



CANON BUSINESS PROCESS SERVICES

DOCUMENT MANAGEMENT BEST PRACTICES

## **FIVE STEPS TO REDUCING YOUR PAPER PROCESSING COSTS**

Streamline Document Processing with Intelligent Data Capture

### **Are your document processing activities efficient?**

Intelligent data capture can save valuable time in end-to-end document processing to drive lower overall costs. Integrating useful information earlier and more reliably can provide an improved customer experience, less manual data entry that in turn improves data quality, and greater security controls.

# INTRODUCTION

## Increase Efficiency and Reduce Document Processing Costs with Intelligent Data Capture

Effectively managing document processing activities is critical to an efficiently run department and your business success. It can help raise employee productivity; reduce the cost associated with processing documents such as applications, forms and invoices; and strengthen security and improve cycle times in customer service. Intelligent data capture is an approach that can help you streamline document processing activities and realize these and other goals.



# DEFINING INTELLIGENT DATA CAPTURE

**Before looking in more detail at the business benefits and implementation strategies connected with intelligent data capture, let's start with how Canon Business Process Services (Canon) defines the term.**

Intelligent data capture is the process of automatically scanning traditional long-form documents or electronic pages, extracting specific data (regardless of its structure) and integrating information with downstream workflow processes and systems

When implemented correctly, this approach can save valuable time in end-to-end document processing, which in turn can drive lower overall costs. Other benefits of integrating useful information earlier and more reliably in the document management process include, as spotlighted earlier, an improved customer experience, less manual data entry that in turn improves data quality, and greater security controls enabling an enterprise to better meet today's compliance requirements.

## **Capture the Right Information to Improve and Optimize Your Team's Workflow**

With this definition in mind, let's begin our closer look at intelligent data capture with why we at Canon believe it matters. According to our own industry research and discussions with clients and subject matter experts, approximately 80 percent of an organization's information assets exist as unstructured content. In a sense this means that the vast majority of your company's "intelligence" is trapped inside unstructured, unmanaged documents. In this scenario, a tremendous amount of useful information is not maximized via processes and technologies that can efficiently move it into workflow systems that support and help grow the business.

Trapped does not necessarily mean stuck forever. In reality, the data is caught in a bottleneck that hinders its flow to downstream systems, some of which can include:

- Enterprise risk management systems
- Client management software
- Business intelligence dashboards
- Private or public cloud repositories that are used for collaboration

# YOUR STRUCTURED VS. UNSTRUCTURED DATA IMPACTS DOWNSTREAM PROCESSES

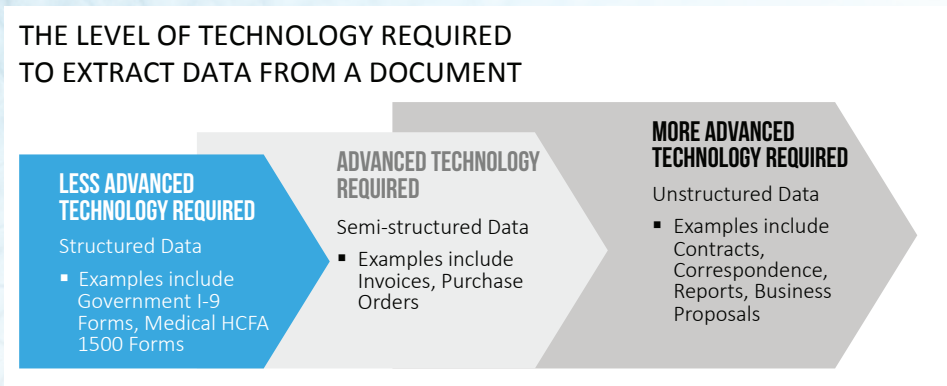


Figure 1

The key to unlocking this trapped data so it can more effectively support downstream processes is to extract it. We'll look at this process in more detail shortly. First, it's important to understand two concepts. The first is illustrated by Figure 1, which contains a line graph representing the level of technology required to extract data from a document.

