

The Five Phases of Capture

Where do you land in the capture continuum?

By Kevin Craine, MBA
Author of the book "Designing a Document Strategy"

You've got a scanner, and you have a lot of paper to manage. Now what?

Despite technology, most companies continue to struggle to manage the burden of paper in many important business processes. Sure, you can scan documents to create a digital image as a first step to eliminate paper; but this "scan and store" approach falls short of the full potential of modern document and data capture systems. You may be missing significant opportunity for more savings and efficiencies; but with all the technological solutions available how do you know which one is right for you?

After more than 20 years, capture systems have evolved from simple solutions for basic scanning into sophisticated and expensive systems for enterprise wide document automation. This paper explores capture systems in five categories – the Five Phases of Capture – and describes the features and benefits of each. Every phase along the continuum can bring value to an organization based on its document strategy. This paper is designed to help you meet your objectives while providing a guide for understanding the complete spectrum of capture capabilities.

The Five Phases of Capture

Around the world a quiet revolution is taking place as organizations make changes to long-standing paper-bound functions, especially as the need to control costs becomes more pressing and the expectations associated with regulatory compliance become more challenging. Scanning documents to eliminate paper is a great place to start, but the goal of current best practices is the fully automated extraction of all relevant data from whatever the data source may be (e.g., a physical document, e-mail, file attachments, faxes, data streams, etc.). Automated validation that the data is correct is another key success factor, as is seamless integration with any number of back-end storage systems and applications.

Not everyone needs a complete capture system, however. In fact, the right solution may fall somewhere between basic scanning and all-inclusive capture enterprise-wide. The question becomes: where do you need to be on the "capture continuum" to best help your organization reduce reliance on paper, lower labor costs, and integrate data and documents across key business activities?



Phase One – Scan and Store

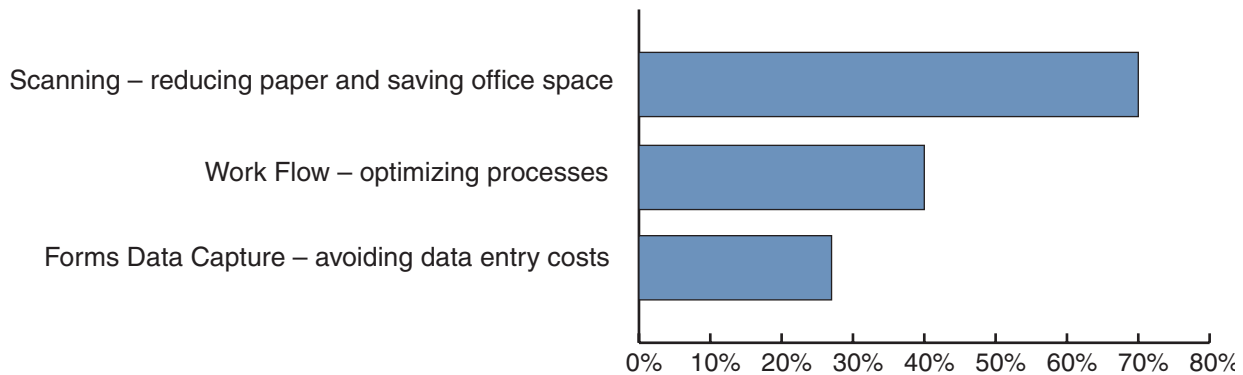
Phase One is characterized by the ability to scan paper documents and store them in digital form. Also known as “imaging,” this phase is the entry level of the capture continuum. It involves scanning hardware and rudimentary software that allows the user to manually enter basic information about the document. By typing keyword data into fields on a screen, users can tuck away the document with a title, a date, and perhaps a brief description – “Acme Company invoice, March 18, 2011,” for example. From there the document image is stored on a computer hard drive or file server for later retrieval.

