Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

4003 Wood Street ● Erie, PA 16509 ● PH (814) 866-2247 ● FX (412) 291-1352 ● http://www.documentimagingreport.com

August 17, 2007

THIS JUST IN!

THIRD ANNUAL HSA CONFERENCE SEPT. 5-6

Harvey Spencer Associates' third annual Document Capture conference is scheduled for Sept. 5-6 at the Glen Cove Mansion on Long Island. This year's event will cover a wide range of topics aimed at vendors of document capture hardware and software. These include XML, taxonomy, embedded software, voice recognition, the emerging Indian market, as well as Spencer's annual state of the industry address.

"Last year, we talked more about technology," noted Spencer. "This year, we are focused on solutions. For example, while last year we had a speaker discuss XML, this year, we have a speaker focusing on XBRL, a business reporting language being promoted by the **SEC** to all public companies. This is a very pertinent topic for vendors selling into the accounts payable and receivable spaces. We are also looking at ways vendors can leverage technology like taxonomy and voice recognition within their capture applications."

Like last year, the event kicks off with a networking dinner, set for Wednesday evening, Sept. 5. A full day of speakers has been slated for Thursday, Sept. 6, followed by an evening barbeque. More than 60 people attended Document Capture 2006 [*see DIR 10/20/06*].

For more information: http://www.harveyspencer.com/documentcapture/index.html; erin.dempsey@hsassocs.com.

OneTouch Drives Strong Quarter

Visioneer's document scanning solutions strategy is starting to pay off. The company recently reported a record second quarter, highlighted by 30% growth over the second quarter of 2006. Driving this growth were a pair

CONTINUED ON PAGE 8

The State Of Invoice Processing

Hottest growing market segment has come a long way in past three years.

Welcome to *DIR's* first issue dedicated primarily to one market, or sub-market, if you will. Yes, this issue tackles invoice processing, which has emerged as one of the fastest growing segments in document imaging. Leading the way is IDR (intelligent document recognition) technology used to automatically extract relevant data from invoices and drop it into accounting/ERP systems. It seems like every week, we hear of at least two or three significant new deals involving invoices and IDR.

Just a few years back, I struggled to round-up a sufficient number of end users to fill out a story on invoice processing in *Transform* magazine. Installations were so scarce that every time I would talk to vendors like **AnyDoc**, who was one of the North American pioneers in this space, I'd ask how many installations they had. Well, as you can see from the chart we've compiled on pages 4-5, there are now plenty of automated invoice capture installations to go around.

Even if you take these vendor-supplied numbers with a grain of salt, it looks like there are more than 500 invoice capture installations on U.S. soil. I'll estimate this is 10 times the amount there were just three years ago. That gives us a compound annual growth rate well over 100% for 2005-2007, which explains why batch capture vendors like **Kofax** are so serious about transitioning their focus to IDR [*see DIR 7/20/07*].

While not quite as easy as shooting fish in a barrel, finding end users who've applied IDR to tackle their invoice capture problems is much easier than it once was. The challenge now is finding end users doing line-item matching. This involves locating and reading individually listed items on invoices for matching against purchase orders. And, while almost every vendor will tell you they offer line-item capture, finding customers to talk about it is still challenging.

We did find one user willing to share its line-item story. However, even though their vendor talks a good game

about getting high recognition rates for line items right out of the box, this customer apparently had to do some serious tweaking to get its system from something like a 17% straight-through processing rate (meaning a person does not have to look at a document image before the data is passed on to the accounting system) to its current level of 84% for invoices that have well-formed P.O.'s. Granted, this was an older generation of the vendor's software, and the end user is very happy with the application, having achieved a greater ROI than it had originally envisioned, but, the fact remains, that while capturing header data (vendor name, P.O. number, date, remit to address, etc.) has evolved into more of an exact science than it was a few years back, capturing line-item details still seems to be a bit of a black art.

