

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>

April 20, 2007

THIS JUST IN!

ABBYY INTRODUCES NEW APP

At this week's **AIIM Conference and Expo**, **ABBYY USA** will be introducing a new version of its forms processing application. *FlexiCapture Professional 8.0* combines ABBYY's structured and variably structured forms processing tools into a single application. ABBYY has traditionally used its *FormReader* (structured) and *FlexiCapture* (unstructured) applications as showcases for its automated data capture toolkit.

"We've improved our classification capabilities, so that less document preparation is required," said Tsiling Merrem, director of marketing at ABBYY USA. "Business users don't talk about documents vs. batches. They just know they have a pile of invoices that needs to be scanned. *FlexiCapture* can now automatically identify which documents in that stack are invoices, and which are attachments."

There have also been improvements to the GUI, and *FlexiCapture* can now better capture data from tables that span multiple pages. Merrem stressed that *FlexiCapture* is not limited to invoices. "There has been a barrier to entry for automated data capture in the SMB space, because users simply don't have the budget to invest in a solution to capture 100 invoices per week," said Merrem. "We made our solution flexible enough that it can be used in multiple departments to capture various types of documents."

For more info: <http://www.abbyy.com/flexicapture/>

VOLUME NUMBER CORRECTION

We forgot to update the volume number in our March 23, 2007 edition. Instead of Volume 6, we incorrectly identified it as Volume 5. The Volume number represents how many issues have been published to date in particular calendar year. This is Volume 8, as we publish twice per month and this is April's second issue. Unfortunately, our mistake carried over to the print edition of the April 6 issue, but we caught it before the electronic edition went out. We apologize for any inconvenience caused by this mistake.

Kodak Adds Teeth To Distributed Scanner Program

New VP of sales details plans to beef up sales of lower volume scanners.

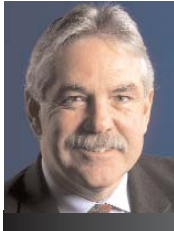
It's no secret that **Kodak** has mostly been on the outside looking in when it comes to workgroup scanner sales. This is despite, over the past five years, Kodak's release of a slew of new scanners aimed specifically at the segment it refers to as distributed scanning. (This roughly translates into scanners with a list price below \$10,000.) Yes, while Kodak remains the undisputed leader in unit sales of high and mid-volume scanners, it has mostly been treading water in the distributed segments—growing at approximately market pace, but missing out on the large volumes captured by competitors like **HP** and **Fujitsu**.

Enter former Fujitsu and **Visioneer** sales exec Don McMahan, who was brought in by Kodak specifically to reverse its fortunes in the distributed segments. In January, McMahan was named VP of sales for U.S. and Canada. "Historically, Kodak has put its emphasis on production scanning at the expense of the distributed segments," McMahan told *DIR*. "The overall market growth in those higher volume segments is essentially in the single digits or flat. Kodak's business model has not enabled it to compete effectively in the higher growth segments of the market."

Kodak's sales team has traditionally been focused primarily on driving sales through its KAIR (Kodak Authorized Imaging Reseller) channel. The KAIR channel consists of approximately 200 VARs, many of whom broke into the imaging business when mid- and high-volume scanner sales were king, and workgroup scanners represented just a blip on the radar screen. Over the past five years, however, workgroup scanner unit sales have increased 10-fold, and it's McMahan's opinion that Kodak's sales team has not transitioned to fully exploit this change in market dynamics.

"When I joined Kodak Document Imaging, we had one sales team that was geographically aligned," said McMahan. "That sales team supported both the KAIR

channel, as well as the KDSR [Kodak Desktop Scanner Reseller] program. The KDSR program was launched a couple years ago to drive distributed scanner sales, but at the end of the day, smart sales reps are going to gravitate toward what works best for them. More often than not, under our legacy sales structure, that turned out to be supporting the KAIRs. This meant our reps didn't have time to grow sales outside of the KAIR community at the rate we would have liked."



"Historically, Kodak has put its emphasis on production scanning at the expense of the distributed segments."

