SearchContentManagement.com

E-Book

The new best practices for document imaging, OCR and scanning

Successful document management begins with efficient and well-managed document imaging, scanning and capture processes. Without a sound approach to document imaging and scanning, for example, a document management system is likely to end up being a source of frustration for business users looking to find important information in corporate documents. This eBook, intended for document management and IT professionals, examines the latest trends in document imaging and scanning as well as optical character recognition (OCR) and provides expert advice on best practices in those areas. It also explores common document imaging challenges and offers a checklist of action items plus tips on key mistakes to avoid.

Sponsored By:

NUANCE

SearchContentManagement.com
E-Book

The new best practices for document imaging, OCR and scanning

Table of Contents

NUANCE

Document imaging systems mature, offer expanded options to users

By Catherine LaCroix, SearchContentManagement.com Contributor

The basic functions of document imaging, scanning and capture technologies are not new. But the form and frequency with which imaging and scanning tools are being applied today is providing both large and small organizations – and those in between – with increased options to automate costly manual tasks as part of their document management strategies.

The strongest drivers for using imaging, scanning and capture technology are improved search capability and knowledge sharing across the enterprise, according to a study released last year by the industry group AIIM, titled *Capture and Business Process: Drivers and Experiences of Content Driven Processes*. Those critical functionalities are followed by additional drivers such as accelerated productivity, reduced costs and improved customer service, AIIM said.

The economy, unsurprisingly, has had a big impact on accelerating the adoption of document imaging systems and scanning equipment as a means of reducing labor and operating costs. Businesses continue to be pressed to cut costs while remaining profitable. "One way to do that is by automating processes," said Ann Valaitis, associate director of image scanning trends for InfoTrends, a market research and consulting firm based in Weymouth, Mass.

Applying new document imaging and scanning technologies to internal workflows has resulted in reduced cycle times and improved processes for many companies, Valaitis added.

One potential path to achieving that kind of modernization is via distributed imaging and scanning. "If we look at the product cycle of scanners over the last few years, scanning was once very centralized," said Valaitis. Companies would often have a production scanning environment using machines installed in a central location. Business users from different departments would bundle scanning tasks and then deposit the scanned documents in some

kind of archival storage system. According to Valaitis, medium- to high-volume scanning systems best fit that type of application.

Cost of document imaging systems drops as quality improves

In the last few years, however, there has been a boom in sales of lower-volume scanners that can be distributed throughout the enterprise for various document capture purposes. That means a drop in prices, "but the features are improving" at the same time, Valaitis said.

Newer scanners offer improved image quality and color, good optical character recognition (OCR), image rotation, duplex capabilities for double-sided scanning and many of the other features that were once the exclusive domain of production-class document scanning systems. And, Valaitis said, "distributed capture can be much more far-reaching in an organization in terms of capturing various types of content across the board."

Although the cost of scanners has come down and their use has become more widespread, some analysts question where the value-add is in deploying them without a comprehensive document capture and document imaging management strategy.

"You can buy a document scanner for $400, and some think that's all they need in order to do document conversion, but when you get into it, you realize that the difference between that and any sort of enterprise capture technology is huge," said Chris Riley, senior enterprise content management and document capture architect at consulting firm ShareSquared Inc. in Pasadena, Calif.

To go beyond the basic conversion and storage process, you have to engage the entire document workflow, according to Riley. The real value, he said, is in taking the information on scanned documents and making it available as quickly as possible to those people within an organization who can use it. To do that, document imaging and capture must take place at the point of or very close to the origination of data. Then, all the information should be made easily searchable via metadata.

Such setups aren't so simple to implement, however, and some businesses struggle to reach that point. "Enterprise capture software costs hundreds of thousands [of dollars] and requires a minimum of three months to get up and running," said Riley. "There's a lot of rebellion from the user space because of that. It's harder than people want it to be."

Not so basic: the benefits of document imaging systems

Valaitis is more optimistic about small-scale implementations. She noted that the majority of individual users are already scanning documents to email and marrying that process with their own workflows, enabling them to share the documents with other knowledge workers.

"The percentage of users who scan to email is over 70%," Valaitis said, adding that the simplicity of doing so has helped the practice proliferate. But now she sees a desire in many organizations to move beyond simple document imaging and scanning and basic collaboration.

"We're at the point where we need to pull content off those [scanned] pages and we need to make it actionable," Valaitis said. "We need to be able to search off that content; it needs to be intelligent and more sophisticated. The software is there now that can do that, the hardware is there that can map and marry with the software – it can all be done."

