

# Distributed Capture

## *How Remote Scanning And Indexing Could Benefit Your Organization*

### Introduction

In the early 1990's, when Electronic Data Interchange (EDI) was just becoming established among organizations and their largest vendors, paper-based transactions dominated. Since that time, the volume of electronic transactions has increased dramatically. Yet, in spite of predictions to the contrary, paper-based business documents remain a significant burden for organizations today.

The efficiency and speed of handling electronic data raises substantial challenges for processors of paper. Concurrent with the increase in electronic transactions is an improvement in productivity, which alters customer expectations and forces more efficiency in the handling of paper. No matter what industry or government agency, the pressure to process transactions quickly and efficiently will continue to increase.

Today's organizations need to minimize time spent physically moving the paper – “mail float” as it is called in the financial industry. Moving paper is expensive, risky, and time consuming. The answer for many organizations is to transform it into an electronic image as early in the cycle as possible. There are a variety of solutions available today – incorporating low cost scanners, capture technology and the Internet – that deliver the capabilities to scan and index documents anywhere in the world.

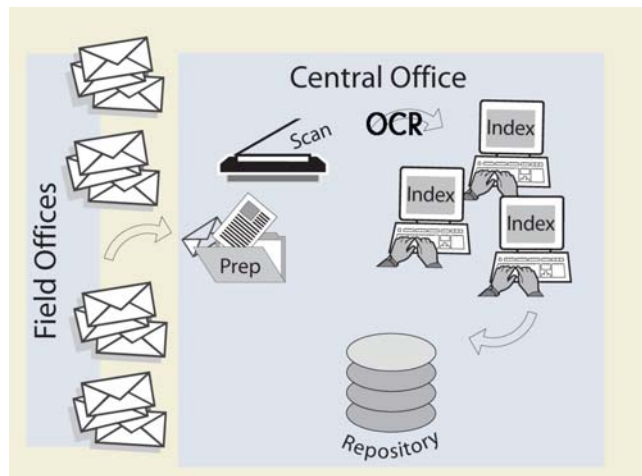
### Methods Of Distributed Capture

Data and document capture involves the scanning of paper to create a digital image. Then recognition technologies extract data from the image and/or assign index values to the image to make it easy to locate the document image after it has been stored in a repository.

There are three main methods in use to capture paper business documents and their associated data or indexes:

**1. Centralized Capture -** Paper documents are collected or shipped to a central scanning point where they are converted into a digital image. Automated indexing is performed by a central server, which may be supplemented by local or remote key entry. Centralized capture is efficient because it allows for each function to have dedicated operations — a production line approach (see Figure 1).

Dedicated staff can prep paper at rates of up to 10,000 sheets per hour, scanners can generate images at rates of more than 100 sheets/minute, optical character recognition (OCR) and other recognition



**Figure 1 – Centralized Capture Environment**

technologies can reduce or eliminate the need to key data, while dedicated key entry operators being paid \$10/hour will key at 8,000 keystrokes/hour or more.

If the paper documents *can* be received centrally, centralized capture is the most efficient and cost-effective way to process documents and data. However, there are many industries where documents are generated locally and dispersed over a wide area, where there is an advantage in rapid processing. Increasingly, organizations are augmenting their capture strategies with what has become known as Distributed or Remote Capture.

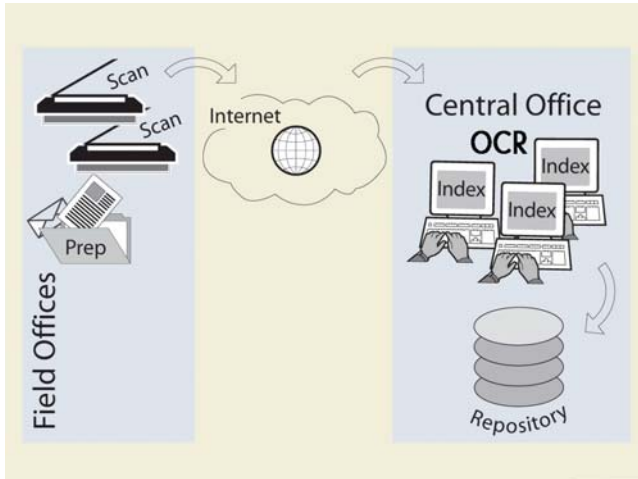


Figure 2 – Distributed Scanning/Central Indexing

**2. Distributed Scanning with Centralized Indexing** - The paper documents are scanned or faxed at a local point and then the image is forwarded to a central point where it is indexed and placed in a centralized repository. In many cases, the main indexes are provided as batch separators, inserted between documents locally at scan time. This approach works well where a third party can do the scanning. For example, insurance or other brokers, when filing a claim or an application, may not know the indexing or storage requirements or they may not have access to all the documents in the transaction set. In these cases, it is

best to scan the paper into an image at the place of receipt, transmit it to a centralized location, and then apply the index data or extract the database data (See Figure 2.).

