2024
Global Capture & IDP Software
Vendor Matrix Report Excerpt

By Ralph Gammon

Senior Analyst

Infosource Software

May 2024

CONTENTS

TABLES AND FIGURES

BACKGROUND AND SCOPE

Infosource has a long tradition of assessing and ranking key vendors in the Capture and Intelligent Document Processing (IDP) software market. As recently as 2016, we were ranking less than 10 significant vendors out of a global total of probably less than 100. It was also about that time ISVs started leveraging AI to develop IDP.

Infosource has historically defined Capture as software used to ingest information into business processes. To accomplish this, Capture SW acquires, classifies, and converts unstructured and semi-structured information into enhanced usable data. The captured information can be used for business transactions, analytics, records management, discovery, and compliance applications. Further, Capture SW understands and extracts meaningful, accurate, and usable information from multi-channel inputs.

The acronym IDP started to come into use with the widespread introduction of AI into the document processing market. While traditional Capture relies on OCR for character recognition, along with techniques such as drawing boxes to define fields and the use of rules and regular expressions to increase accuracy, AI enables more of a learn by example approach. We first saw AI used as a tool in “Advanced” or “Intelligent” Capture applications for auto-classification, as well as some extraction from semi-structured forms (document types that contain similar information in variable locations, so traditional templating doesn’t work).

Eventually, however, as AI became more accessible, in large part due to the increased processing power the cloud brings to bear, we saw new AI-first applications being built specifically for document classification and extraction. Some of these applications even utilize new OCR/ICR engines that were developed with AI. As these types of applications emerged, they were labeled as IDP.

Since then, many people have started to link the term “Capture” with paper scanning. And while the Capture SW market does have many roots in the conversion of paper to electronic formats, Infosource moved beyond covering just scanning input well before IDP was introduced. As an acknowledgement of the rise of the term “IDP” in the market, and its close relationship with what we have termed “Capture,” we have coupled the terms in what we now define as the Capture & IDP market.

This year’s Global Capture Matrix ranks more than 20 of the leading vendors in the Capture & IDP market. Slightly more than half have history that predates the emergence of IDP, but we also have a strong representation of newer vendors, and this includes some of our Stars.

Clusters of Capture & IDP Vendors

When looking at the landscape of the Capture & IDP market, there are multiple clusters we can group vendors into:

- Recognition Engines/Toolkits - offerings include OCR/ICR engines and cloud-based IDP services
- Document Processing Services - vendors with typically cloud-based AI models trained for data extraction of specific document types, such as invoices or receipts
- Capture and IDP Applications - can be run on-premises, in a private cloud, or in a SaaS environment, these manage document workflows, including input, classification, and data extraction and are typically integrated with third-party line of business systems
- Content Services/ECM - vendors with systems that manage content with Capture & IDP providing an efficient avenue for entering and cataloging that content
- Workflow/BPM/BPA - systems for managing processes that often require documents and/or the data contained in those documents
- RPA - processes automation of structured data that can be repeatedly transferred from one system to another; IDP enables input of unstructured data

- Line of Business applications - An ERP, procure-to-pay, expense management or vertical system focused on a single area of business which utilizes information from documents

While we currently include all Capture & IDP vendors in a single matrix, it would be possible to create a separate matrix for each of the listed clusters. It is not something we have had the bandwidth to do yet, but something we are considering for the future.

That said, there is definitely a lot of crossover and multiple vendors could be represented in multiple clusters. In the past, we have sometimes tried to put each vendor we cover into a single cluster based on their major area of focus, but as the market evolves, aligned with our vision toward solutions that better address end-to-end automation (See Figure 1: Infosource Market Maturity Progression), assigning a vendor to single cluster has become difficult. In the summaries, we have tried to list each cluster a vendor participates in.