It's up to document imaging and management professionals to demonstrate all of that to business executives and end users who can benefit, she said, "and show them how they can really get an ROI" from investments in document imaging software and scanning equipment.

About the author:
Catherine LaCroix is a freelance writer based in Portland, Ore. She covers technology used in business, education and health care.

Tips on deploying document imaging and document scanning systems

By Catherine LaCroix, SearchContentManagement.com Contributor

Beyond static document imaging, scanning and capture lies the realm of automating systems and processes for the best use of information and more efficient business practices. By standardizing the processing of forms such as invoices, for example, an organization can automate the sending of reminders about late payments, making it more efficient in completing mundane tasks that help improve the bottom line.

Although the truly paperless office might remain a myth, modernizing systems to read invoices, bills of lading and even resumes via the use of document imaging software and scanning equipment can move a company closer to that ideal.

But getting there can present certain challenges. "This is not install once and walk away," said Alan Pelz-Sharpe, a principal analyst and director at consulting firm Real Story Group in Olney, Md. "First, [you need to] ensure that you have the technical skills to configure and maintain a powerful capture system." Then, he said, it's critical to introduce procedures and policies defining standard practices so they become the rule internally.

A key element of setting up effective document imaging and document scanning systems is determining the image and indexing quality needed to best serve organizational needs. "Where companies fall down is they either don't index their information correctly or the people who are doing the indexing don't choose the right information," said Alan Weintraub, a principal analyst at Forrester Research Inc. in Cambridge, Mass.

Great expectations for document imaging and document scanning systems

High expectations for imaging and capture accuracy also need to be part of the plan. "You should go beyond [optical character recognition] and expect very high accuracy levels for capture, 97% or more," said Pelz-Sharpe. "I find too many organizations fail to utilize forms recognition, for example, and put up with chronically poor accuracy levels." If your imaging,

scanning and capture system isn't delivering higher than 90% accuracy, he added, there likely is something wrong.

In addition, companies "need to have checks and balances in both reviewing the document from a quality perspective and to be sure that they have the right indexing values in there so the document can be retrieved later," Pelz-Sharpe said.

There are multiple ways of doing this as part of a document imaging management strategy. One is to have a quality check on indexing, via a workflow that is kicked off when you scan a document. "I grab information off that document and key it into the metadata fields that I've pre-defined, and the next step might be a secondary index, so there's a verification step there," Pelz-Sharpe said. "And once I'm done with that, I release that document to the document management system for processing."

But the best way to ensure that scanned document images are fully available once they're in the system is to plan for standardized tagging, categorizing and indexing as part of the implementation right from the start, according to Pelz-Sharpe. By definition, this varies enormously from organization to organization, but he recommends auto-tagging at the capture stage.

"Keep it simple: have an agreed set of metadata that you want to capture and automate that process as much as you possibly can," he said. If an automated system captures 80% of the metadata that it was intended to find, that's probably 70% better than would have been achieved if the process was left to a manual approach.

The outsourcing option on document imaging and scanning

Another way companies can ensure that proper implementation of document imaging systems and scanning and capture technologies takes place is to leave it to someone else. "Too often, a lot of money and time is spent on a workflow or document management system, and far too little on the accompanying capture project," said Pelz-Sharpe. "This is specialist stuff and can deliver huge efficiencies, so involve experts and ensure you keep them involved to some degree throughout."

Weintraub agrees. "When you do have a lot of paper, you don't build the scanning function in-house, you outsource it," he said. For example, Weintraub added, a financial services firm or a mortgage processing company might outsource all document imaging and scanning operations to a third party that would also do some basic indexing derived from certain metadata. Then all that information would be shipped to the company and imported into its document management system for processing.

The primary reason for outsourcing document imaging and capture processes among companies that have done so is to cut costs, according to a 2010 study conducted by the industry group AIIM and titled *Capture and Business Process: Drivers and Experiences of Content Driven Processes*.

As for selecting document imaging technology, Pelz-Sharpe recommends that companies undertake a thorough procurement process involving product demonstrations and detailed evaluations. "Don't just go for the so-called market leader," he said. "Though in principal these software options all do the same thing, in reality they all do the same thing differently. You need to find the right fit for your needs."

About the author:
Catherine LaCroix is a freelance writer based in Portland, Ore. She covers technology used in business, education and health care.