Benefits

“Scan and store” begins the process of eliminating paper from paper-intensive operations. In today’s competitive economic environment, companies that continue to conduct business on paper will struggle to control the inherent costs and inefficiencies and find themselves at a disadvantage. Despite the increasing digitization of business today, many office workers continue to spend a great deal of time “pushing paper.” Indeed, a 2010 study conducted by the Association for Information and Image Management found that the consumption and use of paper is still growing in 27% of the organizations they surveyed. Companies with more mature scanning and capture operations report a steady transition, however; with over 50% showing a reduction in paper. Out of the 493 member organizations surveyed, 60% reached a positive payback on their investments in capture technology within 18 months.

Companies that continue to conduct business on paper will find themselves at a disadvantage.

Which aspects of scan and capture have produced the highest return on investment?



Capture and Business Process. AIIM Market Intelligence. 2010

Constraints

Scanning a paper document to an image file (e.g., PDF or TIFF) is a straightforward way to begin to eliminate the burden of paper. However, finding documents in a Windows file directory is often difficult, especially as the number of documents (and related documents) continues to grow. Additional labor is required to separate and presort document types, physically scan the pages, and manually enter index keywords. It is important to consider these additional process needs when adopting a Phase One approach. “Scan and store” reduces your reliance on paper and provides more immediate access to documents, but the limitations of manual data entry and the often hodge-podge nature of document storage means you will need to justify a number of labor and productivity issues.

Moving to the next phase

You may find further benefits by applying recognition, sometimes called OCR, to this basic “scan and store” process. Questions to consider:

- Are scanned documents difficult to find?
- Do you spend unwanted time manually entering index information?
- Does it require multiple employees to scan and index documents?
- Is important information buried among the many scanned documents you have stored electronically?
- Could you gain even more benefits by automating the process further?

If you answered “yes” to any of these questions, it is likely that you will find additional process improvements and further return on investment by moving to Phase Two.

Phase Two – Basic Recognition

This phase brings technologies into play that can “read” data from document images and enable you to leverage that information in new ways. Optical Character Recognition (OCR) will read and capture key bits of information and automatically populate a variety of index fields. Standard data like PO number, amount, date, or claims number are common examples of captured data in Phase Two. Often pre-set templates drive the process on structured documents like invoices, claims forms or bills of lading where specific zones are automatically gleaned of information that would otherwise need to be keyed in manually.

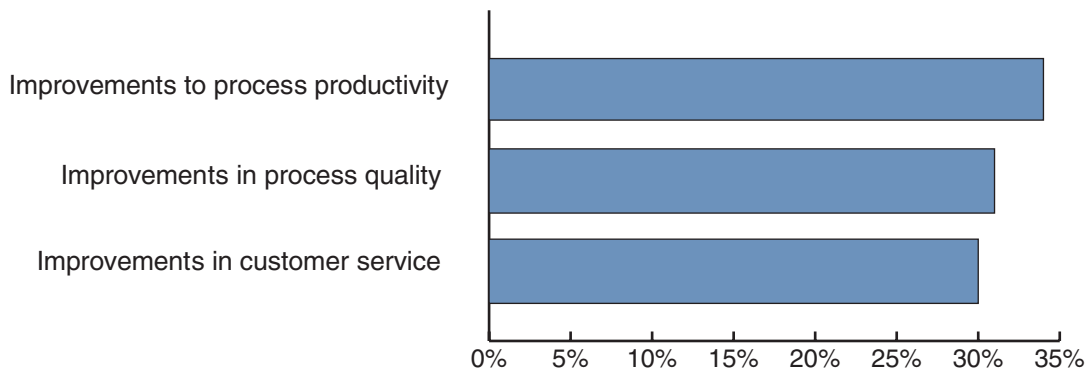
Other advances at this stage include the use of barcodes, which can be interpreted quickly and reveal a more extensive set of data. Optical Mark Recognition (OMR) technology can read and compile check mark responses on a variety of pre-defined forms. Simple business rules can be applied that bring further automation to the workflow and eliminate the labor and effort associated with “paper-pushing” in routine business processes.