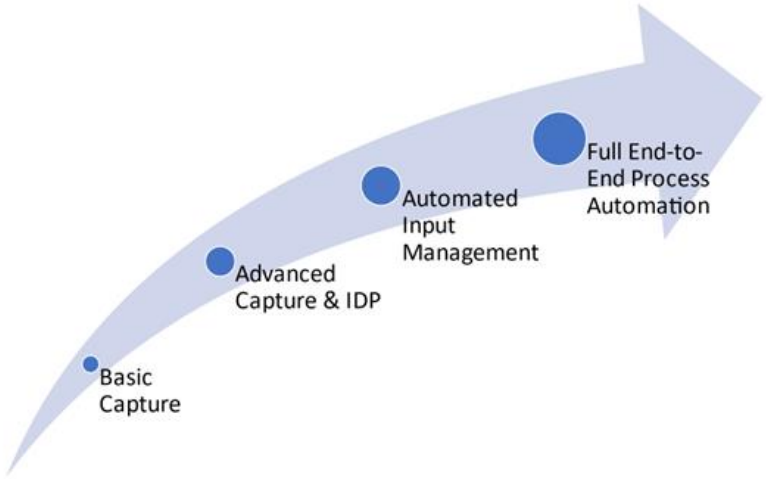

Figure 1: Infosource Market Maturity Progression

METHODOLOGY

Criteria for Ranking

Our assessment of vendors is based on multiple criteria that we align on two axes:

- **Strategy and Capabilities (Y-axis**): This reflects the vendor's vision and its technical capabilities to deliver on that vision.
- **Execution in the Market (X-axis):** This reflects the vendor's ability to deliver Capture & IDP solutions to the market.

For the y-Axis, we surveyed the vendors and researched their offerings and roadmaps and compared the input against our vision for Ideal Future Capture and IDP Solutions, which should automate the processing of all inputs required for business transactions, including unstructured and semi-structured inputs from any source, in any format. Ideal Capture & IDP solutions enable end-to-end automation of transactional processes and can also address compliance and analytics applications. Ideal solutions ingest omni-channel inputs, classify them and identify the related business processes. They then extract the relevant data, validate and augment it, and apply it to the appropriate systems and workflows, which may involve

initiating or advancing a process. Scalability and accuracy is important. AI and machine learning are increasingly utilized to reduce implementation time and effort. A containerized standardized cloud services architecture should be applied to increase flexibility in deployment of Capture & IDP as well as improve integration with cloud platforms.

For the X-axis, we also relied on surveys and research, as well as our Capture & IDP market share information. Due to their strong install bases, vendors with legacy Capture technology made up a large portion of our market share leaders for 2022 (we are currently compiling our 2023 numbers), and several are ranked highly in our 2024 matrix. Despite the emergence of hundreds of IDP vendors in recent years, many of the vendors with a background in traditional OCR-based Capture have been hard to unseat, especially in high-volume complex document processing environments. End users are often reluctant to replace mission critical software applications, and therefore will often look at IDP primarily for greenfield opportunities.

This is not saying that AI-first vendors are not having success in the market. The opportunity many of them saw for getting into the market in the first place was to expand the size of the traditional Capture market by addressing documents that OCR-based document processing wasn't effective on, as well as being able to reach more uses by reducing set up and configuration times. IDP vendors have been effective on both these fronts, enabling automated data capture from more complex and less structured documents, including document types like contracts, bank statements, and financial reports, as well as those containing handwritten information.

IDP technology has also created a lower barrier to entry into the market for vendors in other areas of automation like RPA (Robotic Process Automation) and BPM (Business Process Management). These types of vendors represent some of the most successful IDP players, as they have been able to sell IDP as an extension of an overall enterprise automation strategy.

Weighting Scale

Following is a list of the criteria considered under each category along with the weighting applied to establish the position on the axis (L=Low, M=Medium, H=High):

Strategy and Capabilities (Y-axis)

- **Vision (M):** This is what the vendor aims to deliver related to Capture & IDP today, as well as in the future. We are looking for something that aligns with our vision for Ideal Capture & IDP solutions.
- **Cloud (M):** Traditionally Capture & IDP has been primarily delivered on-premises, but cloud-based delivery has consistently increased for the past 10 years and, with impetus from the pandemic, hosted Capture & IDP services doubled as a percentage of the market from 2019 to 2022. That said, on-prem and private cloud implementations are still common, so we consider all a vendor's delivery options and their ability to address the needs of their target market when ranking them.
- **Technology Stack, including AI and Security (H):** This is where the rubber hits the road, in terms of the ability to deliver a product that meets the vision. Vendors are rewarded for both a deep technology stack, with expertise in specific areas (such as tuning AI models for document processing), and breadth, meaning they have multiple elements they can bring together in a solution. Having your own technology is good, but increasingly partnerships, especially with AI experts and hyperscalers, can be equally effective.
- **Multi-Channel Capture Capabilities (L):** Many Capture & IDP use cases require input from multiple sources. In addition to documents, which can be captured through scanners, MFPs, faxes, and mobile apps, as well as submitted electronically through channels like e-mail, relevant

information can come in from sources like voice, video, SMS, and more. We are looking at a vendor's ability to process all this unstructured input in a business workflow.