Document imaging management missteps: Problems and pitfalls to avoid

By Steve Weissman, SearchContentManagement.com Contributor

It really is magic: push the paper through the scanner and watch it appear on the screen!

Document imaging actually can be that easy when the process is properly planned for and executed. But document imaging management initiatives are often not as carefully planned or executed as they should be, for reasons that are – or should be – readily avoidable.

Perhaps the most basic problem is that people can get so excited about the electronic path they've embarked upon that they skip right over certain key fundamentals when they look at improving their document management strategies. As a result, they run into trouble and often find that the cost of getting back on track proves to be substantial.

In fact, it has been estimated that 40% of all document imaging and scanning work is rework. That's a significant proportion when you consider the efforts that go into deploying document imaging systems, so it's well worth avoiding the major pitfalls in the first place.

Here are some of the common mistakes that can derail implementations of document imaging software and scanning equipment:

Poor or absent document naming conventions. One of the most significant issues might be the simplest to understand and correct. "The biggest sin is failure to name the document properly," said imaging expert Arthur Gingrande, a Lexington, Mass.-based partner with Imerge Consulting. "People tend to use any handy name that comes to mind instead of one that logically locates the document so it can be easily found later."

This particular pitfall speaks to so elementary an issue that you don't even need an imaging system to be tripped up by it. But automating the scanning process can go far toward avoiding it by applying naming conventions to documents according to their function, department or type.

For instance, why not set things up so incoming forms – for vacation requests, say – are automatically given names in the format of "vacrequest-username-datereceived," with the *username* and *date received* information pulled straight from the system? This is the kind of tagging computers are good at, and yet the result is still readily understood by humans – and it's further applicable to documents that live outside the system.

Unrealistic expectations. Having unrealistic expectations for optical character recognition (OCR) is another frequent issue within organizations. In truth, there is no magic involved in the prototypical next step in the imaging process. "People still think it's 100% accurate, and it's not," said Ralph Gammon, editor and publisher of the *Document Imaging Report* newsletter and website. "But some people then junk it, and that only makes the mistake worse."

Extracting text via the scanning of documents can be anywhere from 85% to 95% accurate, depending on how much, if any, fine-tuning the OCR system has received. However, those accuracy rates are far higher than what is typically achieved by humans retyping content to get documents into an electronically searchable database, so scrapping a system simply because it doesn't get everything right is short-sighted.

Failure to define needs. It may be that the biggest pitfall to avoid is rushing out and buying a document imaging and scanning system – or hiring a scanning service bureau – without first figuring out what the company's needs are. Knowing how many documents need to be "electronified," what kinds, paper sizes, whether they're multicolored and who's responsible for managing them – those are just some of the critical issues that should be factored into a technology purchase decision. Overlooking company needs could lead to catastrophic outcomes, financially as well as operationally.

Say, for instance, your documents contain text that is highlighted in places. A black-and-white scan will likely turn the highlighted areas black, and thus make them unreadable, while using gray scale should leave them readable and also require much less storage space than a full-color scan would. These are the types of scenarios that need to be sorted out well before any money is spent to ensure that it is being put toward capabilities that actually meet business requirements.

Forgetting the physical parts of the process. The corollary to the previous point is the tendency to forget that much of the work associated with document imaging is still quite physical. Paper will need to be prepared before scanning by unfolding it where necessary, removing staples and paper clips, smoothing badly wrinkled pages, etc. The good news is that this step is virtually impossible to skip and that it takes only a matter of minutes to perform them at the scanner. The "gotcha" is that the process takes longer – and, usually, costs climb – because of the sudden need to do more work than anticipated.

When all is said and done, document imaging and document scanning systems can be beneficial to any paper-laden organization. Getting all that information on to the computer screen makes it immediately more findable, searchable and shareable and ultimately more valuable to the organization – but only if the proper preparations are made. Otherwise, it is yet another great way to spend lots of money to little effect, and nobody wants that, no matter how enthusiastic they may have been when the project started.

About the author:
Steve Weissman is a certified instructor in enterprise content and business process management, president of the AIIM New England Chapter and principal consultant at Holly Group, where he advises, writes, teaches and speaks regularly on strategy, needs assessment and RFP development. He can be reached at **sweissman@hollygroup.com***.*

Maintaining document imaging software deployments for ongoing success

By Catherine LaCroix, SearchContentManagement.com Contributor

Once an enterprise has its new document imaging software and scanning equipment up and running, it's business as usual, right? Wrong. The only way to maintain the investment of time and money in the imaging and scanning system is by constantly revisiting it.