-Don McMahan, Kodak

To rectify this, McMahan has created two separate sales teams. "We've taken the top sales people from our previous structure and made them strategic account managers focused on KAIRs," he said. "We've realigned our geographic regions into five territories of KAIRs and have started the process of reducing our number of KAIRs by about 30%. When all is said and done, we expect to have about 150 KAIRs who will have to meet minimum sales requirements to keep their status. While our strategic accounts team will market our entire product line to KAIRs, their primary focus will be leveraging their channel to help us defend and grow our production scanner sales."

Cultivating a distributed channel

To promote the KDSR program, McMahan has created a new team of channel sales managers. This team will be headed by Todd LeVeque, who was recently named Kodak Document Imaging's director of national sales for distribution

KODAK REACHING OUT TO ISVs

In addition to launching a new sales team focused on distributed capture sales, **Kodak Document Imaging** is increasing the breadth of its ISV partnership efforts. "We have not been as aggressive in our ISV program as we should be," said Don McMahan, Kodak VP of sales for U.S. and Canada. "With other vendors I've worked for, ISVs have been a larger focus. We currently are interviewing to add resources in this area. We need to improve our business development efforts and increase our sales engagements with ISVs.

"You are also going to see a bigger vertical push by Kodak. Also, we are going to be more aggressively taking our check scanning products to market through our strategic account managers [KAIR sales team]. Up to this point, the check scanning products have been mainly a marketing exercise. I'm really excited about the breadth of the product line I have here at Kodak."

<http://www.kodak.com/US/en/dpq/site/TKX/name/ISVsValueAdd>

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. 8

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>

Copyright © 2007 by RMG Enterprises, Inc.

Federal copyright law prohibits unauthorized reproduction by any means including photocopying or facsimile distribution of this copyrighted newsletter. Such copyright infringement is subject to fines of up to \$25,000. Because subscriptions are our main source of income, newsletter publishers take copyright violations seriously. Some publishers have prosecuted and won enormous settlements for infringement. To encourage you to adhere to this law, we make multiple-copy subscriptions available at a substantially reduced price.

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

capture. LeVeque is a former **Kofax** and **Cranel** sales executive who certainly understands the two-tier distribution model prevalent in distributed scanner sales.

“Todd will oversee sales managers in six geographical regions, plus a federal territory,” said McMahan. “We’ve left Canada alone, as the sales structure there is working out well. We have a couple regional manager openings in Todd’s group that we expect to fill before the end of the month. Todd’s group will be focused on recruiting new members for, and developing the KDSR channel.”

According to McMahan, the KDSR program, which was launched in 2005, had 250 members when he joined Kodak. “In two years, Kodak was able to recruit 250 resellers,” he said. “In the two months I’ve been here, we’ve recruited 250 more and by AIIM, 500 will be well in the rearview mirror. By the end of the year, our goal is to hit 1,000. Of course, part of running a successful reseller program is sorting out your keepers, but an initial push to sign up people at least gets them opted-in to your marketing programs and increases your chances of finding keepers.”

To increase interest in the KDSR program, McMahan has made it more lucrative, while also making it easier to enlist. “One thing we’ve done is introduce back-end rebates that are more competitive with the rest of the industry,” he said. “We’ve also gone from a three-page application form that required four forms of identification and rights to your first born, to a single-page form. We introduced the new KDSR program at the recent **Ingram Micro** event, where it was well received.”

Outside inspiration

McMahan concluded by saying he learned a lot during his time spent as a consultant between jobs at Visioneer [which he left this summer] and Kodak. “I looked at some best practices in other industries,” he said. “In particular I did some work for **Zebra Technologies**, which markets bar code printers. They have a lot of large direct accounts in the retail sector, as well as a two-tier distribution model for VARs. I looked at how they have integrated those two sales models, and I think it gave me a better understanding of some things that might work for Kodak.

“The bottom line is that we now have a dedicated KDSR team that isn’t getting any money from the KAIR channel. When they get up in the morning, their only focus will be driving business from KDSRs. It’s going to be a painful split for some people, but change is never comfortable. By the end of the year, I really expect a lot of people to be surprised and

amazed at what we’ve accomplished.”

For more information:
<http://www.kodak.com/US/en/dpq/site/TKX/name/KDSR>;
http://www.kodak.com/US/en/dpq/site/TKX/name/kairValueAdd_us

Are E-Mortgages A Threat To Imaging?

