

Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

4003 Wood Street ● Erie, PA 16509 ● PH (814) 866-2247 ● FX (800) 507-8981 ● <http://www.documentimagingreport.com>

December 2, 2005

THIS JUST IN!

NAMING NAMES

You probably all saw the recent announcement by **Kofax** that it has signed an OEM agreement for **Indicius** with a “major ECM vendor.” Well, there has been some talk floating around that this could be **FileNet**, and we’d have to agree that this certainly makes the most sense. *Indicius*’ forms processing capabilities would seem to be an ideal complement to the document capture application FileNet has already developed on its own [see *DIR* 11/16/05].

The only other ECM vendors that we can think of that would significantly strengthen Kofax’s “position in the enterprise, fortune-1000 market” are **Documentum**, **IBM**, and **Open Text**. Documentum, of course, is spoken for, and IBM and Open Text both have significant partnerships involving Kofax *Ascent*, so it wouldn’t make much sense for them to OEM only *Indicius*.

So, the big question is, will the FileNet agreement eventually lead to an acquisition of Kofax? Despite some rumors, we’d have to say the answer is no, for several reasons. First off, with less than \$400 million in the bank, FileNet does not have nearly the resources of **EMC**, which has offered \$275 million in cash for Kofax competitor **Captiva**. Second, FileNet and Kofax sell their products through very different channels. Unlike Captiva, which sells primarily direct, Kofax has a large VAR channel. This channel is a huge asset to Kofax, but would be of little value to FileNet—based on past history that includes a disastrous foray into the mid-market with the acquisition of **Watermark** in the mid-1990s. Finally, what would FileNet do with Kofax parent **Dicom**’s distribution business (and what would Dicom do without Kofax)?

The reason FileNet wasn’t named probably has to do with its desire to present a unified ECM solution, including all elements of capture. After all, this is what EMC/Documentum will be doing sometime following the first of the year.

Westbrook Beating The SOA Drum

Publishes white paper designed to increase market awareness of potential of Web services

“SOA is the be all and end all—the alpha and the omega.” These are the types of things we’ve heard (we think half in jest) over the past year from ECM executives when we’ve asked about service oriented architecture. There seems to be no question that SOA is going to be important—but in what way? So far, we haven’t heard a lot of specifics about SOA’s ability to improve document imaging. **Westbrook Technologies** is trying to change that.

Westbrook recently contracted long-time industry and RFP consultant Bud Porter-Roth to author a white paper titled *Introduction to Service Oriented Architectures for Electronic Document Management Systems*. Westbrook President and CEO Paul Lord refers to the paper as “A Dummy’s Guide to Web Services.” Westbrook, which is currently developing an SOA-based application, is trying to establish itself as a thought leader in the emerging SOA landscape.

“There is an ongoing drumbeat surrounding SOA,” Lord told *DIR*. “We want to establish ourselves as part of that drumbeat and create some mindshare in the SOA arena. We have challenges to overcome, because historically, Westbrook hasn’t necessarily been considered a cutting-edge company. Publishing these types of papers will help us recreate our image. In addition, it’s important to have people understand Web services rather than be scared away by them. We plan to follow up with another paper by a technology consultant that gives specific examples of how SOA standards can be implemented in conjunction with our document management application.”

Porter-Roth agreed there is a need for explanation

of SOA within the document imaging market. "I was recently talking to the lead integrator for one of the major companies in the industry," related Porter-Roth. "I asked him what he thought about SOA. He replied by asking me what it was."

Standards Make Integration Easier

Porter-Roth's white paper can be found on our Web site at http://www.documentimagingreport.com/fileadmin/PDF_Content/SOA_Architecture_White_Paper.pdf. It does a good job describing the basic components of an SOA-environment, as well as some of SOA's advantages. It provides a specific document management-related example of a loan being processed. Steps in the process include scanning, OCR, indexing, archiving, new account creation, and approval workflow. None of these is revolutionary on its own. It's their integration, and how they work in connection with multiple line-of-business applications, that show the power of SOA.