But few organizations continue to benchmark the performance of their document imaging systems or fine-tune them after the go-live date, and that can be costly, said Alan Pelz-Sharpe, a principal analyst and director at Real Story Group, a content management consulting firm based in Olney, Md.

 "In the world of capture, [program success] is highly dependent on how you scan the document – no two implementations are the same, so there's no way to say how long it will take or how complicated it will be," said Chris Riley, senior enterprise content management and document capture architect at consulting firm ShareSquared Inc. in Pasadena, Calif.

Companies would do well to set goals for getting a return on investment (ROI) and reducing the costs associated with working on and maintaining paper documents, according to Riley. "The ROI could be to process an invoice in a day versus two weeks," he said. "Determine how much money you want to save by using this technology." But any ROI target should be set high, he said, especially if the imaging and scanning technology is being used to help eliminate manual data entry and reduce staff.

Plan for exceptions in using document imaging software

Something that's often neglected after implementation is the reasonable expectation for exceptions – documents that don't conform to pre-defined rules and have to be entered manually.

"This is the destroyer of capture projects," said Riley, adding that optical character recognition (OCR) and document conversion tools "are interpretive technologies, so they're

never 100% [accurate]." In addition, the core accuracy level of these technologies is already at or near its peak, according to Riley: "It's not going to get much better, so companies need to consider the exceptions."

Many companies start to see an ROI right away from automated document imaging management processes, and there's a window where they're enthralled with the technology, Riley said. But then an invoice or bill of lading comes in that breaks all the rules that have been set up in the imaging system, and the company has to start over.

"What happens is that they get an immediate ROI with the technology, spin several sessions and add exceptions to the [imaging and capture] logic, and then lose the ROI," said Riley. "They have to be willing to process some oddities manually." By modifying the system to handle the exceptions, "you jeopardize all the logic you have," he added. "It's as much of a science as an art."

As part of the implementation process, it's important to determine what percentage of exceptions you are willing to accept. Riley recommended that companies examine all the types of documents they intend to scan and create logic for 80% of the volume.

Tracking and ensuring document imaging success

In document imaging and capture, accuracy is paramount – so it's important to set up and implement imaging and document scanning systems with anticipated levels of accuracy and the metadata values you want to capture. For example, the values might be invoice number, total cost, purchase order number – whatever is pertinent to the particular workflow.

"It's all based on strategy and the definition process of the workflow and the document," said Alan Weintraub, a principal analyst at Forrester Research Inc. in Cambridge, Mass. "It's not a one-size-fits-all. That's part of the up-front definition process you go through when you're defining the system."

In addition, determining and adhering to checks and balances along the way can help to ensure that the information is readable and indexed correctly. "Companies often take the

time to define the metadata, but they don't put enough checks and balances in to make sure they're reaching that quality level," Weintraub said. "You need to make sure you can find the documents." An internal quality assessment process can help ensure that the document management team or external imaging and scanning provider maintains the service level a company requires.

"Keep in mind that in some OCR implementations, making changes may demonstrate improvement in one area and detriment in another," Riley said. "The most likely case is settings for one document type are not working well for another."

Finally, "always refer back to your goal list to see where your production is and what you would like to improve," Riley said. Once a document imaging system is up and running, he added, close monitoring, soliciting feedback from the imaging team and fine-tuning the system as needed can help enterprises get the ROI they were expecting on an ongoing basis.

About the author:
Catherine LaCroix is a freelance writer based in Portland, Ore. She covers technology used in business, education and health care.

Resources from Nuance

How Office Copiers Can Pay for Themselves – and Then Some

Unleashing BPM: Eliminate process bottlenecks created by paper

Integrating paper-based information for Sarbanes-Oxley Section 404

About Nuance

At Nuance, we design and deliver intuitive technologies that help people live and work more intelligently. Whether you're looking to boost personal productivity, departmental efficiency, or enterprise document workflows, Nuance's document scanning, OCR, and conversion solutions Increase operational efficiency - turning everyday tasks into routine processes that take less time, involve fewer steps, and enable office workers to accomplish more. Make business applications more productive - Your team gets more done in less time by transforming paper into digital files they can edit, share, and store securely and quickly. Deliver financial advantage to the bottom line - Nuance-enabled document scanning services stop the inefficiency inherent in paper files and deliver greater productivity, better information management, and lower overhead costs.