The left side of the illustration indicates that less advanced technology is required to extract data while, as we move further toward the right, the technology becomes increasingly more robust to accommodate the need to make more complex decisions. The reason for this is that within structured documents (lower left), the essential information is consistently located in the same place. Examples of these documents include government I-9 and medical HCFA (Health Care Financing Administration) 1500 forms.

Within semi-structured documents (middle of the line), information generally is found in the same location with some variation in how data is displayed, depending on the individual company. Documents such as invoices and purchase orders fall into this category. Unstructured documents (upper right) display information with the most variation. Data could be located anywhere within these documents, which typically include contracts, correspondence, reports and business proposals. Therefore, extracting data from these documents requires more complex decisions, driving the need for more sophisticated technology.

Besides document structure, the second concept that companies need to be clear about is that data enters an enterprise in many different formats. Ironically, technology

has actually increased the number of these formats, not reduced them as was anticipated a decade ago. Data is received in the form of paper documents as well as faxes, e-faxes and emails with attachments. Beyond that, information is transmitted as Microsoft Office files (Word, Excel and PowerPoint), PDF documents and more. This is compounded by the fact that data can enter your company from different devices (desktop, laptop or mobile) and multiple locations within and external to your organization.

# THE IMPORTANCE OF DATA EXTRACTION

As pointed out earlier, businesses extract data in order to make downstream systems and workflow processes more efficient and cost effective. This in turn enables an organization to grow, maintain a competitive edge and realize other significant benefits. One example of the value of data extraction, illustrated in Figure 2, concerns an accounts payable solution.

Canon teamed with an investment bank that wanted to improve its extraction process and technology. The company had moved its approval process offshore, but wanted to reduce the cycle time for getting invoices into its payment system. Additionally, the bank saw an opportunity to reduce the staff required to manage this process so it could redeploy resources to more strategic areas of the operation.

While our project included using OCR (optical character recognition) technology to extract and index a variety of key data points such as vendor name, invoice number and line item detail, the purchase order number was especially important. Once validated, purchase order invoices could be paid without a departmental approver, which could reduce processing cycle time by as much as a week.

Also, the information for all invoices, including those that required approval, would flow more quickly into our client's ERP system, which triggered the approval workflow. Due to the enhanced extraction process and technology we put into place, the bank reduced processing cycle time from seven days to one and reassigned 10 staff members to areas that more directly supported its core business. These results indicate the potential business value of effective data extraction.

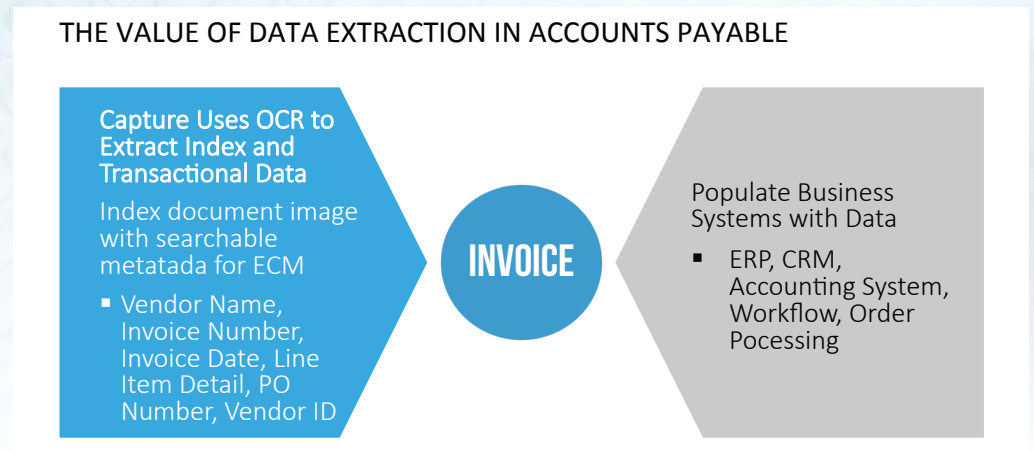


Figure 2

# FIVE STEPS TO IMPLEMENTING THE INTELLIGENT DATA CAPTURE SOLUTIONS

Now that we've established its potential, let's examine intelligent data capture at work, including five key steps to implementing a solution.

