IBM Software

The digital mail room is dead

by Scott Blau, world wide director of ECM capture solutions

IBM

Executive summary

Once there was a vision of the "paperless office," made possible by the "digital mailroom," where all paper arriving in the mailroom would be immediately scanned, identified and routed as an image or electronic file to the appropriate employee or department for processing. No more filing cabinets, no more interoffice couriers wheeling wagons of mail around, no more riffling through manila folders looking for missing paperwork.

The foundation for this next generation document strategy would be gleaming high-tech imaging centers with scanners churning through stacks of incoming documents. It was supposed to transform government agencies and corporate offices worldwide from populations of paper-lugging Neanderthals to digital knowledge workers who would have access to all data and documents that they needed—when they needed them—at the touch of a button.

This new era was supposed to arrive beginning around 1995. But something must have come up, because the schedule slipped. Then we had the distractions of Y2K, so we lost sight for a while until the anthrax scare of late 2001, where U.S. media and government figures received envelopes of the highly lethal chemical. Suddenly, the digital mail room was back and bigger than ever, with a new security twist. Every year since, the importance of a digital mail room has been an annual prediction among analysts and prognosticators around the globe. But guess what? While the quest for the paperless office lives on, the digital mail room quietly became obsolete.

Why?

Documents no longer arrive at the office solely via mail or courier. As a result, there is no longer a single, controllable portal for documents to enter an organization, and it is upsetting long-standing document strategies. Fax was the first thing that began to chip away at the old institutions. Most organizations put the fax machine in the mailroom, so you could still centralize incoming paper. As fax networks grew, the fax server provided an electronic way to receive faxes and route them to secure mailboxes and still control the flow of documents, which, after all, often contain the business transactions of an organization.

But it was the Internet that provided new channels for incoming documents and data. Slowly and steadily, web-based electronic forms and email have increased to the point where now email has become a major source of incoming documents, especially the electronic documents sent as attachments. Studies show that volumes of paper documents are flattening (and even decreasing slightly) in offices today, while electronic documents and records are increasing rapidly.

While the paperless office remains a viable, if somewhat elusive, dream, the digital mail room will no longer deliver on its promise. What is needed now is a strategy to capture and transform all incoming documents, whether it's paper in the mailroom, faxes on a fax server, electronic forms, images on an FTP server, emails and electronic attachments, or electronic data interchange (EDI) and whatever media becomes the next conduit for business-related documents and data.

IBM Datacap Taskmaster Capture software is one of the only enterprise capture platforms built to operate as a universal capture portal, with document input solutions designed to seamlessly merge all forms of documents entering an organization. So it's time to say goodbye to the digital mail room and hello to the universal capture portal.

"It is time to concentrate on how mail, email, fax and other inputs can be brought together through a single funnel that automates the extraction of time-critical business information."

—Scott Blau, world wide director of ECM capture solutions

Document production is growing

A funny thing happened on the way to the digital mail room. The digital mailroom never materialized. Yes, there are isolated implementations represented as digital mailrooms, but, for the most part, things have not changed very much. If you stop to think about it, that is the big news: not much change in the volume of physical mail. Even the U.S. Postal Service is struggling to stay afloat as mail delivery (and stamp revenue!) declines.

Volume of Paper vs Electronic Documents

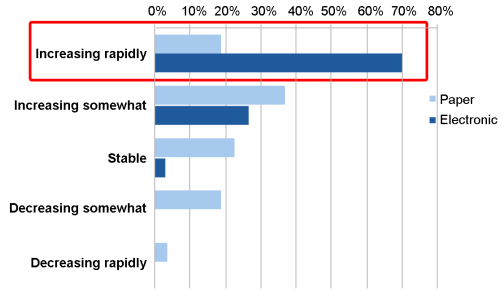

Figure 1: A recent survey, entitled "Electronic vs. Paper," by the Association for Image and Information Management (AIIM) queried 768 organizations on paper and electronic document production. The results reveal that the volume of electronic documents is increasingly rapidly among 70 percent of participants. The study suggests that paper records are still increasing in 56 percent of organizations

Where is paper going?