- **Low-code/No code (M):** This applies not only to the ability to configure and adjust Capture & IDP workflows, but also to the ability to integrate to third-party systems with minimal coding. Using less code is clearly a trend in the market, but the ability to utilize some coding to enable customization can be considered a plus as well.
- **End-to-End Automation (M):** As mentioned earlier, this is where the market is ultimately headed – how well can a Capture & IDP solution enable completion of entire business processes and what is the range of processes it can be used to automate? There are multiple factors to consider, including the breadth of a vendor's complementary technology stack and their partnerships. Elements like vertical and horizontal depth and focus are also considered.

Execution in the Market (X-axis)

- **Competitive Advantages (M):** We are looking at what a vendor does well or differently that motivates end users to choose its solution and how it stands out from the competition.
- **Customer Base (H):** Considers factors like the number of customers, the size and complexity of the vendor's implementations, and their regional, vertical and use case distribution.
- **Partner Strategy (M):** Looking at how well a vendor plays with others. This includes consideration of its channel partner as well as technology partners and how they help the vendor bring its product successfully to market.
- **Ease of Use (M):** Conders how much configuration and professional services are associated with an implementation and how fast can a user or a partner get a solution up and running.
- **Demand Generation (M):** Looks at what a vendor is doing to drive interest and ultimately sales of its product. This includes considering the channels and techniques the vendor is using and their influence in the market.

Influence of Generative AI

AI is now table stakes in the Capture & IDP market. Every vendor we have assessed is using it in some form. Custom trained models have been utilized for several years, but within the last two years, we've also seen the emergence of Generative AI. Gen AI is based on Large Language Models (LLMs), which are trained on billions (even 10s of billions) of parameters. As a result, they are very powerful, and in addition to being able to compose answers to questions and create images and videos from natural language prompts, they can do document extraction and summarization, without any additional training.

LLMs or Generative AI platforms should not be considered complete Capture or IDP applications. They do not have production UIs, any workflow functionality, validation or QA capabilities (including confidence level readings) and their security varies depending on how they are implemented. But, that doesn't mean they can't be useful tools. So far, we've seen Generative AI/LLMs used for functionality like document summarization, on demand interrogation, reducing set up times by finding matching pairs, recommending workflow designs and automating QA and validation steps. This has all been done leveraging commercially available LLMs.

Part of the excitement around LLMs is that they are advancing so rapidly. After they were introduced, they immediately became part of the public consciousness and billions of dollars in capital investments followed. As a result, they are becoming faster and more powerful. In addition, we are starting to see some customizations that could be especially relevant to Capture & IDP:

- **The ability to run LLMs on prem:** This is important for highly regulated, risk averse industries like financial services, insurance and government, which happen to also be three of the top verticals for Capture & IDP SW.
- **Vertically-trained LLMs**: These should improve summarization capabilities and data capture rates.
- **Ability to train LLMs on your own documents:** This will improve accuracy even more. This capability is especially important for production Capture & IDP environments in which businesses are often receiving high volumes of the same document types.

LLMs and Generative AI are just starting to emerge as factors in our industry and there are several vendors in our matrix who have yet to apply them to their Capture & IDP offerings. These vendors are instead focused on improving their dedicated IDP AI models. But, there will come a point when the economies of scale related to LLMs will be hard to ignore. Currently, cost is still a factor when considering utilizing LLMs for Capture & IDP, but Gen AI's widespread use across multiple applications is going to drive costs down. And the technology will improve faster than any dedicated models will be able to improve.

We're not saying LLMs are going to dominate Capture & IDP SW market going forward, but they are certainly going to be an important factor to consider. There is still a need for complementary technologies and techniques utilized in coordination with LLMs to make them a functional piece of a successful Capture & IDP software solution. And in this year's rankings, we'll say their influence was limited. But going forward, we expect the ability to successfully leverage Gen AI and LLMs for Capture, IDP, and related applications is going to help define the market and empower the most successful vendors.