**3. Distributed Capture with Localized Indexing & Control.** The paper documents are scanned at a local point and indexed at that same point; kept locally using distributed document management and also made available with its indexes from a shared (normally centralized) repository. This type of operation works well where the local organization understand the indexes and data needs and has a requirement to work on the documents locally at the point of receipt (See Figure 3).

## When To Distribute Capture?

In an e-business environment, increasing numbers of forms and supporting business transactions are generated and entered directly or transmitted via e-mail (see Figure 4). However, there are still many transactions that have resisted the shift to electronic forms. Either they contain signatures, or they are in older industries that still insist on paper forms.

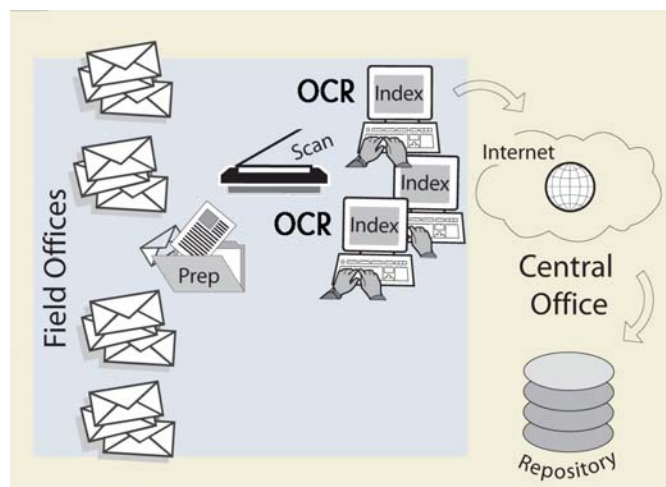
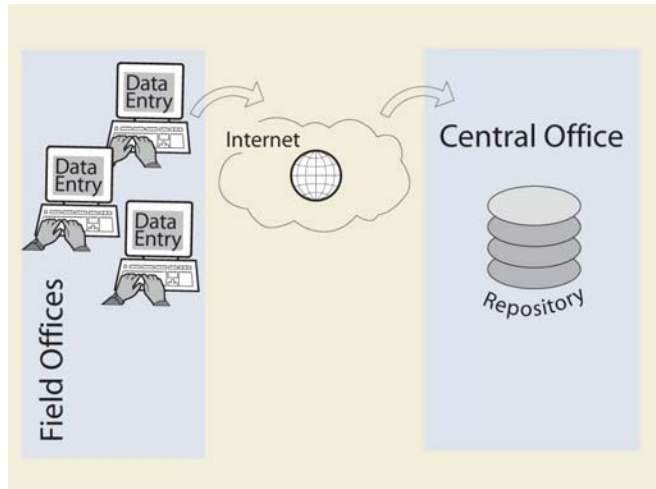


Figure 3 – Remote Scanning and Indexing

Nevertheless, to remain competitive, companies need to process and manage the use of paper at similar speeds to electronic documents. The mail or courier services are too slow and too expensive and there is a small risk of loss or damage to the documents. The solution is to truncate the paper at the first point of entry into the organization. This point of entry varies depending on the type of business and the type of transaction.



**Figure 4 – E-Forms**

Many companies have centralized the collection of some or all of their transactional and other documents – to optimize the process, they are often pre-sorted using P.O. Box numbers. Orders, payments, applications, inquiries etc. are often handled this way and centralized batch scanning is the most effective way to handle these types of documents. But some documents cannot be easily received centrally or there are business reasons why this may not be convenient.

and/or signed by employees in remote offices where they work; insurance applications may be completed by independent brokers; mortgage applications may be completed in a bank branch or mortgage company; shipping documents often have to move with the goods and be signed off; car rental documents are left with the car where it is delivered; local government officials collect immigration or other documents.

For example, Human Resource documents are frequently completed

When documents are distributed, the reality is that, in many cases, it is impossible to get the user to deliver the document quickly to the scanning center. Processing is delayed while documents are sent via courier services and the U.S. mail to the centralized point for capture, which also adds to the cost.