Figure 3 shows the basic flow of the capture process, beginning with how data typically enters an organization in many different formats that are sent from multiple locations, as noted earlier.

The information is captured using scanning technology, comprising hardware and software. OCR is applied, and at this point the real power of the approach begins to be harnessed. Data is extracted using business rules, which specify such details as how a document is recognized: what information is needed from a particular document type as well as how this information is identified, where it is sent and in what format.

This is a broad view of how the process works, but how can an organization plan to implement an intelligent data capture solution? At Canon we have determined five essential steps that can support a successful implementation.

## 1. Centralize Intake

Step one is to centralize the intake of information. This is perhaps the most important step in the process of implementing a data capture solution. It revolves around understanding how and where useful data is being received, and in what formats.

Another key to this step is for the organization to identify the type of business model it wants to use for the capture process: internally managed or outsourced. We'll examine some pros and cons of each model later in this paper. The point for now is that, depending on the model chosen, the enterprise will want to designate a point of data capture and begin to consolidate

information streams through that point. The reason for this is to help ensure that all useful data will be handled by the new workflow and that duplication of efforts will be avoided as much as possible.

## BASIC FLOW OF THE CAPTURE PROCESS

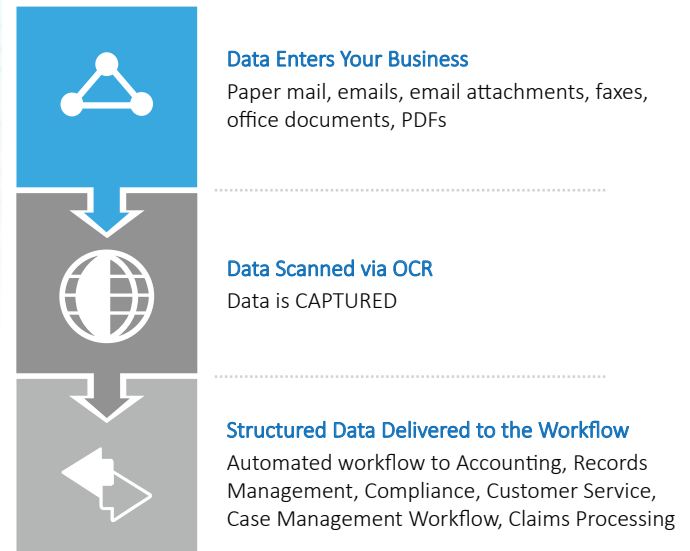


Figure 3

## 5 STEPS TO IMPLEMENTING THE COST-CUTTING INTELLIGENT DATA CAPTURE SOLUTIONS

It is important to note that centralizing intake does not have to disrupt ongoing processes that the company wants to retain. For example, perhaps the established practice is to send invoices to different locations. The organization might want to continue this approach so that the locations can efficiently maintain current vendor relationships. In this case, the invoices could be scanned locally at multifunctional devices, and then the electronic images could be automatically routed to a central point for processing, keeping the local workflow intact.

### 2. Automate Capture

Step two involves automating the document capture process using OCR and business rules. Implementing new capture systems of this type can require a substantial investment. Some of the more complex software systems are not easy to deploy and many IT departments lack the expertise to install these solutions. Consequently, service costs from vendors can accumulate.

One way to contain these expenses is to use a phased approach. This includes identifying

the important channels and/or locations and bringing them online one at a time. In this manner a company can purchase fewer licenses and reduce installation time and associated costs.

Another possibility is to leverage the expertise and technology that the right managed services provider can deliver. This approach can work well regardless of the centralization model, but is especially helpful for organizations deploying an offsite solution because they can take advantage of the provider's technology and infrastructure.

### 3. Classify Documents

The goal of step three is to virtually eliminate manual document sorting by classifying document types, beginning with the most important. One challenge in this step is to ensure that explicit rules exist to identify or classify the document types. In our experience working with clients, we often find that manual processes, especially those in place for a long time, may not have specific written rules for document classification. For example, how does the company identify purchase order and non-

purchase-order invoices? Is the purchase order number required to be on each invoice? Another challenge is that it may be a regular practice for processors to read some unstructured documents in order to determine their type, such as a contract or proposal.