If the volume of mail isn't growing, then the value to an organization of digitizing that activity is going to fall. Where is it all going? The world economy, for all the dramatic ups and downs of the last decade, has grown considerably. So all that activity, all those application forms, those bills, those notices sent by consumers to businesses, by businesses to other businesses, and by businesses to consumers to fill out and return have not gone away. They've just quietly gone digital and skipped the mailroom altogether, because they are sitting in an email inbox.

Simple problems, complex solutions

Turning paper documents electronic is proven technology

Document capture has been transforming paper to digital form for more than 20 years, but make no mistake—it is a complicated process. Document scanners have evolved nicely. You can get one for a reasonable price that scans one full letter-size page per second. But it can jam, it needs cleaning, and the rollers need to be replaced. In the end, you have a finicky piece of equipment that locks down a full-time operator, which is an ongoing expense.

That's just where the challenges begin. Next, the images have to be grouped into documents, the documents identified, and any relevant data extracted, both for storage indexing and possibly for delivery to a line-of-business application.

Paper Document Capture Process

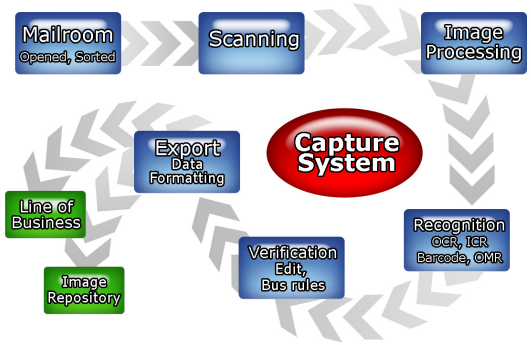

Figure 2: In a typical document capture system, paper arrives in the mailroom, where it is prepped for scanning. Images are processed to "enhance" image quality and identify page and document type, using a variety of different techniques. OCR, intelligent character recognition (ICR), optical mark recognition (OMR) and bar code recognition extract data, which is then verified using business rules and quality assurance operators viewing the document and data. Finally, images and data are formatted for export to line-of-business applications and image repositories for efficient storage and retrieval.

Still require people to complete

All of these technologies share one characteristic—they are imperfect. That means that someone has to be prepared to verify the results of recognition. The optical character recognition (OCR) "engines" are good enough, for the most

part, to know when they are unsure. That makes verification a little easier because the user can jump confidently from one low-confidence result to the next. But it still requires effort.

And only when all this effort is complete is the document ready to move from the capture phase into an organization's circulatory system, that is, a workflow.

EDI helps eliminates paper

EDI to the rescue?

Electronic data interchange is the holy grail of the digitized business. By sending an invoice electronically, a company can save US$10—12 per invoice (Source Aberdeen Group study, "A Comparison of Supplier Enablement around the World," 2008). It is not just the cost of the stamp that is being saved—that's almost irrelevant—it is the cost of some invoices getting lost (the post office has never been perfect in that regard), the cost of some being misfiled at the customer location, or, perhaps worse from an accounts receivable cost standpoint, the customer incorrectly or partially paying for some. If full payment is ever received on those invoices that go astray, a lot of money will have gone out in the calls and lost time. Even customer satisfaction can be affected.

EDI works both ways. As counterintuitive as it may seem (because it seems that some customers never want to pay their bills), those receiving the same invoice electronically save money as well. The savings are also primarily a result of productivity improvements. The accounts payable department can concentrate on maximizing cash and discounts instead of chasing errant invoices and placating upset vendors.