A look at line items

Because line-item capture is such a hot topic and often looked at as a differentiator among invoice processing products, we'll open the Q&A portion of our issue by discussing it. Following are responses we received from vendors that we view as the leaders in North American invoice processing technology, when we asked them: **Do you offer automated line item capture? If so, what percentage of your customers utilize this feature?**

EMC Captiva: EMC Captiva's *InputAccel for Invoices* leverages *Dispatcher's* [*This is the SWT software, which Captiva acquired in 2005, see DIR 5/20/05. Previously, Captiva had an OEM license for* ***Océ Document Technologies'*** *(ODT) DOKuStar invoice processing toolkit.*] free-form capability to provide organizations with the ability to capture line-item details and complex tables for both recurring and non-recurring vendor invoices. All implementations of *InputAccel for Invoices* include line-item capture capabilities; however, in certain instances (such as insurance for example) this particular feature may not be a requirement and is left un-utilized. EMC Captiva does not track who does or does not utilize this function, but it would be safe to assume that more than 70% of implementations are using line-item capture.

Datacap: To date, about 75% of our users are capturing line items. Some of them, like **Kuehne + Nagel** [*see DIR 8/3/07*] and **ista N.A.**, bought us because of line-item capture.

AnyDoc: Yes, one of *AnyDoc®INVOICE's* greatest strengths is its ability to capture detail lines. Approximately 70% of our customers take advantage of this feature.

ReadSoft: Yes, we have advanced line-item capture, which 42% of our customers use.

Top Image Systems (TIS): Yes, around 70%.

BancTec: Yes, approximately 50% of our customers capture and match at the line level.

I.R.I.S.: Yes, we do offer automated line-item capture in the last version of *IRISCapture Pro 8.5 for Invoices*. It was

Document Imaging Report

Business Trends On Converting Paper Processes To Electronic Format

DIR is the leading executive report on managing documents for e-business. Areas we cover include:

1. Document Capture
2. Image Processing
3. Forms Processing/OCR/ICR
4. Enterprise Content Management
5. Records Management
6. Document Output
7. Storage

DIR brings you the inside story behind the deals and decisions that affect your business.

Vol. 17, No. 16

Editor: Ralph Gammon
4003 Wood Street
Erie, PA 16509
PH (814) 866-2247
FX (800) 507-8981
ralphg@documentimagingreport.com

Managing Editor:
Rick Morgan
PH (814) 866-1146
rickm@scandcr.com

Publisher and Circulation Manager
Larry Roberts
RMG Enterprises, Inc.
5905 Beacon Hill Lane
Erie, PA 16509
PH (412) 480-5116
FX (412) 291-1352
larry@rmgenterprises.com

DIR is published 24x per year, on the 1st & 3rd Fridays of the month, by:
RMG Enterprises, Inc.
5905 Beacon Hill Lane
Erie, PA 16509
PH (412) 480-5116
http://www.documentimagingreport.com

Subscriptions: $597 (electronic) or $670 (paper) per year.

available on previous versions of the software for specific projects/configurations. Very few of our customers really utilize this specific feature: a lot of them ask for it when they contact us, but we often realize after talking with them about their needs that they don't really need it.

Cardiff: Extracting line-item detail is a built-in part of our advanced invoice solution, and so our customers can deploy this capability. Interestingly, while most customers initially indicate a desire to capture line-item detail from all documents, the reality is that many initially pursue the extraction of core invoice fields, while leaving line-item extraction as a "phase 2." The reasons for this vary, but perhaps the most common is that these organizations were not previously keying in the line-item detail in their manual processing system and may not have the facility to use line-item information right away. Companies in manufacturing and who deal with a lot of partial orders are more likely to implement line-item detail extraction from the beginning.

Kofax: Yes, we offer automated line-item capture. Typically, the larger companies automate line-item capture, whereas smaller companies only capture totals.

Brainware: Brainware's structure-free approach to finding and capturing data means that line-item details stored in lengthy tables (running 25, 50, or more pages) are accurately processed in seconds. All of our customers utilize this feature.

ODT: ODT offers line-item capture, however it is seldom used. This came initially as a surprise, as we considered the automatic recognition rate as exceptionally high. The explanation is this: The only purpose of line-item capture is matching against P.O. items in those invoices where P.O. number and total amount do not match. Through our software, accounting for this type of discrepancy can be done very simply through one, or a few, clicks. The additional system complexity [*related to setting up automated line-item capture*] does not pay off against this minor piece of work.

DIR opinion: It seems that some vendors may be overly optimistic about the number of end users taking advantage of their line-item technology. On more than one occasion, we've found ourselves on the phone with an end-user whom a vendor has recommended to us as using its line-item technology, only to find that while the user may have experimented with it, line-item capture is far from in production.