2024 INFOSOURCE GLOBAL CAPTURE & IDP SW VENDOR MATRIX

Based on their scores in each category, the ranked vendors earned a position on each axis, which translates to a position within one of four categories:

- **Star** - proven market leader; strong in product strategy and capabilities, as well as execution
- **Disruptor** - strong strategy and product capabilities; need to optimize sales and marketing execution to become a Star
- **Competitor -** strong sales and marketing but limited vision and strategy required for future market scenarios; can enhance their Capture & IDP vision and strategy to reach Star level
- **Explorer -** early in Capture & IDP journey but showing good potential

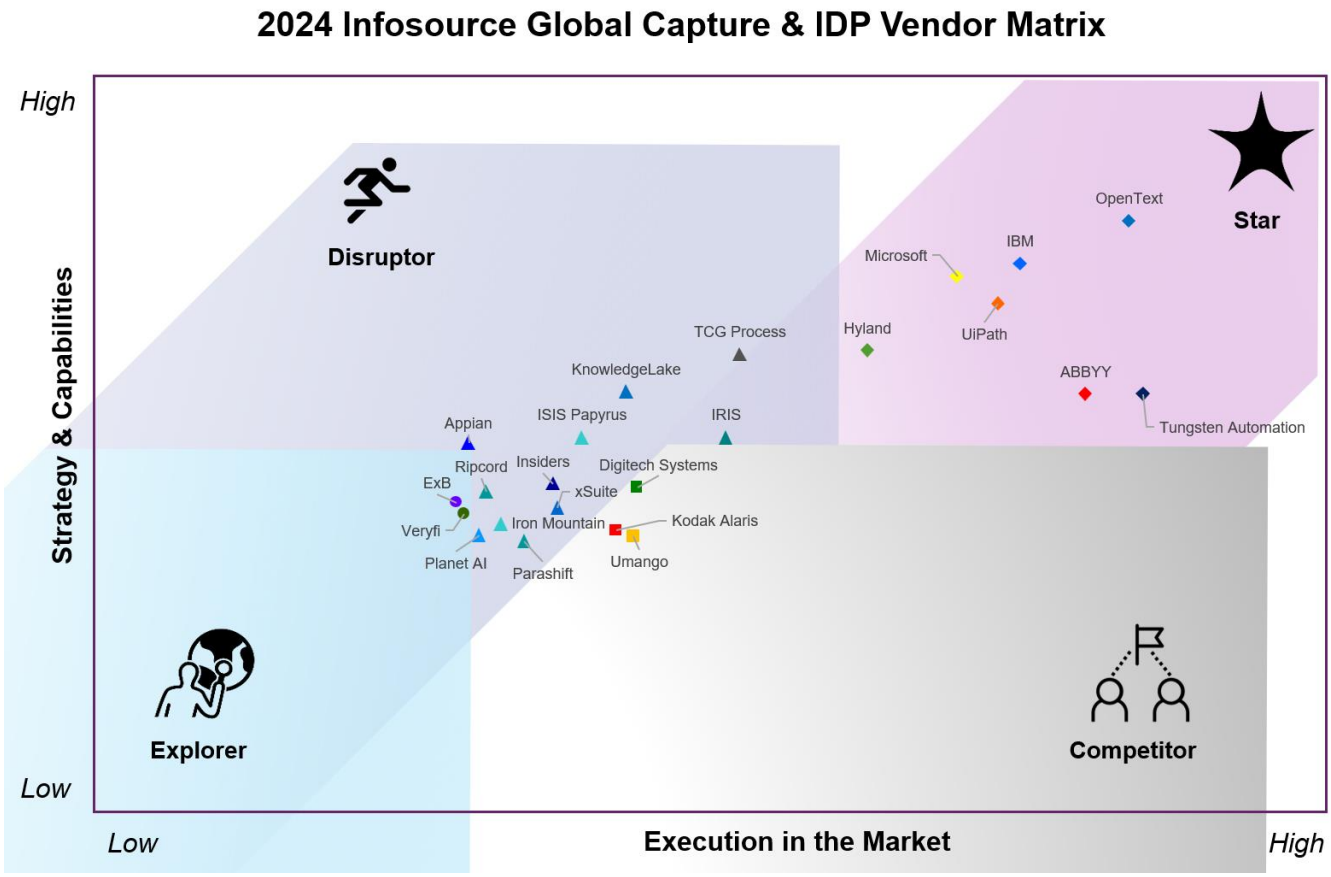

Figure 2: 2024 Capture & IDP SW Vendor Matrix

Matrix assessment

This year we have seven vendors in the Star category, up from three the previous year. Following is a brief explanation of how they were able to advance.