"This is the first time in the history of document management that a product has been built from the beginning to collaborate with other products."

— Paul Lord, Westbrooke

Yes, we've had integration between ECM and ERP systems for years, but in an SOA environment, this integration is based on standards. XML is one of those standards. Others include SOAP (simple object access protocol), WSDL (Web Services Description Language), and UDDI (Universal Description, Discovery and Integration). "Because SOA is based on standards, SOA applications are designed to be re-used in multiple environments," Porter-Roth told *DIR*. "Instead of one-to-one integration, like you get through APIs, SOA applications benefit from a one-to-many framework. This reduces software developers' costs.

"In addition, SOA integration is more robust. The standards-based data and protocols are not reliant on fragile API connections. In other words, if the ERP application breaks down in the loan processing example, it won't shut down the whole process. The user still has data that can be understood by other applications."

Lord adds that the potential for interaction among SOA applications goes further than the walls of the traditional enterprise. "This is the first time in the history of document management that a product has been built from the beginning to collaborate with other products," Lord told *DIR*. "For the past 20 years, when there has been a handshake with another application, it's always been within the network. SOA enables you to take that integration outside the enterprise."

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. 15, No. 23

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 © 2005 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.

In the loan processing example, a connection with an outside credit rating service is shown. *DIR* has also spoken with forms processing vendors whose data extraction technology could be incorporated through an SOA-connection [see *DIR* 10/21/05].

Of course, when dealing with standards, like SOA does, it's important that everyone agrees on the protocols. Lord recommends looking to organizations such as **OASIS** (<http://www.oasisopen.org>) for guidance. "OASIS is assembling blueprints for specific vertical implementations of SOA," said Lord. "For example, if you want to participate with Wal-Mart in retail-centric SOA transactions, there are certain XML tags and SOA protocols you must use."

In addition to SOA flavors being built around vertical markets, large vendors like **Microsoft**, **Oracle**, **IBM**, and **SAP** are trying to promote their own flavors of SOA. "Microsoft is trying to drive everything toward its .NET and C# programming languages, while IBM is pushing everything toward *WebSphere*; Oracle has its own programming language, and SAP is pushing its NetWeaver environment," said Porter-Roth. "The good thing is that if SOA is implemented correctly, it shouldn't matter if you are running an IBM *Content Manager* application and a Westbrook document management system—they should be able to talk to each other."

Lord admitted there is little immediate advantage gained by implementing an SOA application in a non-SOA world. "To take full advantage of an SOA-based ECM system, your line-of-business or ERP application should be SOA-based as well," he told *DIR*. "However, that's just a matter of time. The advantages of SOA are too great to be ignored."

As it relates to ECM, SOA should mean reduced operating costs for end users. "We are targeting the reduction of transaction processing costs," said Lord. "If you look at the financial reports of any public company, G&A costs are somewhere around 10-12% of expenditures. Accounts payable makes up at least a couple of percentage points alone—not to mention the increased burden that Sarbanes-Oxley is putting on accounting departments. By automating transaction processing leveraging an SOA-based document management system, users are eliminating a significant portion of their AP costs—which goes straight to the bottom line. They are driving down corporate head counts."

Lord has put his money where his mouth is by heavily funding Westbrook's SOA development initiative. When Lord arrived at the Branford, CT-based software developer as COO in the fall of 2003,

the company was in the midst of rewriting its flagship *Fortis* application to run on a .NET platform. Early this year, Lord left the company briefly to take a position as the CEO of Web Services library specialist **Infravio** (<http://www.infravio.com>). When he returned as Westbrook's CEO this summer, Lord had seen the light.

"Just because a product is based on .NET, does not mean it is built fully around SOA standards," explained Lord. "We wrote off more than \$1 million in R&D and started over again. As a result, we are going to have one of the only products in the industry truly built around SOA from the ground up. I think we need to come up with a term to differentiate between true SOA architected products and those that are legacy applications with an SOA wrapper."