That said, adoption of line-item capture is on the rise. Yes, it still takes some extra work to implement, but in certain situations, the payoff is worth it. And, based on demos we've seen recently, we expect line-item capture to continue to evolve and become much more of a plug-and-play option in the next couple years. This should lead to more widespread adoption, as the capture of invoice header data followed a similar path to market.

When is a template a template?

Prior to line-item capture's rise to prominence, the most pressing question surrounding invoice capture had to do with templates. Everyone who plays in this space wants to brag about having template-less technology. Of course, the validity of this claim depends on your definition of templates.

The fact is, we really don't see many truly template-less applications. What vendors have mostly done is set up a series of rules that enable users to more quickly define templates. Yeah, you can get away with going template-less if you really want, but, the bottom line is that to improve performance, users are typically employing templates (call them what you want) for their leading suppliers.

Here are the answers we got to the request: Briefly describe how your system is set up for invoice processing. (Do you need to design templates, run sample documents, etc…?)

Datacap: Let's break down the challenge of invoice capture: An AP department gets hundreds, even thousands, of different invoice layouts from their vendor base, yet the same data is required from all of them. The challenge is data location. Datacap takes a multi-part approach that we call Dynamic Natural Analysis (DNA).

First, let's separate the invoice into two sections:
1. the header information with P.O. number, date, remit to address, etc.
2. the detail lines with items, quantities, and total due

The header information (*for a particular vendor*) is almost always consistently in the same place, while the line items below it (and on subsequent pages) will vary with each invoice.

Datacap begins by doing a pattern analysis on the header – taking an impression of the header – the logo, the layout of data, the use of columns and lines, etc. This analysis, called a Fingerprint, is stored in a database and given a name, i.e. "ABC Company." The Vendor ID number assigned to ABC Company is associated with the fingerprint. Next, zones can be assigned and linked to each field in the header, and this is also associated with the Fingerprint. At implementation time, we work with

Invoice Processing Apps

Company name/Invoice product	Total	In U.S.	In Europe	Rest of world	Marquee/Global 2000 customer names
ReadSoft: *INVOICES*	2,557	228	2,214	115	Allstate, Sony, Volvo, Volkswagen, Lockheed Martin, CSC, Glaxo Smith Kline, Kellogg's, Dupont, Time Warner Cable, Siemens worldwide (all process over 200,000 invoices a year)
Kofax: *Ascent for Invoices*	350	130	210	10-plus	The Coca-Cola Company (greater than 500,000 invoices per year), Corporate Express (> 1 million per year), Carrefour Group (tens of millions per year) Bell Canada (>500,000 per year); Exelon (up to 500,000 per year)
Brainware: *A/P-distiller*	appx. 285	appx. 110	appx. 110	appx. 65	Large transportation manufacturer (3.24 million pages/year), large construction contractor (2 million docs/year), large cleaning products vendor (1 million pages/year), leading hygiene products vendor (2 million pages/year), large oil company (3.6 million pages/year), large electricity producer (6 million pages/year)
AnyDoc Software: *AnyDoc®INVOICE*	150	120	30	2	Global 2000 customers include American Electric Power (approximately 50,000 invoices per month), Hagemeyer North America, Inc., Northeast Utilities and Pepsi Cola. Other customers include HG Commerciale, Fastenal, Coop Prozesse Warenwirtschaft and Duke Realty Corporation.
EMC Captiva: *InputAccel for Invoices*	70+	50+	20+	N/A	Brown-Forman (12,000 per month), Renault (700,00 annually), Eloir (900,000 annually).
Datacap: *Taskmaster for Invoices*	Approaching 50	Approaching 50	0	0	Northrop Grumman, Goodyear, DHL, Kuehne + Nagel, Georgia-Pacific
Top Image Systems: *Freedom for Semi-Structured Applications & Invoice Reader*	appx.120	12	90	18	Hormel Foods, Agfa, Network Rail (UK), Italy's largest retail chain
BancTec-*AP Master powered by Contempus* (BancTec packages the Contempus products, combined with its own products to provide a complete invoice imaging and automated purchase to pay solution.)	165	5	160 Europe and rest of world	NA	Dixons (2 Million per year), Yum Brands, Scandinavian Airline Systems
Peladon Software: *DocXP4 Invoices*	3	1	2	0	O'Charley's
Océ Document Technologies: *Invoice Center*	less than 100 direct sales	No end-user business in U.S.	less than 100	0	Deutsche Post, Deutsche Lufthansa, Beiersdorf, Swiss Post

Invoice processing is one of the hottest markets in document capture. Every week, we hear of several new deals in this rapidly growing and evolving space. Indeed, after a couple years of false starts, IDR (intelligent document recognition's) first killer app is a reality. This charts features some metrics related to North America's leading invoice processing software vendors.