1. **Hyland:** Hyland has been one of our market share leaders with multiple Capture offerings and a strong end-to-end technology stack. Its introduction of a new IDP offer, based on proven technology, has pushed it into the Star category.
2. **Microsoft:** Microsoft has made progress by introducing IDP into its Office 365 ecosystem. Also, its handwriting recognition continues to get strong reviews.
3. **Tungsten Automation (formerly Kofax):** Tungsten continues to be strong in the Execution phase as they are our 2022 market share leader. The ISV has improved in its Strategy and Capabilities rating by expanding its focus to better address end-to-end automation opportunities and announcing a broad-reaching AI strategy.
4. **UiPath:** UiPath continues to improve its product, adding capabilities including innovative use of Generative AI. It has also rapidly gained market share by selling into its global base of RPA customers.

This year's Global matrix contains a handful of new vendors, including a couple featured only in our EMEA matrix last year:

- **Digitech:** A long-time Capture & IDP ISV that is especially strong in the service bureau market
- **ExB** (EMEA only last year): Leveraging expertise in NLP and AI to create a self-service cloud-based IDP service
- **Iron Mountain:** Primarily known for analog records storage, Iron Mountain has brought to market an open IDP platform that leverages third-party recognition services.
- **Kodak Alaris:** High-volume production scanner market leader has had Capture SW for a long time, but recently fleshed out the capabilities of its Info Input IDP Solution
- **Parashift** (EMEA only last year): A SaaS IDP vendor with no-code classification and extraction for a wide range of document types
- **Umango:** Australian based vendor that focuses on Capture & IDP enabling the MFP market
- **Veryfi:** an OEM vendor with a cloud service and proprietary mobile capture technology

Dropping from our coverage this year are Hyperscience, Rossum, UST, Automation Anywhere, BIS and Parascript. These vendors all continue to compete in the market, but we didn't have enough visibility into their current product offerings to rank them properly.

OPENTEXT ★

Global HQ: Waterloo, ON
Capture & IDP Products: OpenText™ Intelligent Capture, OpenText™ Core Capture, OpenText™ Magellan™, OpenText™ AppWorks™
Clusters: Capture and IDP Applications, Content Services/ECM, Workflow/BPM/BPA, Recognition Engines/Toolkits
(https://www.opentext.com/products/intelligent-capture)
(https://www.opentext.com/products/core-capture)
(https://www.opentext.com/solutions/intelligent-document-processing)

Overview
OpenText™ is a long-time leader in the market which in recent years has complemented its deep stack of Capture technologies with AI from both internal and third-party sources. The multi-billion dollar ISV has an extensive portfolio of content services and information management technologies that help it deliver end-to-end solutions to its customers incorporating Capture & IDP. It also has partnerships with enterprise software market leaders SAP and Salesforce, which have OpenText Capture & IDP integrated into their products.

Strategy and Capabilities

Vision

OpenText's vision includes using intelligent capture, AI, and its content services stack to help customers fulfill their requirements for end-to-end solutions. "Our vision is to be the single vendor where organizations can get proven IDP solutions that seamlessly integrate information capture, AI and machine learning, and process automation – to extract actionable data and accelerate downstream workflows in line-of-business applications and process automation tools."

Specific to Capture, OpenText's vision is to provide omni-channel input of structured, semi-structured and unstructured content, with built-in and optional AI technologies and integrations with BPA and RPA technologies.

Cloud

OpenText has a large install base of high-volume customers that run its software in-house, but the ISV is transitioning toward an increasing number of private cloud and SaaS implementations. OpenText Intelligent Capture (which has its roots in the Captiva line) can be deployed in the OpenText Cloud Platform (as a private instance) or in an environment hosted by a hyperscaler.