The real estate bubble has been very good for the document imaging industry. Mortgages create paperwork, which is often scanned at least once for archiving and servicing. In recent years, we’ve done stories on businesses like **SourceCorp**, **Kofax**, **Hyland**, and even **Green Button Technologies** that have built successful products and/or practices around mortgages.

That’s why last month when we saw a handful of stories moving across the wires that dealt with the implementation of electronic mortgages, we thought it’d be worth investigating. After all, e-mortgages are a potential threat to one of the lifebloods of the document imaging industry—or are they?

eMortgages are nothing new

The most significant story we ran down had to do with **Adobe** and **MISMO** (the Mortgage Industry Standards Maintenance Organization) announcing the release of guidelines for electronic signatures on PDF mortgage documents. “Currently, there is a lot of electronic transfer of information at both ends of the mortgage process,” said Harry Gardner, senior director of industry technology for the **Mortgage Bankers Association**, MISMO’s parent. “However, there is a hole in the middle when documents get printed and signed at the closing. One of our goals is to create a seamless electronic process from the loan application through the closing, all the way through to the servicing and sale of the loan.”

Gardner pointed out that several years ago MISMO published a specification for a SMART (Secure, Manageable, Archivable, Retrievable, Transferable) Doc format for creating electronic mortgages. “SMART Doc is an openly published spec, based on XML,” he said. “There are actually several categories of SMART Docs, with Category One being the smartest, if you will. Category One Smart Docs include links between data elements in an image file like a scanned deed, for example, and the XML data these elements correspond to.”

Gardner noted that the U.S.’s two largest mortgage financiers, **Fannie Mae** and **Freddie Mac**, require Category One SMART Docs be submitted by anyone sending them electronically formatted mortgages. “However, SMART Docs represent only a very small

percentage of the mortgage volume in the U.S.," said Gardner. "It's probably less than 1%.

"Of course, the benefits of using electronic mortgages can be tremendous, especially downstream, when a mortgage changes hands. That's because electronic mortgages have data that travels with them, and that data can be used to update systems of records on the fly."

Making PDF an option

According to Gardner, the main purpose of the new MISMO guidance on PDF signatures is to promote PDF as a viable option for electronic mortgages. "A significant number of lenders are already using electronic signatures in conjunction with PDFs for internal purposes," he said. "We decided to come up with some guidance that would facilitate the implementation of electronic signatures with PDF mortgages in a consistent format."

Like electronically signed SMART Docs, PDFs need to be registered on the national MERS System to guard against fraud. "In the paper world, the physical possession of a note has the connotation of ownership," said Gardner. "In the electronic world, it's fairly easy to create an unlimited number of copies. The MERS System acts as a master database that tracks when a note was closed, and who owned the loan when the borrower's signature was applied. So, when a lender delivers an e-loan to Fannie Mae or Freddie Mac, they can compare it to the information in the MERS System to ensure the transaction is valid."

We concluded by asking Gardner if MISMO is offering any guidance on the construction of PDF e-loans in addition to the digital signature protocols. "There are ways to leverage XML wrappers to make PDF documents smarter," he said. "But that is not a subject we have addressed yet."

What does it mean?

The bottom line seems to be that adoption of e-loans, although technically possible, is still a long way from gaining mainstream adoption. For the near future, batch scanning applications for mortgages appear safe. That said, in another 20 years, when everyone in the business world is fully comfortable working with Web documents, why wouldn't we transition to e-loans with e-signatures?

However, even after this transition occurs, there will no doubt still be plenty of paper documents that need to be on-ramped into electronic loan files—W-2s, paper deeds, correspondence, etc. This means that as the batch scanning market for loans disappears, a market for ad hoc scanning will take its

place. Visionary capture vendors are already preparing for this transition from batch to ad hoc scanning and should be in good shape to adjust with the dynamics of the mortgage market. As for the dynamics of the real estate bubble... that's covered in another newsletter.