## **Distributed Scanning With Central Indexing**

While not fully distributing the capture process, distributing the scanning process can eliminate courier and mail costs and remove the risk of losing documents in transit. In simple indexing applications, document separators, featuring barcodes, can provide the automated indexing which is then supplemented with OCR and/or key entry at the centralized site. This method works well where the scanning is taking place at a third party site such as an agent or broker.

Early applications used fax machines to provide a remote distributed image capture capability, and fax is still being used in a number of applications. Effectively the fax machine acts as a digitizing scanner, but it has problems:

- The feeders tend to be unreliable causing skew on the paper, sometimes parts get eliminated and thin papers jam or double feed
- Standard mode FAX is just 100 x 200 dpi, which is inadequate for good OCR recognition and may cause legibility problems with small fonts
- If an image sensor goes bad, the sender never knows until the recipient calls to report the problem
- Noise on the line can introduce artifacts, which make the image unreadable

Today these types of distributed solutions use portable scanners and the Internet for delivery. This solves the problems posed by fax machines, and distributed capture with centralized index control is an important option for some applications. Distributed scanning is also a way to transition from a paper-based forms system to a more efficient electronic forms system.

T-Mobile, the cell phone provider, is in the process of converting all of its human resources forms to electronic format – a process they estimate will take 3-5 years to complete. In the meantime, T-Mobile is scanning all its historic H-R forms at their remote sites and then storing them on a centralized FileNet system. Distributed scanning saves moving the paper for centralized scanning, which is expensive and runs the risk of loss in transportation. Staff from key departments – legal, H-R and executive management – can access a person's file with appropriate security, while anyone who needs to access their own personnel records can do so in their local H-R department.

SIRVA, a major shipping company consisting of North American Van Lines, Allied Van Lines, and Global Van Lines, is distributing the capture of its shipping documents to 340 third party agents' offices, reducing their processing time for new shipments *by an average of 2 days* and resulting in greater efficiency and customer service.

When a moving vehicle driver returns to the agent's office, the documents are scanned using software supplied by SIRVA. The agents pay for the scanning equipment, as they benefit with faster payment from SIRVA and more immediate access for queries. An average shipment consists of 20-25 pages of details – each of which is identifiable electronically by either a barcode or an identifying character. Each document is preceded by an index separator sheet, which uses a barcode to automatically attach the contract number, and other “meta data.”

A secondary benefit of distributed capture is realized if claims are lodged with the shipper's insurance. Now the insurer can access the documents immediately over the Internet, rather than the previous method where a clerk had to extract the documents, copy them, make up a folder and then fax or overnight them to the insurer.

## **Distributed Scanning And Indexing**

Another solution is to scan and index the documents at the remote site and then transmit the images and associated indexes to a centralized repository or make them available from a distributed database. This works well in environments where the recipient of the paper documents is the one who knows how it needs to be filed; how it needs to be accessed; and the relationship to other papers or electronic documents in that transaction set.

Additional benefits include:

- Allows for almost instant entry into the databases
- Provides the ability to use the standard infrastructure
- Takes advantage of equipment that the office worker is familiar with, such as the office copier as the scanning device
- Provides integration with a simple desk top document storage and retrieval system allowing reduction in physical filing
- Enables the potential to use mobile scanning devices connected through a wireless link into the Internet

Merrill Lynch is currently piloting the idea of using Xerox copiers to scan documents received in the mail at each of their 800 branches. Scanned images would be routed through each employee's personal desktop management system and transmitted to the appropriate destination.

This will replace the current method, which requires copying the mail, then batching the documents and sending them by courier for centralized scanning.

For Merrill Lynch, the potential savings and benefits are plenty:

- No transportation costs (\$1 Million/year)
- Reduced copying (\$800,000/year)
- No extra paper generated and elimination of the associated storage costs
- Reduction of lost or missing papers; reducing the need to communicate to sort out errors

Merrill Lynch estimates savings of around \$2 Million AND improvement in efficiency and customer service. Like SIRVA, the new system shaves 1-2 days off document cycle time. Merrill Lynch employees will be able to collaborate on documents within the field offices. The only downside is an estimated 30-35% increase in network bandwidth to support the images.

The problem is that clerical staff earning \$20, \$30 or more now scan and index the documents, rather than specialized scanning and indexing staff earning \$10 or less. It is therefore critical to install a system, which minimizes their labor. Ideally, they will be able to put a stack of papers into the auto-feeder, press start and be sure that the system knows what to do with each document and can index or extract most of the needed data. Systems are now available to do this.