In heavily regulated industries with complex document management and workflow requirements, such as financial services, healthcare and public sector, private cloud is the more common route. For organizations whose needs are met by SaaS (public cloud, multi-tenant), OpenText Core Capture, a cloud-native Capture platform, is available either as a SaaS app or as cloud services (RESTful APIs). "The trend with our strategic partners, such as SAP and Salesforce, is to lead with cloud solutions. As a result, we also offer two partner-specific cloud offerings: OpenText Core Capture for SAP Solutions and Open Text Core Capture for Salesforce."

Technology Stack, including AI

OpenText has an extensive technology stack, including its own OCR, market leading scanner driver technology, the AppWorks low-code integration platform and leading content services platforms. OpenText continues to introduce new technology, including in 2023 “smart classification” for its Information Extraction Engine (IEE) which incorporates multiple technologies that utilize continuous machine-learning. OpenText incorporates multiple flavors of AI and machine learning including its own internally developed Magellan™ and IEE technologies as well as third-party LLMs.

“Instead of relying upon a single machine learning technology to apply in all situations, we provide multiple technologies so that the best fit for a specific situation can be employed to maximize accuracy when recognizing documents and accelerating business processes. We incorporate a number of AI technologies, including NLP and neural network technology, that are trained by us and embedded into our products. Examples of capture functionality driven by additional AI technology include OCR, image enhancement, document classification, sentiment and tone analysis, content summarization and named entity recognition.”

On OpenText’s Capture roadmap is the use of LLM technology to augment functionality like document parsing and object recognition. However, data extraction and continuous machine learning functionality will continue to be internally developed. “This will enable us to achieve higher accuracy than using LLM technology alone and also provide the greatest degree of data privacy for our customers.”

In addition, OpenText takes the position that GenAI can benefit from high-quality Capture. Along these lines, in 2023, it announced OpenText™ Content Aviator with generative AI and LLM technology for conversational search, content discovery, summarization, and translation within their content services platforms. “Content Aviator boosts productivity and revolutionizes how people extract insights and interact with content – and does this even better with access to the accurate metadata and text-searchable files that are provided by OpenText information capture and data extraction products.”

Multi-Channel Capture Capabilities

OpenText has capabilities for processing a wide-variety of input types, primarily document-based. These include scanned paper, faxes, electronically submitted PDFs, emails (body and attachments), e-invoices, mobile input, social media, images, SAP IDoc and XML/EDI.

Low-code/No-code Capabilities

OpenText CaptureFlow Designer enables users to set up workflows through a drag-and-drop UI that minimizes scripting and coding for features like auto-population and validations of fields.
In addition, OpenText Core Capture services are integrated into the Salesforce Automation framework, which enables administrators to automate processes through Flows using clicks, not code.

OpenText AppWorks™ is a process automation platform which employs a low-code/no-code approach for integrating capture capabilities with the rest of a customer’s information management ecosystem and their existing line of business systems. “It enables technical business developers to rapidly build and iterate on applications while providing the infrastructure and governance that professional developers need to extend the applications with custom code. End users are also able to define their own processes where appropriate using a ‘no-code’ case/process design experience. Developers have access to a wide range

of ‘building blocks’ which are added to an application to enrich it with capabilities such as data models, discussions, business history tracking, integrated outbound/inbound email correspondence, UI design, business rules, content management, RPA and reporting.”

Also, as part of its Aviator strategy for introducing AI to its product set, OpenText plans to introduce an Aviator for developers into AppWorks in 2024. This will accelerate development of applications by automating the generation of data models, workflows and other elements through natural language conversations. There will also be an Aviator for end users to streamline the creation of workflows.

End-to-End Automation Capabilities

OpenText has a comprehensive end-to-end automation strategy including Capture and IDP, a market leading portfolio of content services platforms, RPA technology, the AppWorks low-code/no-code integration platform, and integration with enterprise applications from Salesforce and SAP, as well as third-party content repositories. OpenText seeks to help organizations “capture, connect and control content and information – with the goal of improving employee experience and productivity, optimizing operational excellence, and protecting information.”

OpenText also has a Process Intelligence component within AppWorks, which supports advanced analytics on running processes. One goal is to introduce predictive AI features such as suggesting the next best task in a process, recommending process optimizations or delivering domain specific insights, such as the assessment of risk for insurance policies.