Westbrook plans to release its SOA product for beta testing in the first quarter of 2006. "We expect it to open up some new channels for us," Lord told *DIR*. "Currently, our partners tend to be smaller VARs and office equipment providers, and we don't compete very often with big players like **FileNet** and **Documentum**. A Web services based product should make us more attractive to the large systems integrators that are moving their practices toward SOA implementations.

"Because of the ease-of-integration of SOA products, we also see potential for increasing our OEM business. **Hyland**, for example, generates a significant amount of revenue from an OEM relationship with **Fiserv**. I'm not saying we expect to unseat Hyland with Fiserv, but there are plenty of other opportunities with large technology systems vendors that could potentially leverage SOA protocols to add document management to their offerings."

For more information: <http://www.westbrooktech.com>

Patented Technology Covers Distributed Data Entry

As we all know, there is more to a capture application than making electronic images of documents. So, shouldn't the transition to distributed capture involve more than the creation of remote scanning stations? How about distributing the data capture process? **Orbograph** thinks it has come up with a unique approach to this and describes it in U.S. Patent Number 6,931,589—entitled "distributed document processing."

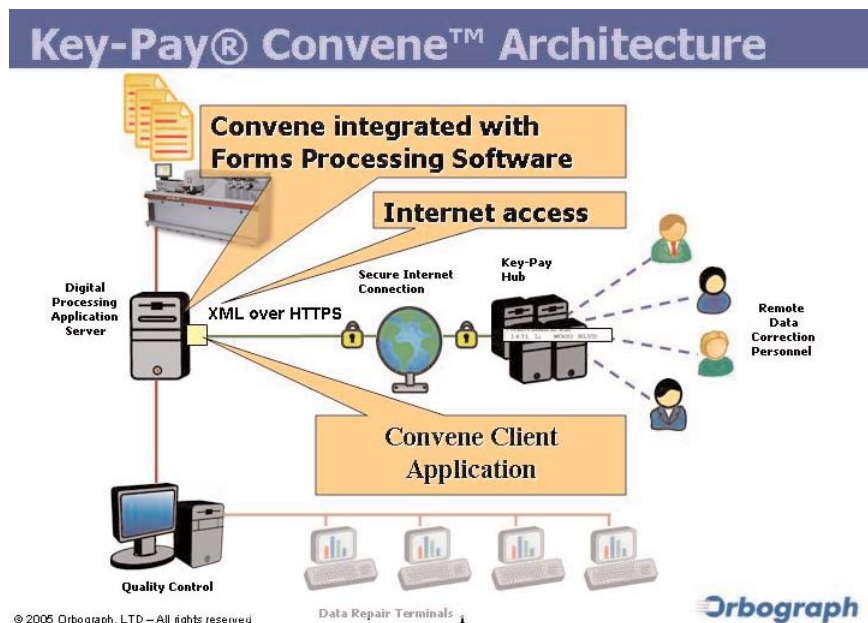
Yes, we realize the concept of capturing a document at one site and sending it overseas for inexpensive data entry is not new or innovative. Orbograph's patented technology, however, is designed to take this process to the next level. It manages the distribution of snippets taken from imaged documents.

"Our technology takes distributed capture from a one-to-one process to a many-to-many process," explained Roni Boker, product manager for Orbograph's *Key-Pay Convene*. "We realize there are other applications that enable distribution of snippets for data entry. However, they all involve transmission from one batch of documents to one set of end users. Our technology offers tighter control and enables multiple batches of images to leverage multiple keyers on the other end."

Orbograph's patented technology has been enveloped in an application called *Key-Pay*. *Key-Pay* has the ability to match specific data entry personnel with specific snippets of information. "Key-Pay can automatically train users on specific types of data," said Boker. "After a trained user logs on, our system automatically goes into the queue and matches him or her up with the appropriate snippets. The system knows which keyers are authorized to handle which snippets. *Key-Pay* also features a monitoring system, a performance tracking system, and a 'cop,' which eliminates 'problematic' performers."