Benefits

In addition to eliminating paper and reclaiming storage space, significant improvements in cycle time, error rates and staffing requirements are hard-dollar savings that can be attained by applying automation to the scan and store process. Elementary procedural rules can help drive the verification and validation of the data collected. For example, if the system has a low level of confidence in what was captured, the fields can be highlighted for analysis by a verification clerk. Searching for more detailed information contained in scanned documents can be made easier and faster by using full page OCR to create a searchable PDF file. Rather than wasting time searching for information, employees are empowered with a new level of access and responsiveness that can cut costs and dramatically improve workflow.

Organizations can benefit from a document strategy that guides a forward progression through this “capture continuum.”

Which aspects of data recognition provided the best financial return on your investment



Capture and Business Process. AIIM Market Intelligence. 2010

Constraints

While this phase is a step up from the basic scan and store functions of Phase One, it still requires a fair amount of labor to implement. Staff members will need to presort and separate documents, insert separator sheets, then physically scan the documents and verify the integrity of the data collected. Many capture systems that offer zonal OCR require document templates to be built by programmers which can result in additional expense and effort. In addition, many capture vendors charge a “per click” fee on top of the cost of software, so adding more documents to the system may add to your cost of ownership.

Moving to the next phase

More advanced data extraction techniques can extend the power of scope of information captured from your scanned documents. Questions to consider:

- Do you need to re-key information contained on these documents into back end systems?
- Could you do more to capture “business intelligence” using the information collected?
- Are you spending significant time trying to locate lost documents or fix errors in data entry?

If you answered “yes” to any of these questions, it is likely that you will find additional process improvements and further return on investment by moving to Phase Three.

Phase Three – Intelligent Data Extraction

Once known as “forms processing,” Phase Three is characterized by systems that go beyond capturing basic index information to the extraction of multiple fields and line item detail. Rather than simply gather data for indexing, intelligent data extraction strategies pull more detailed data from documents where data locations may not be known ahead of time, such as ad hoc invoices or healthcare documents. From there, these systems deliver that information to a variety of back end databases and applications (such as an ERP system or a medical claims adjudication system). Advanced data extraction moves beyond structured documents to handle multiple document types and intelligent data recognition technologies allow the system to read handwriting, to map fields, and leverage advanced data codes and glyphs.

A two pronged delivery characterizes this phase as well. Scanned images and the captured index data are passed to imaging storage systems like IBM FileNet Content Manager or IBM Content Manager, while captured line item information is passed to any number of ERP or business application systems. In conjunction with this dual delivery, more complex business rules can be applied and much more complex and automated decisions can be made.

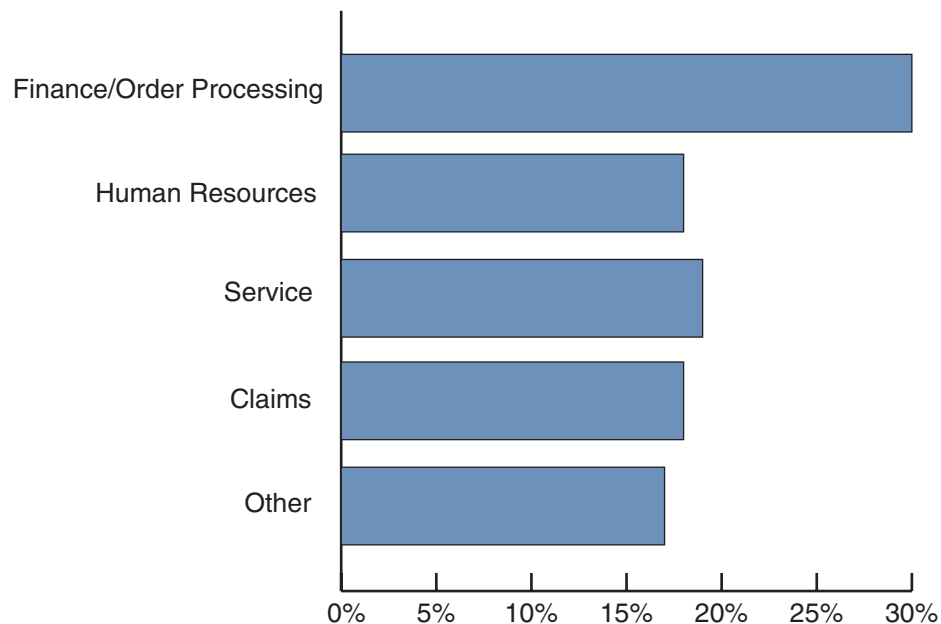
For example, when processing an invoice you may want to capture and utilize a handful of index fields to facilitate retrieval, but business database applications (in SAP or Oracle, for instance) are now able to receive and utilize that same data as well. Ultimately, with the ability to employ advanced business rules, people's roles can change dramatically; away from manual data entry toward higher value roles of exception managers and experts focusing on data quality. Gaining benefits in Phase Three comes down to thoughtful rules that are modular, re-usable, and in a design framework that makes them easy to create, manage and deploy.