Sales and Marketing Execution

Competitive Advantages

OpenText has a wide breadth of traditional Capture and Content Services technology complemented increasingly by AI-based technologies. For its FY 2023 (ended June 30, 2023), OpenText had total revenues of $4.5 billion. The company has a global base of more than 120,000 customers and hundreds of reseller and systems integration partners. “When we present our IDP solutions, we talk specifically about combining the benefits of information capture, AI and machine learning, and process automation to extract useful data from all content and to accelerate downstream business processes. When competing against vendors that focus exclusively on IDP, having a broad portfolio of products and services is a winning competitive advantage for OpenText. Organizations often prioritize buying from a single vendor that offers all the pieces of Content Services that they need now or may need in the future.”

OpenText touts advantages like proven scalability and performance, deep integrations, innovation in Capture, flexible deployment options, and its broad portfolio of products and solutions.

Customer Base, including Case Studies

For several years, OpenText has ranked highly in Infosource’s global market share numbers. The ISV has a good number of high-volume implementations, several that are processing hundreds of millions of pages per year. It has customers spread across multiple regions and verticals. Case studies are diverse and typically feature more than standalone Capture, detailing integration with OpenText’s end-to-end automation stack.

Partner Strategy

OpenText sells through a combination of direct sales and partners and has multiple blue chip technology partners. OpenText utilizes a direct sales team for large global customers and addresses emerging geographies and the SMB market through partners. “We are committed to supporting customers and sales prospects with solutions that offer public cloud, private cloud, hybrid, and on premises options; compatibility with leading cloud hyperscalers (Google Cloud, AWS, and Microsoft® Azure); and deep integrations with strategic partners like SAP and Salesforce.”

Hundreds of organizations in OpenText’s Global Partner Program offer Capture and IDP. This includes resellers, support and services partners, distributors and others. This also includes SAP, which, in 2023 cited OpenText as its top technology partner. SAP offers a multitude of OpenText content services products through a resale agreement, including Vendor Invoice Management and Core Capture for SAP Solutions.

Ease of Use

OpenText’s software-to-services ratio can vary depending on requirements and deployment strategy (on-prem, cloud, hybrid, etc.). Basic Capture and IDP solutions using OpenText™ Core Capture or OpenText™ Core Capture for SAP® Solutions can be up and running in a matter of minutes. OpenText offers pre-packaged Capture & IDP solutions for both SAP and Salesforce customers and also as part of focused content services solutions in the Life Sciences, Financial Services and Government verticals.

Demand Generation

OpenText leverages corporate marketing, global campaigns, regional programs and partner marketing teams to generate demand for OpenText Capture and IDP solutions. These teams use approaches such as in-person events, digital programs and webinars, SEM, SEO and email communications. Currently, OpenText has a campaign around SaaS, that includes Core Capture, and it is also promoting its Salesforce partnership.

“A key differentiator is that OpenText has a broad customer base for Information Management, with the reach to deliver market-leading capture and IDP solutions to that customer base.”

Summary

OpenText continues to grow its Capture & IDP business at a steady pace. The ISV understood early on the value of advanced capture as part of an end-to-end solution and continues to build on that through internal development and strategic partnerships and acquisitions. Its large install base is fertile ground for increasing Capture & IDP use.

We expect OpenText to continue to invest in the Capture & IDP market and to leverage more AI to increase both the power and breadth of its solutions. We expect the ISV to look for additional enterprise software partnerships. Acquisitions will continue to be part of OpenText’s strategy for both increasing its market share, as well as adding complementary technology.

ABOUT OUR ANALYSTS

Ralph Gammon is a Senior Analyst for Infosource Software. In this role he covers market sizing, trends, opportunities and leaders in the Capture & Intelligent Document Processing software space. Ralph joined Infosource in 2017 as Regional Manager for the Americas. From 1998-2020, he served as editor and publisher of *the Document Imaging Report*, the premier insider newsletter covering the ECM/Content Services space.

Petra Beck is a Senior Analyst in the Infosource Software division, where she is responsible for analyzing and forecasting the Intelligent Capture Software and Robotic Process Automation markets. Petra has over 25 years of experience in the Information Management market. Prior to joining Infosource Mrs. Beck held various global positions with Kodak Alaris leading Business Research, Strategic Planning as well as thought leadership functions. Petra has gathered experience working with several multi-national companies in Market Intelligence assignments.

COPYRIGHT

Infosource (USA) Inc.

1013 Centre Rd. - Suite 403A

Wilmington, DE 19805

+1 (302) 467-2255