For more information: <http://www.mersinc.com>;
<http://www.mortgagebankers.org/eMortgage>
<http://www.mismo.org/default.html>
http://www.adobe.com/financial/pdfs/emortgage_sb.pdf

New Network Scanning Interfaces Debut At AIIM

Nuance showing EIP application; Kodak partners with NSi

At **AIIM On Demand** this week, look for a couple important announcements regarding capture from networked devices. **Nuance Communications** will be introducing one of the first capture applications to run on **Xerox's EIP** (Extensible Interface Platform.) Also, Kodak and **NSi (Notable Solutions, Inc.)** have co-developed a new capture platform for the Scan Station 100.

EIP is the embedded application interface for MFPs that was previewed by Xerox last fall [see *DIR* 11/3/06]. Nuance, which develops imaging technology that carries over from its days as ScanSoft (a spin-off from Xerox) will be announcing the *PaperPort Personalized Scanning Platform (PSP)*. "PSP is more than an application for EIP," noted Robert Weideman, general manager and senior VP of Nuance Productivity Applications. "It's designed as a plug-in to operate with any open embedded environment.

"With *PSP*, we've introduced into MFPs the concept of personalized scanning. Historically, when users have tried to scan from MFPs, they have been presented with several generic buttons, like scan-to fax, e-mail, print, or a file folder. To do anything useful, they have had to drill down through folders or type in e-mail addresses at the device. In these environments, users will often follow the path of least resistance and end up e-mailing images to themselves, so they can work with the files on their desktops. Any personalization of these interfaces typically requires help from the IT department or a solutions provider.

"With *PSP*, at their desktop, anyone can use a GUI to personalize and control the scanning interface they will see when they log on to an MFP device.

They can set up specific folders as scanning destinations or configure workflows: like scan, convert to Word, and then store files in a specific *SharePoint* folder. We've even applied for a patent around this scanning method. By allowing users themselves to personalize what they will see on the MFP touchscreen, we think we have significantly reduced the cost of ownership of scanning from MFPs."

Nuance is no stranger to MFP capture. The company already has OEM deals with the likes of **Xerox**, **Kyocera Mita**, **EFI**, and **Konica Minolta**, all of which use some form of *PaperPort* as a desktop client for scanned documents. "With *PSP*, we've gone one step further than our previous integrations and created two-way communication with the MFP," noted Weideman. "This is part of the evolution of the MFP market. A few years ago, scanning from an MFP was a pretty dumb process. Now, vendors like Xerox are incorporating more intelligence and capabilities in their devices. This enables application developers like Nuance to offer more functionality directly through the device.

"Platforms like EIP, which enables ISVs to deploy applications through a Web interface, are going to change the dynamics of the capture market. Leveraging EIP, with *PSP*, Nuance has been able to make it very easy for an office worker to complete scanning tasks at the MFP. In the same way we revolutionized dedicated scanning with our patented OneTouch technology, we think we can revolutionize network scanning through *PSP*.

"And through EIP, we have been able to achieve this without going through middleware from a vendor like **eCopy**, NSi, or **Omtool**. If middleware vendors are going to survive in the future, they are going to have to offer a lot more functionality than connections to other applications. With the emerging paradigm of embedded MFP platforms, more ISVs will choose to integrate directly with these platforms, which could marginalize the need for middleware."

For more information:

<http://www.nuance.com/scansoft/>

AutoStore for Scan Station

The NSi and Kodak deal has been in the works for some time. Kodak introduced the Scan Station last year, specifically targeting **HP's** Digital Sender market [see *DIR* 3/17/06]. It's probably no coincidence that NSi's *AutoStore* was originally developed as a scanning platform for the Digital Sender.

"The integration with NSi will enable Scan Station

customers to utilize the entire *AutoStore* ecosystem," said Will Hebert, who was recently named Kodak's portfolio business manager for distributed capture. "This includes *AutoStore's* document processing workflow, which can perform tasks like OCR/ICR, bar code reading, encryption, and authentication. *AutoStore* also has direct links with Exchange and Notes and supports the leading fax server products. *AutoStore* for the Scan Station will even enable image previews on the device touchscreen."

AutoStore features back-end connections to some 30 ECM systems. "It enables indexing right at the touchscreen at the time of scanning," said Hebert. "There is also an API available to ISVs or resellers who want to do custom development and possibly output documents from NSi to a new destination.