Companies can meet these challenges by, first, writing rules for document classification, starting with documents that are either most important or easiest to classify. Second, the implementation team can undertake an analysis designed to help gain better control of unstructured documents. This could involve talking to processors and determining what key words or phrases they look for when classifying documents. Third, the team can investigate the possibility of leveraging a solution that is commonly referred to as clustering. This is a machine learning technology that involves scanning and automatically grouping documents together that have similar characteristics. The document groups can be reviewed and corrected, and the clustering system continually adjusted, so that over time it can identify document types without the need to code a series of complex rules.

## 5 STEPS TO IMPLEMENTING THE COST-CUTTING INTELLIGENT DATA CAPTURE SOLUTIONS

### 4. Extract Data

Step four revolves around cost-effectively extracting data. This means reducing labor costs associated with data entry. The key point is that in its current document processing, an organization may use manual data entry which, being labor intensive, causes the overall document management process to be more costly and error prone than it needs to be. Additionally, manual data entry keeps the document processing system from being scalable because there is a limit on how many characters an individual can key in a given period of time. Finally, if more processors are required, the learning curve for training them is going to differ for each person.

To resolve this situation, an enterprise can take a few initiatives. One is to determine critical data points and the rules for finding them. Then, leverage software to automatically extract those data points based on clear rules. Finally, assign manual steps only to exception processing.

### 5. Export Data

The objective of step five is to distribute—or export—the extracted data to areas within the organization where it will be used. As with the previous steps, there is at least one key challenge involved. In this case, creating custom links between capture systems to downstream platforms, such as an ERP system, can be expensive and usually require professional service support. There are a few ways to help alleviate this pitfall:

- **Document the requirements for the downstream systems.** These systems may be able to accept information in more formats than expected, enabling an “out of the box” export from the capture system to work.
- **Use a phased approach, connecting to one system at a time.** This potentially reduces the amount of professional services needed at once.
- **Leverage a managed services provider with the right experience and technology.** The latter might include a connector that works “as is” or could work with a few changes.

# CASE HISTORY EXAMPLES

Now that we've examined five steps that can help support a successful implementation, let's see how the intelligent data capture process can work in real world scenarios via three brief case history examples.

## Insurance Company Case History

The first, as illustrated in Figure 4, concerns an insurance company that needed to capture data from the approximately 10,000 claims it received monthly.

The company receives the forms in several formats (PDFs, scanned typewritten forms, handwritten forms and faxes). Because the insurer receives the forms at its main location, centralizing the inbound document stream was a relatively simple step. Additionally, as part of its information intake strategy the company decided to use an offsite processing model in order to maximize space and contain cost.

The other elements of the solution included automatically capturing data from typewritten and electronically generated forms. Then,

business rules were applied to the extracted data, which facilitated moving the majority of the forms efficiently through the approval process workflow. This had the effect of lowering the company's claims settlement costs and reducing processing time to less than 24 hours. Manual activity was only used for processing handwritten forms or forms that required exception processing. This system not only helped raise efficiency and lower costs, it enabled the company to better meet industry regulatory and compliance requirements.

## CAPTURING DATA FROM MEDICAL CLAIMS FORMS

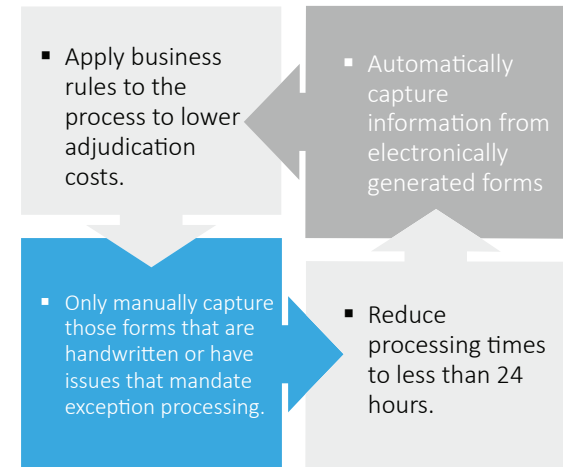


Figure 4

## CASE HISTORY EXAMPLES

### A Manufacturer Case History

In our second example, highlighted in Figure 5, a manufacturer needed to digitize unstructured data in the form of vehicle and dealer files, comprising approximately three million pages.