Benefits

In Phase Three you get the benefits of capture – eliminate paper, reclaim storage space, more efficient retrieval of scanned documents – but you get the added dimension of feeding vital information into a variety of enterprise content management systems and applications. This can be a “magic ingredient” of your document strategy; with thoughtful integration, the advanced data extraction capabilities in Phase Three opens the door for significant business advantages and process improvements that makes “capture” a strategic part of the enterprise

Benefits are gained in Phase Three with rules that are modular, re-usable, and easy to create, manage and deploy.

Which of your enterprise systems are capture-enabled and integrated at a process level?



Capture and Business Process. AIIM Market Intelligence. 2010

Constraints

Many advanced data capture systems require very complex programming to teach the systems about the documents being processed. Business rules often become monolithic as a result, and require specialized expertise to design, deploy and maintain. A number of new solutions are successful in solving this problem however, and can empower business unit personnel with the ability to design complex functionality. The ability to move this responsibility away from often arcane information technology organizations to frontline users is important because of the need to react more quickly to changing business requirements. New health care legislation, environmental cap and trade, and changes in import and export document processing rules are just a few examples of how the world today is becoming increasingly complex in terms of compliance and legislation. A more responsive level of support is often required as a result.

Moving to the next phase

You might benefit by extending capture beyond a centralized processing environment to diverse locations across your enterprise. Questions to consider:

- Do you pack and ship paper documents from remote locations to a centralized facility for scanning?
- Is scanning at the workgroup level fragmented among pockets of standalone scanners or multi-function devices with no cohesive approach?
- Are departments and individuals left in the dark, waiting for batches of documents shipped to the central office to be processed?
- Would you benefit from moving validation tasks to at-home workers?

If you answered “yes” to any of these questions, it is likely that you will find additional process improvements and further return on investment by moving to Phase Four.

Phase Four – Distributed Capture

This phase builds further advantages by scanning documents at the point of origination instead of shipping documents to a central scanning center. This is an important distinction because capturing paper documents into digital form has traditionally been a centralized business function. In contrast, “distributed capture” is a strategy for entering documents into the business process in decentralized locations – where they originate.

For organizations with a multitude of paper documents being generated in multiple locations, moving scanning and indexing functions to the point of origin saves on mail and courier expense, speeds data input, and expedites the processing of critical business documents.

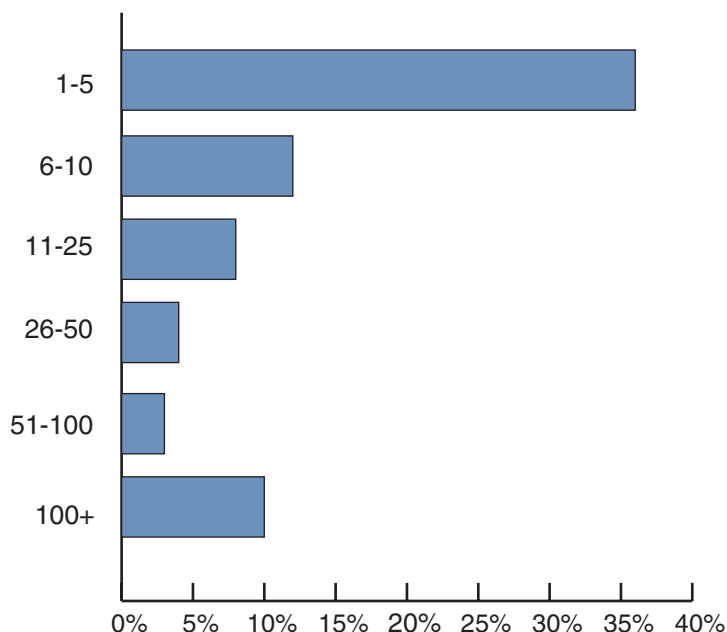
Distributed capture solutions are being used in a wide variety of industries in both large and small document volumes, and the approach provides the ability to establish an efficient document management system across various offices and locations whether throughout various floors of a building or across the world. Verification of data accuracy can also be distributed, typically to line-of-business experts most familiar with the documents and the work requirements. This provides much more responsive and intuitive quality control.