"While we will continue to sell the Scan Station with our own interface [which also has been upgraded], *AutoStore* addresses a different piece of the market. For customers, it offers a richness of features and robustness they can't achieve solely with our operating system. For the channel, it creates a greater opportunity for professional services. It also enables integrators to build customized solutions and more effectively market the Scan Station to specific verticals."

AutoStore is a client/server application which leverages networked scanning devices as clients. Each client seat of *AutoStore* lists for \$600. NSi currently offers two versions of its server—*AutoStore Express*, which is free but has limited capabilities [see *DIR* 3/16/07] and the full server, which lists for \$3,700. Although details of the sales model are still being worked out, Kodak plans to embed the *AutoStore* client in every Scan Station, giving users the option of buying it after a 30-day trial.

State of the market

Word on the street is that initial sales of the Scan Station have been slow. This is not surprising due to Kodak's traditional channel make up, which we discuss in our cover story. As the Scan Station 100 is very much a distributed device, it should benefit from the re-structuring of Kodak's sales team.

However, in addition to channel issues, we were generally unimpressed with the initial operating system on the Scan Station. Yeah, it was fairly easy to use and navigate, but it was clearly a couple generations behind interfaces developed by network scanning market leaders like NSi, eCopy, and Omtool. Obviously, jumping into bed with NSi eliminates this disparity. It also immediately adds more intelligence to the Kodak device—which, as Nuance's Weideman noted, is the wave of the future. However, in this case, it is a middleware

vendor adding the intelligence, with the hardware vendor piggybacking on the work the middleware vendor has already done. Of course, if the hardware vendor was to eventually acquire the middleware vendor...It certainly will be interesting to watch the dynamic between middleware and embedded vendor platforms play out in the market.

For more information: <http://www.nsius.com>;
http://www.kodak.com/US/en/dpq/site/TKX/name/scanStation100_product

ADDENDUM

Canon, Captaris Introduce RightFax MEAP Integration

It's probably worth noting that there was one announcement last week that seems to strike a blow in favor of embedded platforms over middleware. That was availability of *RightFax* MEAP (multifunctional embedded application platform) integration for **Canon** digital copiers. *RightFax* has been primarily integrated to Canon ImageRunners through **eCopy's** *ShareScan* middleware.

Historically, the *Connector for RightFax* has been one of eCopy's best-selling third-party application *Connectors*. "Without *ShareScan*, there has not really been a good way to connect ImageRunners to *RightFax*," said Dennis Amorosano, director and GM, Solutions Business Development Division of Canon USA. "The direct integration through MEAP will be more economically palpable for users who do not necessarily need all the document distribution capabilities of *ShareScan*."

The direct *RightFax* MEAP integration was co-developed by Canon and *RightFax* vendor **Captaris** through a partnership announced in *DIR* last year [see *DIR* 2/3/06]. The *RightFax* server product will continue to be available through Captaris and its reseller channel. The *RightFax* MEAP component will be available from Canon to its distribution channels. Of approximately 15 MEAP applications, Canon only has previously distributed *ShareScan* and a legal services application from **Ribstone**.

"*RightFax* is well-entrenched in many large organizations around the world," said Amorosano. "And, Canon has a healthy number of shared customers with Captaris. This MEAP integration gives our dealers a tool to enable their customers to tie together existing investments in *RightFax* with MFP investments in Canon hardware."

Pricing starts at \$999 for a single MFP unit

integration with *RightFax*, with per-unit prices dropping as volume rises. In contrast, *ShareScan* lists for \$2,995 per device with a *RightFax Connector* listing for another \$1,500.

Two-way communication

Mike Spaccone, general manager of Captaris' MFP practice noted that while ImageRunners aren't the first MFPs *RightFax* has been integrated with, MEAP has enabled a richer integration than previously possible. "The big difference is that through MEAP, we are able to achieve two-way communication," he said. "Historically, users were entering information into *RightFax*, but they couldn't get anything back from it. This means they had to be prepared before they approached the device.

"They needed specific phone numbers, cover sheets, or had to know 15 different fax codes. With the MEAP integration, after the users log in, they can have the same type of *RightFax* experience they would at their desktops. They can choose from phone book entries, billing codes, they can initiate password protection, track the status of jobs, and receive notifications when jobs are complete."