Pricing Model	Range of applications	Median/average sized installation
Based on number of documents per year and if line items are used or not. The number of seats included is usually sufficient.	10,000 invoices per year to over 6,000,000 a year	Typical implementation is 50,000 to 250,000 invoices per year.
Volume-based, meaning a customer would need to buy a software license to process a defined maximum number of pages per month or per year, etc. It is a one-time license fee, not a click charge.	hundreds per day to thousands per day	NA
Flexible, tailored to business needs of customers, which range from direct users to resellers, OEMs, BPOs and Shared Service Centers; range from perpetual licenses (with one-time royalty fees plus maintenance) to click-basis (ongoing royalties as end users process documents).	1,000 invoices a day to as high as 3.5 million-plus per year (many customers are licensed for, and moving toward, much higher volumes)	The bulk of Brainware's customers are large ($100 million annual revenue and up), Global 2000 companies.
Per page and per module.	100 per day to 100,000 per day	Median is 3,000 to 10,000 invoices per day
Based on the volume of invoices being processed, however, the cost of the required *InputAccel* server begins at approximately $15,000 and increases depending on the number of images being captured by the solution.	2,000-15,000 pages per day	Average volume: 6,000 pages per day
Concurrent users.	300-500 per week to 2 million per year (9,000 per day)	NA
The most common model we use is per station (automatic and manual) but we also offer volume-based pricing.	between 50,000 and 2 million invoices a year	Between $100,000 and $1 million, average project is for $200,000
Per user for solution implemented at client site; per invoice for hosted application.	200 invoices per day (48,000-50,000 per year) – to more than 2 million invoices per year.	100 users, 200,000 invoices per year
Trying to stay away from click charges, particularly if customers are allergic to them; ultimately, though, the price has to be proportional to use; the more users (or volume) the higher the price.	NA	Several thousand documents per day
Per invoice (independent of number of pages). Also, we offer royalty and click-based models.	Minimum 40,000 ivoice per year	$200,000

customers to identify the top vendors and set up a Fingerprint for each—setting up a new invoice takes only minutes.

After the Fingerprint has been established in the Fingerprint Library, every invoice from ABC Company will trigger the Vendor ID and data location for their particular invoice layout. But, remember the multi-part approach—if zones don't catch all the data, Datacap's rules are evoked to search for key words or aliases.

If key word searches don't work, there is another back-up plan that we call "Click 'N' Key." Since all the text on the invoice has been captured with OCR, Datacap connects the OCR data with the invoice image on the Verify screen. If a field is not filled-in for any reason, the operator simply clicks on the data on the image to automatically populate the field.

Now let's address the bottom of the invoice – the detail lines. These are typically arrayed according to column headers. Once the first line has been identified with the proper column separators, all the subsequent lines can be captured using location rules. The column formatting is stored with the Fingerprint, but following Datacap's multi-part approach, if the lines aren't captured, the operator uses a feature called Dynamic Details. He or she highlights the first line, sets the column widths, and then all subsequent lines are automatically captured. Math calculations on the line items are compared against the total and act as an extra validation.

When a brand new, never-before-seen invoice arrives, we can use the feature known as Intellocate. Simply put, Intellocate enables the Verify Operator to add a new invoice to the system with just a few clicks—no programming required. For a new invoice format, there is no Fingerprint, so the Operator clicks the "New" button and assigns the Vendor ID from the Master File. Datacap location rules will have located much of the data, but Click 'N' Key can be used to get the rest. Dynamic Details captures the line items and when the image is uploaded, it creates a new Fingerprint and stores it in the Fingerprint Library, with Vendor ID, data locations and line items all memorized for the next time that vendor submits an invoice.