Distributed Capture is a strategy that saves on mail and courier expense, speeds data input, and expedites the processing of critical business documents.

Benefits

Using a Phase Four distributed capture approach organizations can shorten transaction processing time, cut costs in shipping and mailing, and open up new opportunities for increased productivity. While paper-intensive organizations such as transportation firms, banks, insurance companies, and mortgage brokers were early adopters of distributed capture, companies in all verticals and of all sizes find that distributed scanning can help them save time and money, reduce errors, and increase efficiency. Indeed, as companies become more widely dispersed with regional offices, remote employees, and geographically distant customers, distributed capture surfaces as an important tool to ease and facilitate the processing of mission critical information.

How many distributed capture sites do you have across your enterprise?



Distributed Capture Survey Report 2009. Datacap.

Constraints

Performing scan and capture functions away from a centralized processing facility may bring up labor and administrative issues that can be overlooked. After all, you may now be asking high-value “knowledge workers” in business units to process paper documents through scanners, copiers or multi-function devices as part of their daily routine. While this serves to engender a more paperless acumen in your process culture, it can also present additional administrative duties that you must consider. Rather than have an imaging clerk in a dedicated facility presort the documents, insert separator sheets, and verify the integrity of the data, the tasks now will fall to business unit professionals.

Moving to the next phase

Scanning documents to eliminate paper, capturing data for robust search and retrieval, advanced business rules, and ability to leverage extracted data in a variety of systems within a multiplicity of diverse locations -- these are all significant benefits as you advance from phases one through four. What could be left for the final Phase Five? Here are a few questions to consider:

- Do you routinely print e-mail messages and file attachments, only to scan the paper in order to get the information into your content management system?
- Do you need to go to multiple repositories to collect the data and documents you need for the process?
- Does searching critical information scattered among a variety of diverse, complex and isolated storage repositories inhibit your ability to react to the changing demands of your business?

If you answered “yes” to any of these questions, it is likely that you will find additional process improvements and further return on investment by moving to Phase Five.

Phase Five – Enterprise Capture

Phase Five takes advantage of the latest innovations in the form of a universal capture portal where all documents that arrive in an organization are digitized and stored with the least amount of manual labor. This means that the system is sophisticated enough to accept a document in any format it might be delivered in, including on paper via scanners, incoming fax, from multi-function print devices, from a file directory, in the form of desktop documents, data documents like XML, or HTML, via EDI, or in the form of legacy print streams. Solutions at this stage are also capable of reading any kind of barcode or mark recognition system, and there is essentially no limit to the breadth or complexity of business rules that can be employed. Enterprise capture systems also interface with any number of enterprise content management, resource planning and data storage systems.