We asked Spaccone if Captaris was considering integrating with any other vendors' embedded platforms. "Canon is currently our only two-way connector, but we are evaluating what we should do next," he replied.

Why fax servers?

Amorosano noted that *RightFax* is a perfect application for MEAP deployment. "While we would like to have MEAP integration with more applications than we do today, the reality is that for a vendor to invest in MEAP development, there has to be strong market demand," he said. "The *RightFax* integration was customer driven, as fax servers are becoming an integral piece of an MFP's workflow."

Spaccone concluded by pointing us to a white paper recently authored jointly by Captaris and **Cantata Technology** that details the advantages of fax servers over fax boards when working with MFPs. "A fax server gives you total control of your faxed documents as they come in and out of your organization," he said. "In a smaller implementation, using fax cards may save some money, but they create greater risk. Our white paper details how, with as little as five MFPs connected to your fax server, you can start to see an ROI."

For more information:

<http://www.documentimagingreport.com/Fax.1543.0.html>

<http://www.captaris.com/rightfax/>

http://www.usa.canon.com/templatedata/pressrelease/20070410_captaris.html

ReadSoft Invoices Patent Revoked

Top Image Systems (TIS) is celebrating a victory over **ReadSoft** at the European Patent Office (EPO) in Munich. In March, based on a challenge filed by TIS in 2004, the EPO revoked a ReadSoft patent on automated data capture from variably structured forms. The patent potentially covered technology used in the fast-growing invoice and EOB processing, as well as digital mailroom markets.

“We believe the decision will have a positive effect on the invoice processing market,” said Ofir Shalev, VP, R&D and CTO of TIS. “The claimed patent was quite general, and like any general claim, it might seem similar to other technologies available today in the market. Although TIS was the only company to actively address this issue, we believe other companies will benefit from the decision as well.”

The patent in question is number EP 0976092 entitled “Method and Arrangement for Automatic Data Acquisition of Forms.” It was filed and published in 1998 and approved in 2003. It’s our understanding that having a patent revoked by the EPO is much less difficult than achieving a likewise result from the USPTO. ReadSoft has the opportunity to appeal the revocation.

The description of the patent includes the following statements:

- 1) “...automatic data acquisition...forms whose design and information is not known in advance, by input into the said means together with storage of patterns of the same.”
- 2) “A need therefore exists for all who handle invoices and other forms to quickly be able to identify these and/or quickly be able to enter and store new patterns in their invoicing system.”
- 3) “The present invention concerns a system (method and arrangement) for the automatic data acquisition of forms where the system has no prior knowledge of what the form looks like or where on the form the information is to be found. In this way, templates of forms do not have to be defined in advance, but are instead registered as they are submitted to the system, i.e., in real time.”

“The basic idea behind our challenge was that there is a big difference between how general the claims in the patent are and the actual description of the technology for achieving those claims,” said Shalev. “For example, there is a claim describing ‘...automatic data acquisition of forms where the system has *no prior knowledge* of what the form looks like or where on the form the information is to

be found.’ But this claim is not technically supported. TIS has developed a set of advanced technologies that are part of our current version of *eFLOW*, and we know from our own experience that building a recognition system that works on *any* type of form, without any prior knowledge, and without human intervention during the learning process is not a trivial task.

“If indeed there is a clever algorithm that does this, the relevant details should be disclosed as part of the patent. We claimed that there were no relevant details of how to achieve such an ambitious goal, which means it is insufficiently disclosed in the claim. We had other arguments regarding the novelty of the parts that were disclosed, but eventually we did not need to use them at all.”

This ruling would seem to be good news for everyone in the rapidly growing IDR (intelligent data recognition) market, especially after the notorious Millennium patent that plagued the traditional forms processing industry a few years back [see *DIR* 1/9/04]. Granted, there is one major difference between Millennium’s patent and ReadSoft’s—ReadSoft is actually a viable business with a productized implementation of its patented technology.

It’s also important to add that, while ReadSoft did not wish to comment on the patent when we queried them about it earlier in the year, we had not heard anyone say ReadSoft had plans to sue for infringement. In fact, the word we had received was that the patent was more or less filed as a defensive measure to prevent another Millennium from coming along and making claims on IDR. If that was indeed the case, the recent EPO ruling should hopefully achieve some of those same defensive goals.