The company teamed with Canon to help address several key challenges and implement a solution. The challenges included:

- The documents were stored in multiple locations.
- The company needed a secure central repository, but didn't want to install or manage it on the company's premises.
- While some files could be digitized offsite, some sensitive documents had to be scanned onsite.

Canon's solution was to create a temporary scanning operation onsite for the sensitive documents and manage intake for the other documents at an offsite facility. Our team used OCR technology to capture important vehicle file data in full text. This data included automobile VIN number, vehicle type, region and year. We also installed a secure cloud storage system for the company, into which we imported the documents. The repository included separate network access for each location in order to provide added security. The project, scheduled for six months, was completed earlier than planned.

### DIGITAL CAPTURE AND STORAGE OF VEHICLE FILES

Captured Key Field and OCR'd the Vehicles Files in Full Text

- Vehicle Identification Number
- Employee Name
- Vehicle Type
- Year
- Region

Direct Upload to Cloud or Network Storage

Figure 5

## CASE HISTORY EXAMPLES

### An Educational Institution Case History

Our third example (highlighted in Figure 6) concerns an educational institution that needed help in capturing data connected with transcripts and other admissions documents for prospective students.

The admissions documents represented an income stream; as such the decision was made to perform the scanning functions onsite. One challenge was that, while the university had an enterprise repository and capture solution in place, these were not optimized for batch ingestion. As a result, the school was scanning one page or one document at a time. To address this and other challenges, the school engaged Canon to find a solution.

The Canon team redesigned the school's capture process using technology that was

### CAPTURE OF TRANSFER INFORMATION FOR STUDENT APPLICANTS

- Direct intake from mail stream
- Automatically open envelopes and extract contents
- Validate originals (manual step)
- Scan transcripts and supporting documents
- Pull key field data
- Export data to admissions workflow for evaluation

already in place. This included creating a process for receiving mail directly and automatically opening envelopes and extracting the contents. Next came the only pure manual step in the workflow: validating original documents. This step was required in order to meet the internal business rule that original transcripts must follow a different workflow than copies. Original transcripts are frequently stamped with a raised seal that might not be recognized during an automated scanning process.

After the validation, the documents follow a similar scan and data extraction process as highlighted in earlier examples. The data is

then moved further along the workflow to a point where it is accessed by a university team that evaluates the transcripts. It is important to note that previously the school was using highly skilled labor to manually perform these document processing tasks. Because the new automated capture process requires less manual processing and staff, the school was able to reassign employees to perform activities better suited to their skill set.

