

Intelligent Data Capture: A Trend Only Beginning

Part 2

IMERGE Consulting, Inc

Robert Smallwood
September 2006
Robert.smallwood@imergeconsult.com
As Appeared in KM World
www.imergeconsult.com

In the last issue, we covered key trends and some leading firms in the Intelligent Data Capture (IDC) marketplace, which is also referred to as Intelligent Document Recognition (IDR). Both terms refer to the ability to scan documents or electronic pages that have no fixed layout and extract data from specific fields to populate a database or business system. Using IDC/IDR, documents may be unstructured, with varying layouts, or "semi-structured" with some fixed fields but mostly varying formats. Examples of unstructured forms are contracts, resumes and letters; semi-structured documents may be Explanation of Benefits (EOB) forms, insurance claim documents, invoices and the like.

The marketplace is active and dynamic, being characterized by mergers, acquisitions, re-positioning and new entrants. We have covered ABBYY, Brainware, Document Strategies Inc. and their Xerox alliance, EMC Captiva, DICOM's Kofax unit and ReadSoft. Here is a review of a few more leading players in this developing market:

AnyDoc Software

AnyDoc Software (anydocsoftware.com) of Tampa, Florida made its name with its forms processing product, OCR for Forms which gained a reputation in the industry as being a durable, cost-effective, straight-forward product for lower volume demands. Now called OCR for AnyDoc, the company includes the ability to process structured and unstructured forms within the product by including a rules engine where rules are created to assist the system in making determinations of valid data when it is performing the recognition process (e.g. valid date ranges, two-way purchase order-invoice matching). Similar to ReadSoft's product naming convention, the company offers various flavors of the product targeted at vertical or departmental applications like AnyDoc INVOICE, AnyDocEOB (for explanation of benefits applications), AnyDocCLAIM and AnyDocCLASSIFY (for mailroom applications).

AnyApp technology was introduced 4-5 years ago to provide a better way to capture and organize the common data that appears on semi-structured documents. AnyApp has also helped the firm with capturing and exporting data to business systems like SAP. QuickApp technology is used more for variable format/unstructured documents. When processing invoices data such as vendor ID numbers or total amount due are found in different locations in each invoice format, depending on which vendor provided the invoice. Using QuickApp, AnyDocINVOICE facilitates invoice processing. Data fields needed such as "Amount Due," or "PO Number" are defined and the invoices are scanned to find any variation or combination of keywords to locate the data needed. The next time an invoice from that vendor is processed, the software will "remember" where it found the data making data capture even more efficient. Hughes Supply (recently purchased by Home Depot) processes approximately 250,000 invoices per month using AnyDocINVOICE.

AnyDoc reports approximately 70 users of its unstructured form capability, especially focused in the manufacturing & distribution and utility sectors. Over 90% of its sales go through its reseller channel. What's AnyDoc's advantage? "The ability to read data accurately in a complex environment, like lengthy, variable format invoices," says Sam Schrage, VP Operations, Marketing & International Sales.

Datacap, Inc

Datacap, Inc. (datacap.com), based in Tarrytown, NY has been providing document capture and forms processing software solutions to organizations worldwide since 1988. Datacap claims several industry “firsts”:

- The first forms processing solution for the PC in 1989
- The first web-based capture solution in 2000
- The first forms processing components for FileNet Capture in 2003

Datacap introduced a rules processing capability in their Taskmaster product several years ago, giving it the capability to process not only structured forms but also unstructured or “highly variable” documents. “We don’t distinguish between structured and variable format documents – you get the full capability for either when you buy our Taskmaster product,” states Scott Blau, CEO of Datacap. “In fact, some of our customers start with forms processing and then have moved to processing variable documents, like Blue Cross Blue Shield of Arizona or Sharp Healthcare. The advantage is in lower training, implementation and support costs, since it’s one product.”

Taskmaster Web Service won a 2006 AIIM Best of Show Award. Taskmaster Web Service is a set of capture capabilities, including image processing, recognition, validation, and export formatting that are hosted in a Web Service, simplifying integration.

Written in Microsoft VB Script, the rules processing capability also facilitates data export formatting to populate databases like Oracle or SAP. Datacap is certified with SAP’s web-based NetWeaver product, which unifies integration technologies into a single platform and is pre-integrated with business applications, reducing the need for custom integration. The platform is based on industry standards and can be extended with commonly used development tools such as Java 2 Platform, Enterprise Edition (J2EE); Microsoft .NET; and IBM WebSphere.

Datacap reports 50-60 customers using the IDC/IDR capability, out of a customer base of approximately 350. The firm is particularly active in the public and healthcare market sectors and has about 35 resellers.

Esker

Esker (www.esker.com) entered the information technology market in 1985 as a software consulting company in Lyon, France. Over the next 15+ years, Esker has grown into a worldwide provider of fax and document capture software, and host access solutions. Its U.S. offices are based in Madison, WI.

Esker DeliveryWare is a solution that automates the exchange of critical business documents between customers, business partners, and suppliers regardless of source, format, and destination. According to Renee Thomas, Director of Field marketing, “It can route faxes, email, XML documents, print streams, wireless messages and web content to automate business processes end-to-end. As an example, we’re NetWeaver certified to format and input data into SAP R/3, regardless of its origin.” Esker works with partner Merkur for outbound communications from the Oracle Business Suite.

Marking the dynamism of the IDC/IDR market, Esker built meta rules and introduced its dynamic document capture product in March of this year. Esker is working with Whirlpool to take sales orders from faxes to automate the order entry process.

Top Image Systems

Top Image Systems Ltd. (www.topimagesystems.com) was founded in 1991 and is a public company traded on NASDAQ (symbol: TISA). The company is strongest in Europe and operates internationally though its R&D is headquartered in Tel Aviv, Israel. TIS partners include SAP, IBM, Accenture, Atos Origin, Xerox, CACI, Kodak, Unisys and Fujitsu. Some of its larger customers include Armstrong Flooring, FedEx (processing waybills) and Gillette. Top Image Systems just won recently won a contract with Hormel Foods in Minnesota.

Top’s eFLOW Unified Content Platform, provides a single end-to-end solution to capture, classify, process, validate and deliver business-critical key data lying within incoming documents. eFlow uses Artificial Intelligence (AI),

capabilities and is designed to work with various ERP, finance and purchasing systems, allowing organizations to retain and exploit the inherent value in their legacy systems. eFlow integrates with SAP, Oracle, JD Edwards and 170 other business systems. eFLOW customers span a broad range of industries.

“What’s different about us? We developed IDR capabilities in our product from the beginning; and all processing of structured and unstructured forms is accomplished in one platform,” according to Gideon Shmuel, VP of Sales and Marketing. “And we use a minimum of four recognition engines on a field level to provide a high recognition rate and eliminate false positives.”

Other entrants in to the IDR/IDC market include Bantec (www.banctec.com), Saperion (www.saperion.com) and Scantron (www.scantron.com) and there are more – and many to come.

Although automating the accounts payable process was the first application ever to use electronic document imaging, (when Andersen Consulting implemented a custom system at McDonnell Douglas in St. Louis almost 20 years ago) the IDC/IDR market is likely less than 10% penetrated - which leaves a lot of room for growth. New IDC/IDR capabilities further improve efficiencies by cutting out nearly all manual data entry and intelligently formatting data.

The winners in the IDC/IDR race will be those that have the best relationships with business system vendors like SAP and Oracle, have stable, efficient installations and aggressive marketing efforts. Only time will tell who those winners are.