EDI did not solve all our problems either

EDI remains, for most realms of business activity, a distant goal, not a reality. Even where EDI has made significant inroads, typically under government mandate as with Medicare claims, non-EDI transactions (paper and other media) continue to circulate, if not flourish. When EDI and paper transactions exist in parallel, the expense of paper stands out, making a strong justification for automation through scanning and document capture.

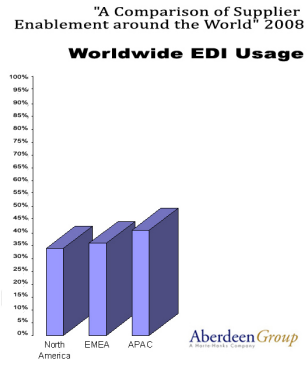

Figure 3: According to the 2008 Aberdeen report, "A Comparison of Supplier Enablement around the World," only 34% of purchase orders and invoices are transmitted electronically in North America. In EMEA, 36% are transmitted electronically and in APAC, 41% are transmitted electronically. There were many predictions that Internet-based EDI would increase reliance on these electronic networks, but the Aberdeen survey indicates that has not happened.

Email—the benefits of EDI without the overhead

Unlock the promise of EDI in your email

For most areas of business, email has quickly filled the gap left unfilled by EDI. Email is, at least superficially, electronic data interchange. It is definitely electronic. The only difference is that it's not as structured as EDI. Email is to EDI as a scribbled request for service from a customer is to a carefully typed service request form. Both may do the job, but it requires much less effort to accurately fulfill the service request form than a barely legible note written on a piece of paper.

Email, however, is what the digital mailroom looks like today. The challenge is to not only manage the flow of emails, filtering out endless streams of spam, but also identify and extract useful information from the email messages and attachments.

The promise of email capture is clear: a vendor sending an invoice via email means one less envelope to open, one less document to scan, recognize, and verify. That can translate into big savings and faster input.

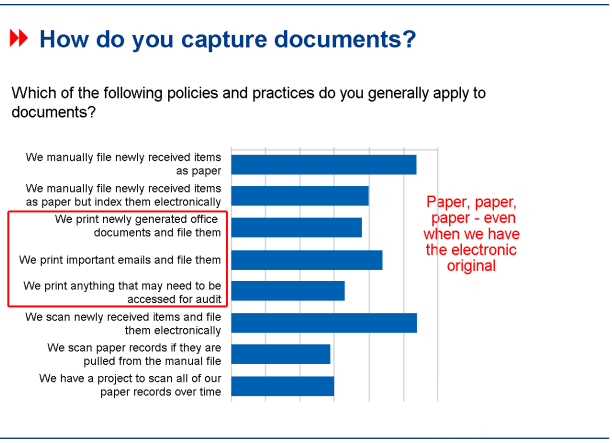

Figure 4: The AIIM survey, "Electronic vs. Paper," 2009, suggests that a significant number of organizations are printing out electronic documents to capture, process and store documents and records. Many are maintaining separate input systems for paper and electronic documents, creating more effort and more disconnection in their document strategy.

However, that is not a widespread practice today. Instead, organizations are printing their electronic documents so they can be scanned and processed with the same capture systems that their paper documents are going through (Source: AIIM Survey. "Paper Vs. Electronic" 2009). This is an unnecessary waste of time and paper.

How to combine the vision of the digital mail room with the new realities of the virtual mail room? You'll need a capture solution that can handle emails and all the things that can be attached to an email just like paper. It is what we call a universal capture portal.

The new digital mailroom—a universal capture portal

Feed emails into your document capture system

With a universal capture portal, an organization maintains a single input system for paper, fax and electronic documents. In one solution, you can give paper documents the flexibility and speed of electronic documents, and apply to electronic documents the benefits of automated document capture—data extraction, automated indexing and document identification.

Economy of scale

Perhaps most importantly, with a mult-modal capture system, you need only one set of processing rules for virtually all your incoming documents, whatever medium brings them into the organization: the postal service, fax, email, and so on.