The snippet model was designed specifically to work in a hosted environment. The *Key-Pay* hub [which is hosted in an **MCI** data center] receives and distributes the snippets. Applications developed by Orbograph capture partners feed the snippets to *Key-Pay*.

As the product name suggests, most of Orbograph's success to date has come in the check processing world, where its handprint recognition technology is already in widespread use. "So far, we've integrated *Key-Pay* into check capture solutions from partners like **AFS, Wausau, Jack Henry, and Fidelity,**" said Boker. "We are currently processing about 1.5 million snippets per day through *Key-Pay* for about 60 different end users. Most are banks or service bureaus that have their own keying personnel who log into *Key-Pay*."



Working with organizations like **Goodwill Industries** and **Service Source**, we've also trained our own data entry personnel that customers can leverage."

Orbograph is starting to expand into the forms processing market with a product called *Key-Pay Convene*. The concept behind *Convene* is the same, but instead of working with check snippets, snippets are taken from forms where automatic recognition is a problem. "We are currently evaluating two beta sites, one with **Captiva** and one with **IBM,**" said Boker. "We have also had talks with companies like **Fairfax Imaging, ReadSoft, and Parascript.**"

Integration with *Key-Pay* is accomplished through an SDK. "It took us about a week to make our full functionality available to IBM's forms processing app," said Boker. "We used their APIs to build a bridge. If APIs are not available, there are various other ways to connect utilizing XML and file servers."

To make it easier for end users to access snippets on the hosted *Key-Pay* server, in the latest version, Orbograph has leveraged standard protocols like MIME objects, XML, and HTTPS. "This has removed some problems with firewalls that we've had in the past," said Boker.

Orbograph's pricing model for *Key-Pay* varies. "Typically for forms processing clients, we have talked about a per character model," said Boker.

Boker concluded by saying that the recent press release, which announced the patent award, was primarily a marketing maneuver. "It has taken us more than three years to get approval for this

patent,” Boker told *DIR*. “When we filed for it, we did not have in mind that others were doing the same thing. Most importantly, we want the world to know we have this technology, and we own it. We hope this announcement will help us further penetrate the market. At the same time, we want to have some protection for what we have.”

For more information:

<http://www.orbograph.com/content.aspx?pageId=22;>

[Sandy.Leavenworth@orbograph.com;](mailto:Sandy.Leavenworth@orbograph.com)

PH (978) 448-8484.

Top Mortgage Lenders Rely On SOURCECORP

SOURCECORP has found a niche in the mortgage banking market. The Dallas-based document services specialist is currently imaging between four and five million pages per day to be stored in its *FASTRIEVE*™ document repository. According to Jim Gerken, general manager of the *FASTRIEVE* business, the majority of those documents are related to mortgages.

Gerken credits (no pun intended) rising home loan interest rates as changing the dynamic of mortgage banking. “When interest rates were dropping, lenders were writing so many mortgages and making so much money, they were able to continually throw more labor at the approval process,” Gerken told *DIR*. “About eight to 10 months ago, the market turned around. It has become more important for mortgage lenders to reduce costs, and going paperless represents a way to do that.”

FASTRIEVE was originally launched in 2001 to service a single mortgage banking customer. [Due to the competitive nature of their business, mortgage bankers are reluctant to have their names used in print.] “We were doing microfilming for this customer and offered to transition them to electronic imaging without increasing their costs,” said Gerken. “They had tried a couple internal imaging installations that didn’t work out. We built our *FASTRIEVE* repository and workflow system around their requirements, and it has mushroomed from there. We now store all sorts of documents related to areas like human resources, accounts payable, insurance claims, and medical records, but the majority are mortgage documents.”

According to Gerken, *FASTRIEVE* services eight of the top 20 lenders in the United States, in addition to companies from several other vertical industries.

“In the past six months, our *FASTRIEVE* volume has grown 25%, and [including backup files] we are managing more than 140 terabytes of data for our mortgage customers.”