Finally, this case is just the latest example of the headaches that software patents can bring to bear. One of our favorite stories coming out of AIIM last year was that long-time automated data capture industry consultant Arthur Gingrande had decided to attend law school to become a patent attorney. He apparently had recognized a significant opportunity after being called as an expert witness in several cases. While no one has ever accused Art of being the most staid of fellows, the fact is, someone with his experience could bring some more sanity to what is often a very confusing process.

For more information:

<http://www.wipo.int/pctdb/en/wo.jsp?IA=SE1998000602&DISPLAY=DESC>
<http://www.topimagesystems.com>

Google Announces Open Source OCR Project

Last week, Web search kingpin **Google** announced it was launching an open source OCR project that will be made available under the Apache 2.0 license. The initiative is being called OCRopus and is sponsored by Google and being worked on by the Kaiserslautern, Germany-based **IUPR** research group. The project is based on the Tesseract OCR system, as well as some handwriting recognition technology software that has been used by the **US Census Bureau**. Google also brings technology to the table in the area of layout analysis.

DIR caught up with OCR/ICR engine analyst Christopher Riley of **Artsyl Technologies** and asked his opinion on OCRopus. "The Tesseract engine has been in Google's possession for quite awhile," Riley told *DIR*. "So, we knew the project was coming. Tesseract has been available as a SourceForge project for several years and was born out of the UNLV testing and **HP**. When you look at the character-level OCR with its limited morphology and dictionary support, it's slightly less accurate than the **Expervision** commercial engine."

Riley views document analysis as the most intriguing piece of the project. "Document analysis is an extremely important part of OCR and, knowing Google's team, within a few years they could potentially develop the most accurate document analysis around," he said. "Only two vendors commercial engines have really mastered this."

Riley went on to say that he does not necessarily

view Google itself as a direct threat to OCR vendors. He does, however, view the open source engine as posing a potential challenge to OCR toolkits. "I'm sure Google will use the technology to complement the Million Book Project and in other tools embedded into Gmail and Google Docs," Riley said. "The real threat to the commercial OCR market could come from independent developers who decide to take the engine and run with it.

"The technology's true power could be unleashed when it is set into motion for a niche type of processing, and fine-tuned to do it well. An example of this could be ID card reading, or OCR of numbers and mathematic equations. Also, as independent developers take the engine from the desired platform of Google (Linux—specifically a modification of Ubuntu 6) into other platforms, then there will be more potential to create some powerful, more commodity OCR and data capture solutions.

"The engine under initial testing is not viable for a production-level environment. It is, however, one of a limited number of Linux engines, and, even in the enterprise, Linux is becoming more common because of security and stability advantages. I suspect we will see many talented developers take the engine to higher levels with some impressive results, but I don't think Google itself, or any other large commercial development pool, will take this engine to market sooner than the next two years."

For more information:
[http://code.google.com/p/ocropus/;](http://code.google.com/p/ocropus/)
[http://www.artsyltech.com/;](http://www.artsyltech.com/)
<http://www.iupr.org/doku.php;>
<http://sourceforge.net/projects/tesseract-ocr>

Subscription Order Form for RMG Enterprises, Inc.

5905 Beacon Hill Lane • Erie, PA 16509 • Phone (412) 480-5116 • Fax (412) 291-1352 • <http://www.documentimagingreport.com>

Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

1 year (24 issues)

- electronic copy @\$597
- paper copy @\$670

SCAN: The DATA CAPTURE Report

Premier Management & Marketing Newsletter of Automatic Data Capture

- electronic copy @\$597
- paper copy @\$670



Name _____ Title _____

Company _____

Street _____

City _____ State _____ Zip _____

Phone (____) _____ Fax (____) _____

E-Mail _____

Please enter / renew the following subscription.
(Add \$33 on all orders outside of Canada and the United States.)

Payment Enclosed (Remit to: RMG Enterprises, Inc., 5905 Beacon Hill Lane, Erie, PA 16509)

Charge My Credit Card (Charge will appear as RMG Enterprises.)

____ AmEx ____ Visa ____ MC ____ Discover _____
 _____ card number _____ expire date

Bill My Organization (Purchase order # optional.) _____