output formats through API. The AI algorithms can extract data from multiple unstructured sources, primarily focusing on loss runs. Extracted data can be enriched with third-party data sources. Publicly announced clients include West Bend, Graham Company, Heffernan Group, and Kraus-Anderson.
- **Groundspeed** offers AI-driven extraction and structuring of data from commercial insurance submission documents. The company flags any low-confidence data for human review, enriches the data set using third-party data, and returns results via API, SFTP, or email. Publicly announced clients include AF Group, Travelers, and Berkshire Hathaway.
- **Intellect AI** offers Magic Submission for the intake and routing of commercial underwriting documents. The company enriches the data from submitted documents with third-party data to increase the speed to quote and bind, standardize underwriting decisions, and improve loss ratios.
- **Mea** offers platforms for submissions and claims ingestion. Mea's pre-trained AI designed specifically for insurance uses enables near-immediate application and rapid scaling. Publicly announced clients include AXIS Capital, Acquinex Limited, Mosaic Insurance, and Rising Edge Insurance.
- **Pibit.ai** offers a solution for commercial underwriting. The solution consists of ingestion, data extraction, and analysis of unstructured, semi-structured, and structured documents during the commercial underwriting process. Current clients include Tokio Marine, Sompo, and Mitsui Sumitomo.
- **Send** offers Smart Submission, a solution for automated submission intake, triage, and enrichment via third-party data. The solution allows for the customization of business rules to manage submission prioritization. It can automatically push submissions meeting an insurer's criteria to an underwriting workbench.

Conclusion

Insurers

- Evaluate fit-for-purpose solutions to accelerate time to value for well-defined use cases and opportunities.
- Consider horizontal solutions to solve a variety of use cases across the enterprise and minimize solution costs. Recognize that time to accuracy may take longer to achieve than fit-for-purpose solutions, and additional resource time will be required.
- Negotiate with solution providers to avoid per-transaction fees, which can quickly become expensive and diminish value proposition.
- Reconsider workflows in underwriting as part of an IDP solution. Integration with risk models and underwriter workbenches can improve underwriter insight into risk.

Vendors

- Continue to invest in AI capabilities to improve accuracy and document type flexibility.
- Consider integration with other third-party solutions, such as analytic models and core systems, to provide a seamless experience for the insurer.
- Invest in model training outside a carrier engagement to accelerate accuracy and time to value.
- Evaluate pricing models to ensure carriers can take maximum advantage of the technology. Pricing carriers out of utilization will diminish value and market penetration over time.

About Datos Insights

Datos Insights is an advisory firm providing mission-critical insights on technology, regulations, strategy, and operations to hundreds of banks, insurers, payments providers, and investment firms—as well as the technology and service providers that support them. Comprising former senior technology, strategy, and operations executives as well as experienced researchers and consultants, our experts provide actionable advice to our client base, leveraging deep insights developed via our extensive network of clients and other industry contacts.

Contact

Research, consulting, and events:

sales@datos-insights.com

Press inquiries:

pr@datos-insights.com

All other inquiries:

info@datos-insights.com

Global headquarters:

6 Liberty Square #2779

Boston, MA 02109

www.datos-insights.com

Author information

Deb Zawisza

dzawisza@datos-insights.com

Contributing author:

Jack Krantz

jkrantz@datos-insights.com

© 2023 Datos Insights or its affiliates. All rights reserved. This publication may not be reproduced or distributed in any form without Datos Insights' prior written permission. It consists of information collected by and the opinions of Datos Insights' research organization, which should not be construed as statements of fact. While we endeavor to provide the most accurate information, Datos Insights' recommendations are advisory only, and we disclaim all warranties as to the accuracy, completeness, adequacy, or fitness of such information. Datos Insights does not provide legal or investment advice, and its research should not be construed or used as such. Your access and use of this publication are further governed by Datos Insights' Terms of Use.