One of **SOURCECORP**’s major competitive advantages is its turnaround time. “We can scan and post loan documentation within 12 hours; I believe our competitors typically offer a 72-hour turnaround,” said Gerken. “We have implemented a workflow process for one customer that enables it to make a funding decision in four hours. This customer believes it has the fastest service in the industry.”

Why is speed so important? “There are thousands of smaller brokers who send out loans for funding by the larger lenders,” said Gerken. “Often times, these brokers will submit a package of loans to four or five mortgage companies and end up selling to whoever responds first. We have developed workflow processes that expedite our customers’ decisions on whether to fund a loan or not.”

According to Gerken, a typical loan file is about 300 pages long with 25-50 different document types or sections. **SOURCECORP**’s largest processing center, in the Los Angeles area, uses **IBML** ImageTracs to capture the images.

Because of its large number of mortgage-related customers, **SOURCECORP** has been able to design some unique integrated processes. “We have one customer whose job is to make sure there are no liens on properties before loans are funded,” said Gerken. “So, when one of our mortgage customers is using the service, we can just extricate the appropriate links and send them to our first customer.”

SOURCECORP can also facilitate image exchanges between mortgage companies. “One of our services involves classifying images into a format suitable for the receiving company,” he said. “For example, one lender might break up a mortgage into 55 document classes, while another might only use 42. Of course, we also print 60,000 pages per night from our repository.”

This enormous print volume is just one sign that paper mortgages are not going away. “We accept electronic mortgages, but the only ones we receive have been scanned by somebody else,” said Gerken. “We have no customers doing paperless mortgages. At least one customer has told us they don’t see paper mortgages going away any time in their lifetime.

“Because of our advanced workflow services, our

business is not totally dependent on paper. That being said, we are pretty confident there will continue to be paper mortgages for a long time.”

For more information:

<http://www.sourcecorp.com/SOURCECORP/Industries/Financial/Finance+Mortgage.htm>

Comfidex Gearing Up For 2006

Like many registered voters, VoteFiler sits out 2005 elections

The first Tuesday after the first Monday in November has come and gone, and **Comfidex** is biding its time. Comfidex is the Manhattan-based document imaging company that introduced *VoteFiler* at the **FOSE** show this April. *VoteFiler* is an electronic-based voting system that offers a paper trail in addition to providing potentially significant cost benefits over popular touchscreen-driven voting systems [see *DIR* 4/22/05].

Comfidex hopes to have *VoteFiler* ready to hit the streets during the first quarter of next year, in time to be installed for the 2006 elections. “We realize there was a lot of federal money given out this year, through the Help America Vote Act, to counties for upgrades to voting systems,” said John Philpit, CTO of Comfidex. “But, the timing just didn’t work out for us to get fully involved in the process. We’re hoping there’s some money left over next year. Remember, the cost of our system is so low, that we even view it as a potential add-on that can be used to create voting express lanes.”

VoteFiler relies on bar-coded documents presented in person at the polling station by the voter. The bar code is scanned and married to data on a server that has been entered previously through an electronic form. (Comfidex holds patents on this procedure.) The encrypted and secure data is only counted as a vote after the bar code is scanned. The document serves as the paper trail.

“The demand for a paper trail continues to increase,” said Philpit. “Many states have enacted their own requirements. Now the national Election Assistance Committee is calling for a paper trail, which should be required for the next presidential election in 2008.”

Philpit reiterated Comfidex’s contention that the practice of attaching a printer to a touch screen system does not constitute a proper, secure paper trail. “In addition, *VoteFiler* uses commercial off-the-shelf hardware, which makes it significantly less

expensive to install and maintain than a touch screen system—especially one with a printer added on,” said Philpit.

In addition to political elections, Comfidex is marketing *VoteFiler* for corporate and organizational elections. “In October, we used our system to tabulate an election for the Pennsylvania Chapter of the **NAACP**,” Philpit told *DIR*. “They got their results within five minutes of the last vote being cast. Because our application is not quite ready for the street, we brought in a turnkey system and ran it for them.”