Brainware: Brainware's approach to auto-classification and auto-learning requires only a small "learning set" of sample documents to begin processing large document volumes. For one of our largest customers (one of the world's leading energy services firms), we required only 31 invoice samples to build the learn sets needed to begin processing up to 2 million invoices a year from approximately 500,000 global vendors (within an **SAP** ERP environment). Field extraction rates of 80% to 90% are fairly routine, "right out of the box."

Once those learning sets are created, Brainware's "supervised learning" approach allows users to quickly expand them when new exceptions are encountered, without requiring the creation or editing of templates or forms. With a few mouse-clicks, users "train" *IDC-distiller* to learn new document types.

AnyDoc: *AnyDoc®INVOICE* can process better than 90% of a document right out of the box, using its default rules and algorithms. The remaining 10% can be captured by tweaking the algorithms (the attributes that allow us to locate and capture the data). Because the attributes are not hard coded, programming isn't required to customize *AnyDoc®INVOICE* to match a customer's existing business processes.

Templates are not required with *AnyDoc®INVOICE*. Instead it uses AnyApp™ technology, or "memorization," to capture data regardless of where it is located on the page from one vendor to the next. This memory technology remembers the location of data on a particular invoice, increasing the speed of processing the next time that invoice format is processed. During setup, *AnyDoc®INVOICE* is also "connected" to the client's back-end accounting or ERP system to perform validation and retrieval lookups using existing databases, such as using a P.O. number, to verify a vender's name and address.

EMC Captiva: *InputAccel for Invoices* is powered by the EMC Captiva *Dispatcher* classification software. *Dispatcher* is an intelligent document recognition engine that automatically identifies incoming documents using a combination of full-page, text-based and image-based analysis, and then routes the documents based upon pre-defined business rules. *InputAccel for Invoices* includes self-learning templates for high-volume vendors and automated matching of invoice line-item details to purchase orders, so that only exceptions are presented to operators

Kofax: Ours is a self-learning system, meaning there are no templates. We create a single knowledge base for all invoices an organization is looking to process. For example, one of our customers implemented an invoice processing system, which has been in place for approximately three years. This customer will typically add about 100-120 new suppliers each month. For the three years they have been utilizing the system to process invoices, they have not needed to change the knowledge base.

ReadSoft: Based on feedback from the operator, the system automatically learns the layout of specific supplier invoices as needed and remembers this for the next time invoices from the supplier are encountered. The information automatically learned can be manually corrected if needed. Supplier layouts that are seldom used or that do not provide good performance can be automatically removed from the system. Also, support for some electronic invoice formats is built-in and easy to extend. The system does not rely on imported information, but the automation level can be improved by importing information about suppliers and buyers.

TIS: Our free-form product, *Freedom for Semi-Structured Applications*, is based on script, and thus automatically identifies various semi-structured documents without the need to predefine the structure of each document and create a template at start–up. To improve performance over time, we added a unique *Automatic dynamic learning* mechanism, which means the performance improves in proportion to the amount of documents the system processes. With this tool, manual corrections in the completion are automatically learned and stored. Next time a similar document enters the system, this information enables the system to efficiently and accurately recognize it. In addition, we offer *Invoice Reader*, a standard, generic invoice capture and approval solution based on *eFLOW* platform. It includes sets of definitions and rules that can serve as a good start for an invoice processing project.

Océ ODT: The capture process of *Invoice CENTER* from Océ Document Technologies uses rule-based free forms recognition together with a learning module (Adaptive Read Technology). The rules base has been developed using greater than 1,000 invoices from more than 20 countries and achieves between 70% and 95% accuracy depending on the type of information. This is the starting point, which the learning module improves upon. If special circumstances require, templates can be defined.

A look at back-end integration

The third big topic when it comes to invoice processing is integration with back-end accounting and ERP systems. Initially, this was not a high priority for the data capture specialists who pioneered invoice capture. However, over the past couple years, as invoice processing adoption has moved from the "innovators" to the "early adopter/early majority" stage of the technology life cycle, integration with other systems has become paramount.