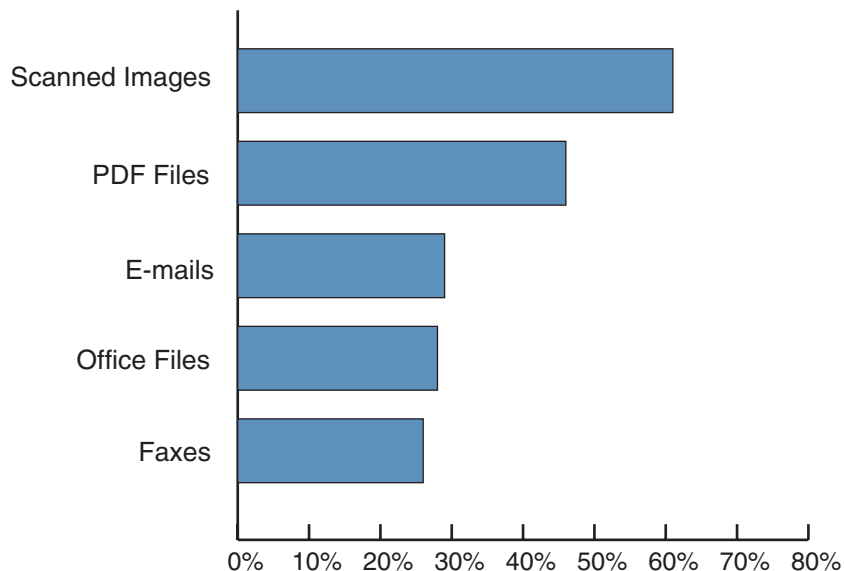
In this phase documents are intelligently identified using advanced text analytics. All relevant data from these documents is captured, validated, and automatically sent to data and document repositories. As a result, only a handful of people are involved to manage the exceptions. A great majority of labor is eliminated from the equation because no document presorting is required, data entry has been eliminated, scanning and verification can be done from anywhere at anytime, and you no longer need to print faxes or e-mail in order to input them into the system.

Benefits

While paper remains a fundamental medium for many organizations, the business environment today has shifted away from physical delivery of documents to include a multiplicity of e-mail, attachments and electronic data. Indeed, this is the real world of document management today; a confluence of paper and digital documents that must be managed. A Phase Five universal capture portal deals with all of those possibilities. As a result, organizations that effectively manage and leverage the entire spectrum of documents will have a competitive performance advantage.

The business environment has shifted to include a variety of electronic data. Stage Five deals with all of those possibilities.

Which of the following content types are you processing for automatic capture of data?



Capture and Business Process. AIIM Market Intelligence. 2010.

Constraints

Without a thoughtful, process-oriented document strategy (versus an entirely technical mindset) organizations using Phase Five capture approaches can miss the mark. It is important to work with technology suppliers that can provide solutions that scale and extend to different departments, different documents and different spheres of application. Be wary of solutions tailored for only one or two phases which may limit your ability to travel down the capture continuum without significant system upgrades and rework. Look for sophisticated partners with a suite of solutions that are modular and flexible, and that will allow you to grow as you move forward.

Capturing Value

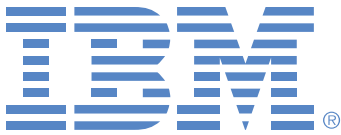
For most organizations a document strategy that leverages the concepts of document and data capture is an undeveloped opportunity to gain business advantage, reduce operating costs and bolster the success of the organization. New affordable and easy to use scanning and data capture technology has brought a new world of opportunity, especially as firms move along the capture continuum with a strategic as well as technological focus.

Use this white paper to guide your evaluation of capture technologies and how the capabilities can enable a document strategy that makes sense. Engage respected partners and look for solutions that are modular, flexible, and that will allow you to grow as you make your journey along the five phases of this continuum. Above all, remember that capturing the value of a document strategy is not about the latest technological advances, but rather how the resulting improvements better the charter of your organization.

About the Author



Kevin Craine is the author of the book “Designing a Document Strategy,” and the host of the Document Strategy Podcast. He has managed document services organizations in the health insurance, aerospace and pharmaceutical industries. He holds an MBA in the Management of Science and Technology, as well as a BA in Communications. Kevin is the managing director of Craine Communications Group. For more information visit www.craingroup.com.



This white paper was commissioned by IBM. IBM's Enterprise Content Management software enables the world's top companies to make better decisions, faster. As a market leader in content, process and compliance software, IBM ECM delivers a broad set of mission-critical solutions that help solve today's most difficult business challenges: managing unstructured content, optimizing business processes and helping satisfy complex compliance requirements through an integrated information infrastructure. More than 13,000 global companies, organizations and governments rely on IBM ECM to improve performance and remain competitive through innovation.