Comfidex’s pricing model is based on an initial installation fee and then a cost per election based on the number of voters. Comfidex hosts the *VoteFiler* server and provides its customers with access through a Web interface. There is also a LAN-based option for customers that don’t want votes traveling over the Internet.

PAPER-HANDLING, WARRANTY, UPGRADED ON 5x20s

In addition to the improved card-feeding and double-feed detection capabilities in the new **FCPA** fi-5x20C models that we mentioned last issue, FCPA product manager Kevin Neal pointed out a couple more improvements. “We’ve also improved the paper handling,” he told *DIR* in an interview last week. “We incorporated a more efficient ADF, which has improved paper separation. We’ve also improved our 300 dpi speeds [to 24 ppm/45 ipm in black-and-white and 24 /24 in color].”

Neal pointed out that FCPA continues to offer the only workgroup scanner that features a straight paper-feeding path and feeds paper through the same path as cards. “To date, our card scanning capabilities have been primarily used in the healthcare industry,” said Neal. “I wouldn’t say a huge percentage are utilized for card scanning, maybe 5-10%, but we have some very large customers. We are also looking at expanding further into the retail/point of sale market, in niches such as automotive rentals. Instead of making paper copies of licenses and credit cards, we see imaging as a more efficient alternative.”

FCPA has upgraded the maintenance package included with the new workgroup line. The company has introduced new drivers that will allow fi-4x20C users to install 5x20s without changing their application. FCPA has also added an advanced exchange option to its standard one-year warranty for the 5x20s. “We have some resources in the areas of service that we wanted to use to our advantage,” said Neal.

<http://www.fcpa.com/products/scanners/>

Philpit said Comfidex hopes to have federal certification for *VoteFiler* well before the 2006 elections. "Of course, with all the red tape you have in the United States, our initial installation might come overseas," he said.

We concluded by asking if Comfidex is still looking for a large integration partner "We view a potential partnership more as an opportunity than a need," Philpit said. "If we could find the right company among the current voting system vendors, it might be a good fit. These vendors already have channels set up to sell to all 51 states. Additional funding for our initial marketing costs would also be helpful. At the same time, we think there are ways to tackle the market ourselves."

For more information: <http://www.comfidex.com>

Medicare PDP Represents Challenge, Opportunity

You've probably seen the stories in the news. Currently, more than 29 million senior citizens are scrambling to get enrolled in Medicare's Part D Prescription Drug Plan (PDP). This is the controversial new program being administered through private health insurance providers. While there has been plenty of ink spilled about the headaches senior citizens are getting trying to navigate the intricacies of choosing the right provider, we've haven't seen much addressing the challenges insurance companies are facing processing all their enrollment forms.

Obviously, the trials of grandma and grandpa make better mass media copy than the struggles of multi-billion dollar insurance firms—but this is *DIR*, where managing paper problems is priority number one. Fortunately, vendors like **Softheon** share this view. The Long Island-based software developer recently released an application designed specifically for dealing with Medicare's Part D PDP. "This is a brand new challenge for health insurers," said Chuck Strahlendorff, EVP at Softheon. "We view it as a tremendous opportunity. This is the first time we will be competing with more established vendors from day one."

Softheon is a high-volume imaging and workflow specialist that has been cultivating a growing business in the health insurance market. Softheon's initial foray into this market was profiled in *DIR* three years ago [see *DIR* 8/2/02]. Since that time, Softheon has added nine more health insurers to its customer list. "We have acquired some pretty meaty customers over the past two years," said

Strahlendorff. "We have established a strong footprint that we can leverage for growth."

To engineer its Part D PDP solution, Softheon has indeed worked very closely with some of its largest customers. "Almost every insurance carrier has a Part D PDP offering, but only 10 are offering national coverage," said Lisa Wenninger, solutions architect for Softheon. "Two of those 10 are our customers."

Part of the challenge of Part D PDP is the time frame. Current Medicare beneficiaries have between Nov. 15, 2005 and May 15, 2006 to enroll in a program to be eligible for benefits in 2006. "We consider our enrollment processing solution to be phase one of a complete Part D PDP package," said Wenninger. "We've already implemented it at some sites.