Figure 5: A quick inventory of all incoming documents usually reveals a broad variety of formats: paper, fax, email, and email attachments, which can be image files such as PDFs, JPGs or TIFFs; Microsoft® Office files such as Microsoft Excel or Microsoft Word documents; or zipped files that can have any of the above. With the universal capture portal, virtually all formats can be fed into Datacap Taskmaster Capture quickly and securely.

Datacap Taskmaster Capture Connector for Email and Electronic Attachments

Breaking it down: What's in an email?

The challenge facing the universal capture portal is that as varied as paper documents are electronic documents are exponentially more variable. Take a second and open your email inbox. There are a bunch of emails that are of interest and, unless you have a great spam filter, maybe as many that are simply digital noise. But let's drill down a bit further into a "good" email.

Perhaps the body of a message has useful information, and/or an attachment or carries the important content. The attachment may be in any number of common formats (or some uncommon ones!), including (using their file extensions as shorthand): .doc (and .docx), .xls (and .xlsx), .pdf, .zip, .tif, .jpg, .png, .bmp, and even a .msg or .eml (emails!).

Opening an email and prepping for capture

An email message is essentially a container. It has some standard content—the address info, subject line and body—as well as attachments. The attachments are also often containers in their own right. For example, a ZIP archive contains another layer of documents. A "simple" Word file has multiple pages, but can also have its own independent content, embedded graphics and embedded documents in, again, various formats.

Although emails deliver information electronically, they are highly unpredictable. The universal capture portal must unpack the contents of an email to find and extract useful information. Datacap Taskmaster Capture Connector for Email and Electronic Attachments does this via the Datacap Rulerunner Service, using rules that methodically open up each level of the contents and apply the appropriate conversion so that all content in the end is rendered in the lowest common denominator format, for example as images.

The universal capture portal helps deliver control over documents

Image is everything

Even after conversion to image format for processing in Datacap Taskmaster, the native format documents that flow into the universal capture portal are maintained in parallel.

Figure 6: The AIIM survey, "Electronic vs. Paper," 2009 suggests that a significant number of organizations are printing out electronic documents to capture, process and store documents and records. Many are maintaining separate input systems for paper and electronic documents, creating more effort and more disconnection in their document strategy.

Typically, one or both (native and image) formats are sent to the repository for storage after index information has successfully been extracted.

A big advantage of having images, even temporarily, for electronically received documents is that then you can use the same rules that are applied to scanned paper documents. The universal capture portal is a gateway into your organization. It is also a bridge between the paper and electronic worlds. It is what the digital mailroom is shaping up to be.

There are a number of reasons to create images out of electronic documents:

- It is easier to display documents onscreen, even in a browser without a special plug-in.
- Extraction technologies can be used to locate data, practically regardless of the underlying (and evolving) native format (for example Word docs in .doc or .docx format).
- It is more convenient to reassemble in any way appropriate to the application, again ignoring the limitations of the original format.

Forget about performing magic in the old, paper mailroom. It is time to concentrate on how mail, email, fax, and other inputs can be brought together through a single funnel that automates the extraction of time-critical business information to help your organization operate more efficiently, improve customer satisfaction, and allow you to adjust to future, and often unexpected, shifts in how people take advantage of digital media.

About Enterprise Content Management

To learn more about IBM Enterprise Content Management Solutions, contact your IBM sales representative or visit:
ibm.com/software/data/content-management

About Datacap Inc., an IBM Company

Datacap provides award-winning enterprise content management (ECM) software for document capture and paper forms processing. These solutions help organizations worldwide improve the speed, accuracy and cost-efficiency of data extraction, making information more accessible. Using the latest technologies and innovations, Datacap solutions help automate a broad range of document-centric business processes in a wide variety of industries.

For more information about Datacap solutions, contact your IBM sales representative or IBM Business Partner, or visit
www.datacap.com

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
March 2011

Datacap
an IBM® Company

IMW14545USEN-00