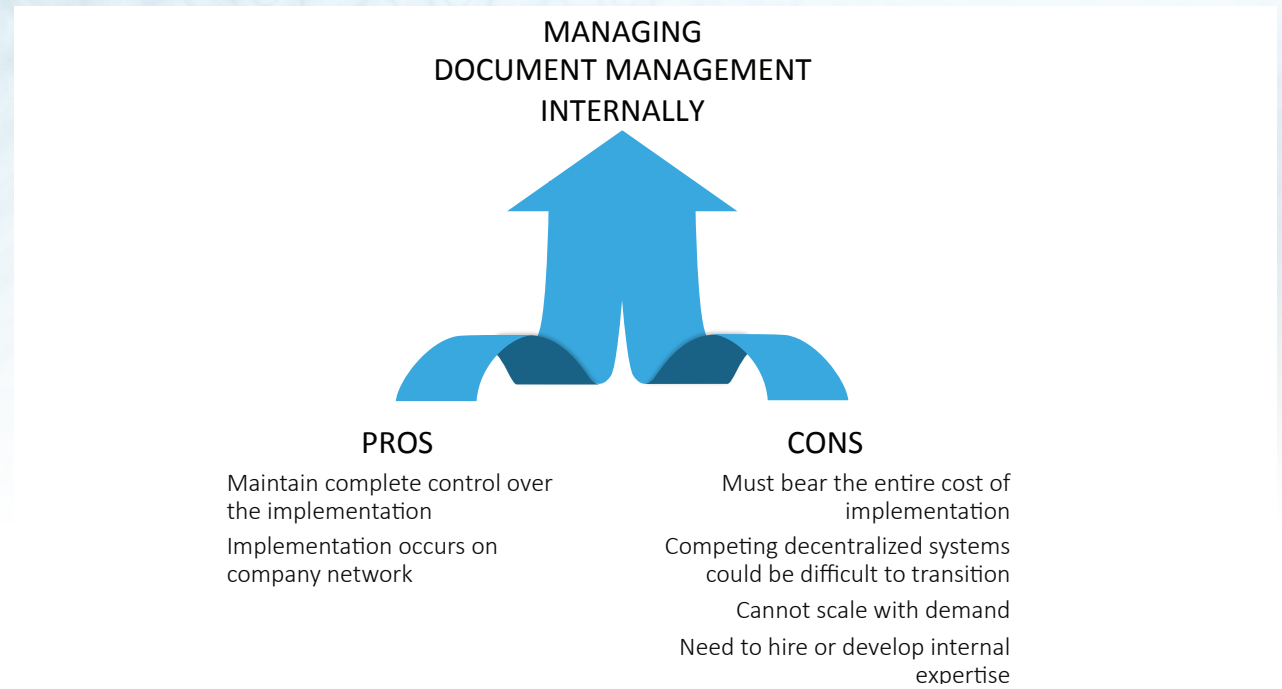
Figure 6

# MANAGING DOCUMENTS INTERNALLY VS. OUTSOURCING

The earlier section highlighting Step 1 (Centralize Intake) noted that a key to this step is for an enterprise to clarify the business model it wants to use for managing the capture process: internally or outsourced. Let's briefly examine some pros and cons of each model. Figure 7 spotlights some of the advantages and disadvantages of managing documents, including the capture process, internally.

On the plus side, the organization maintains control over implementing and maintaining the capture process. Further, the company's IT department controls the installation, which occurs on an internal network.

There are several issues on the minus side. First and foremost, the company has to bear the entire cost for the implementation, including hardware, software, training and project management. Another challenge related to implementation is that many IT departments do



not have the requisite expertise with capture systems to ensure a smooth install process. Additionally, implementation might require transitioning from legacy systems. Depending on the age of the systems, such a transition could be costly. And looking ahead, as the

need to scale up arises due to changes in the business, the company may need to purchase additional equipment and software licenses, and incur other related expenses.

Figure 7

## MANAGING DOCUMENTS INTERNALLY VS. OUTSOURCING

### OUTSOURCING DOCUMENT MANAGEMENT



Figure 8

Figure 8 depicts some of the pros and cons associated with an outsourcing model.

The “pros” of the outsourcing model include the fact that, because an organization is leveraging the expertise, processes, technology and economies of scale connected with a service provider, there is the potential to realize a greater return on investment over time. Additionally, a service provider often provides the latest technology upgrades and security updates as part of an outsourcing agreement, freeing the organization from these concerns. Another significant benefit is access to a service provider’s pool of technology and processing experts, who can help ensure the success of future projects.

The main potential drawback to the outsourcing model is that—due to security concerns for documents containing sensitive data such as employee benefits and compensation—some document management activities may need to occur offsite, away from where employees are located. However, there is a “political” issue connected with offsite document management. Some executives are simply not comfortable with an offsite, outsourced model because they believe that this generally poses a higher security risk. However, in our experience offsite operations managed by a services provider are often protected with greater security than a company’s internal, onsite systems because the provider’s offsite facilities have to meet a much broader range of industry regulations and guidelines.

# SUMMING UP

**Whether your company is looking to reduce document processing costs, boost employee productivity, safeguard against business interruption or respond to internal and external customers faster, intelligent data capture can help achieve these and other critical business objectives.**

Developing a document management strategy that will help you achieve these goals requires expertise and best-in-class technology. Canon Business Process Services has over 20 years of experience providing end-to-end imaging, document conversion and indexing services, and we are here to help you. You will find that intelligent data capture is among the best investments you can make for your business.

Call 888-623-2668 for an assessment or visit [cbps.canon.com](https://cbps.canon.com) to learn more.