"Basically, our system takes control of whatever form of enrollment content comes in—although there has been some talk of using Web forms, mostly it has been paper. Our application is used to capture information from those enrollments and facilitate submissions to the **CMS** [Centers for Medicare & Medicaid Services]. CMS responds with rejections, or approvals and terms of service. Based on CMS' responses, our system can help manage the insurer's responses to the senior citizens."

Once a person is enrolled, insurers have the responsibility of administering the plan. "We already have experience managing privately funded PDPs," said Wenninger. "There are a number of similarities, but also a number of different rules we are addressing."

Softheon's Part D PDP solution represents the latest in a series of template-based applications the company has released targeting different aspects of the health insurance business. Other processes addressed include claims pre- and post-adjudication, grievance and appeals, and underwriting. "You can purchase these applications as part of our total document imaging/workflow repository package, or as standalone systems to hook into existing infrastructures," said Strahlendorff. "When we released our grievance and appeals package earlier this year, for example, we immediately got two dozen responses from interested health insurance companies. Point solutions like this give us an opportunity to get our foot in the door at some of the larger players.

"We have put almost all our focus into the health insurance market and in the next 12 to 24 months expect to realize very strong and aggressive growth."

For more information: <http://www.softheon.com>

Archive Founder Discusses Book Scanning

For a peek into what the **Open Content Alliance's** (OCA) book scanning project could look like when all is said and done, go to <http://www.openlibrary.org>. The site features images of books displayed to look like bound copies. There are also downloadable PDF and DjVu versions of the books, as well as a listening option (which when it's enabled, we presume will contain an audio file). There is also a link to a company that will print and bind a copy of a book and mail it to you.

"There are all of about 14 books up there right now, but that should give you some idea of where we're headed," explained Brewster Kahle, the Internet and search pioneer, entrepreneur, as well as founder and digital librarian of the non-profit **Internet Archive**. "The aim of the OCA project is to create something useful to both scholars and the general public. This means we have to have a coherent collection with appropriate quality and a good selection."

According to Kahle, the OCA project is designed to integrate with the core mission of traditional libraries. "Librarians are really driving this project," he said. "It requires their help and supervision."

Only after discussing the big picture of the project, did Kahle seem comfortable delving into some of the technical details. He acknowledged that both PDF and DjVu, as well as JPEG 2000, are playing roles. He also said the Archive is keeping an uncompressed copy of each image.

Kahle indicated that the OCA has decided to go primarily with a camera system utilizing manual page turners. The system is known as the Scribe, and the *San Francisco Chronicle* described it as looking like a "portable darkroom." "Inside, two digital cameras peer down on a book held in a V-shaped glass cradle." According to an article published by *DevX News*, the Scribe uses 16 megapixel cameras that capture images at 500 dpi.

Scribes are being installed on-site at the libraries where the books are being captured. "We are trying to achieve some very high resolution color images at a cost of 10 cents per page, including labor," said Kahle. "That's a really tough set of constraints, and we've found the Scribe has been the best way to approach it."

Incidentally, the Internet Archive is looking for people to operate the Scribes (<http://www.archive.org/about/textsjobs.php#bookScanner>). The first two points listed under the "skills" section of the job posting are

- 1) High tolerance for repetitive tasks.
- 2) No, really. High tolerance for repetitive tasks.

For more information:
<http://www.opencontentalliance.org/>;
<http://www.archive.org/details/texts>

VISIT OUR BLOG

Remember to visit our blog at <http://documentimagingreport.blogspot.com/>. Actually, do more than visit. Please feel free to post your own comments and contributions. You can even post anonymously if you wish.

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

Mass Storage News

Opportunities and Trends in Data Storage and Retrieval

SCAN: The DATA CAPTURE Report

Premier Management & Marketing Newsletter of Automatic Data Capture

1 year (24 issues)

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

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

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

OUR GUARANTEE TO YOU

If you are not completely satisfied, we will refund your subscription cost for all remaining unserved issues.

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.) _____