We don't think it's any coincidence that as vendors have placed a higher priority on back-end integration, invoice processing software installation rates have soared. Here are the answers we received to the question: **On what level are you seeing integration with back-end systems?**

AnyDoc: Nearly 100% of our customers choose to integrate *AnyDoc®INVOICE* with their back-end accounting or ERP system. The level of integration is fairly evenly split between releasing only the data, releasing both the data and images, and releasing the data, images, and performing two- or three-way matching between a P.O., an invoice, and other database lookups.

ReadSoft: The level differs for different customers and partner solutions but most do the matching in the target system. Our SAP integration, *COCKPIT,* and **Oracle** integration are typical examples that match captured data with system data, point out all problems, and offer quick access to related system information in a well-known environment.

EMC Captiva: EMC's overall solution for automating the purchase-to-pay cycle includes the SAP-certified *Documentum Content Services for SAP* integration that directly links relevant business content. From within the SAP application, users can access all the relevant content related to an AP transaction, whether it is a scanned invoice, vendor contracts, correspondence or other content.

By combining Captiva *InputAccel* for capturing information from paper invoices with *Documentum Content Services for SAP*, organizations can export the captured data as an SAP transaction that triggers an SAP workflow. The Documentum system handles some document validation/approval processes, while others are managed via SAP. *Content Services for SAP* enables users to add descriptive metadata to the SAP system, so SAP transactions can be populated or created using captured information.

This is one example of how EMC solutions integrate with ERP systems. Other integrations are with systems from Oracle, **Microsoft**, **JD Edwards**, and **Sybase**.

TIS: Our ***eFLOW Ability*** is a complete solution for accounts payable process and approval workflow and is designed to integrate with SAP *R/3*. It allows for two-way communication with SAP. It matches the extracted information and verifies it against information that already exits in the SAP, such as vendor details, P.O. information, etc. Then, it delivers the information into SAP for further use.

Our *Invoice Reader* is a standard, generic invoice capture and approval solution that may be deployed and integrated in any enterprise accounting

environment including SAP, Oracle and other financial systems. With this module, extracted information can be automatically verified in assorted ways, including comparison against internal tables or external databases from the ERP systems, as well as comparative calculations.

ODT: Most integrations go deep into the ERP system. Data is imported and a booking record is created. Two- or three-way matching is standard. An approval workflow and a process monitor a mandatory.

Kofax: The vast majority of our customers are matching against P.O.'s. If they pick up line items, they will include them in the match.

Brainware: We estimate about 80% are matching against P.O.s.

More next issue

Well, that's probably enough on invoice processing for this issue. After all, we're running out of space. We do promise more next issue including responses to questions such as, what is the difference between the U.S. and European markets when it comes to invoice processing? What are the hottest verticals? And what new trends do you see in invoice processing and IDR? Until then…

For more information:
http://www.readsoft.com/products/;
http://www.kofax.com/products/ascent/invoices/details.asp;
http://www.brainware.com/invoice.php;
http://www.anydocsoftware.com/software/solutions/invoice/intro.html;
http://www.captivasoftware.com/products/ia4Invoices.asp
http://www.datacap.com/products/invoices/;
http://www.topimagesystems.com/Page.asp?Par=2&id=14;
http://www.banctec.com/solutions/business-process-outsourcing/invoices/;
http://www.peladonsoftware.com/docxp4invoices.htm;
http://www.odt-oce.com/usa/?ul=en.

VISIONEER, FROM PAGE 1

of large deals that feature the company's OneTouch technology for creating a simpler scanning interface.

According to Bill Kouzi, Visioneer's senior VP of sales and services, the **PA Dept of Welfare** recently purchased 300 scanners, and **TCI Tire Centers**, a **Michelin** subsidiary, purchased 180. The Dept. of Welfare deal came through **SRC Solutions**, a Dunmore, PA-based reseller, and features OneTouch integration with **Xerox's** *DocuShare* software. The Tire Centers deal came through Liberty, S.C.-based **KeyMark** and features **Kofax** *Ascent* OneTouch integration. Both deals leverage scanners with Kofax VRS embedded in the OneTouch driver [*see DIR 2/2/07*].

"We now have a product offering that really enables resellers to make scanners an integrated part of their document imaging solutions," Kouzi told *DIR*. "We are very excited about our work with all components of our channel. We are working more closely with our distribution partners, and we also expect shortly to have significant announcements involving OneTouch integration by ISVs."

For more information:
http://support.visioneer.com/products/software/SDK/default.asp