## **A Mixed Environment**

In reality, many applications are using a mix of capture methods: centralized, distributed with centralized indexing, and distributed with localized indexing.

Taylor, Bean and Whittaker (TBW), one of the top mortgage wholesalers in the US, processes loans that can involve 190 different document types. With the recent surge in new mortgages driven by lower interest rates, the organization scans more than 300,000 images a day at its Ocala, Florida processing center. To improve efficiency and reduce shipping costs, in 2003 TBW added distributed capture to its centralized Datacap Taskmaster capture system.

Approximately 25% of total scanning is now done at 12 regional TBW offices throughout the country. Doug Bowman, Vice President of Product Development, reports that by simply reducing the movement of paper, TBW has saved substantial shipping costs. According to Bowman, an average regional office was spending \$30,000 a month on overnight courier services, an expense completely eliminated by remote scanning.

In 2005, TBW is introducing a distributed solution that will enable their customers – independent mortgage brokers and small lending institutions – to scan documents themselves at their own site. The benefit of local scanning to TBW's customers is they receive approvals faster, resulting in greater satisfaction among *their* customers. The benefit to TBW is they can reduce their own processing costs even further.

## **Technology Advances will Increase Use of Distributed Capture**

### ***Recognition, Infrastructure, Equipment***

Most remote offices are equipped with networked PC's and the field offices are usually linked to the corporate IT systems via high speed WANs via an Extranet. This connectivity allows immediate access to databases and other centralized information.

## **Powerful, Low Cost Scanners**

Scanners priced at a few thousand dollars can scan a page in about a second and include high-speed features such as dual-sided scanning, double-feed detection and replaceable rollers. This is all packaged in compact, easy to use 'single button' units, enabling quick and easy operation. Also, digital copiers (often called Multi-Function Products or MFP's) with networked printing and scanning ability are being installed in field offices.

## **Accelerating Recognition Technology**

High speed microprocessors in today's PC's enable optical character recognition (OCR) software to provide much more automated recognition of documents than ever before. This software has the ability to rapidly scan the complete document image, classify it, find the main index fields, and extract transactional data as well as index data. For example, Datacap Taskmaster, which enables browser-based scanning and indexing, can link the Taskmaster Web Server to such processing-based tasks as OCR, ICR, OMR and barcode recognition, as well as automated validations running in the background. After scanning remotely, batches of document images can be run through recognition to automate data entry and indexing.

## **Desktop Document Management**

Desktop document management systems enable personal filing and retrieval, where the image can also be sent up to the main repository content management system and shared through collaboration tools.

As a result, most of the growth rate in scanners (over 20% annually) is occurring in workgroup and departmental scanners, designed to scan a few hundred pages a day. High speed, high volume scanner growth is essentially flat, according to Susan Moyse of InfoTrends/CAP Ventures. These low-end production scanners, together with small personal scanners and fax machines, are being installed in distributed environments.

Currently, only an estimated 5% - 10% of digital copiers are being used for distributed capture, linked to a document management system. The reasons for this limited usage are partly infrastructure – only about 50% are currently connected to the network as shared printing devices – and partly design. Some limitations:

- The display is typically too small to use easily, and completely inadequate for monitoring image quality
- They don't offer duplex (dual-sided) scanning; they only scan both sides by turning the paper over and rerunning it
- Their document feeders are not as robust as those of production scanners
- Features such as endorsers have not been implemented

Much or all of this is set to change, already MFP copiers are becoming JAVA addressable, control panels are getting bigger, even a duplex scanner has been implemented by one vendor. The result is likely to be a major change in the use of these devices in low volume distributed capture during the next few years.

## **Remote Monitoring Essential To Success**

One of the potential issues with distributing capture is controlling the process to ensure that there is no loss of documents or images. Control is also particularly important with the Sarbanes-Oxley and HIPAA regulations, which mandate that access to all financial and health data be limited to those with permission and the path that document images follow must be tracked and

auditable. Browser-based capture systems must provide a useful way to centrally control and monitor distributed scanning and indexing.

Datacap Taskmaster offers administrators complete control over the distributed capture process, beginning with the level of permissions allowed users. System supervisors can limit the classes of documents to which remote index operators can have access, and can monitor the status of batches that each station is processing.

When operators are scanning remotely using Taskmaster, the Taskmaster Web Server creates a new batch at scan time, whether the images are uploaded immediately or not. From that moment on, all activity within that batch is traceable, assuring that the existence and location of document images is known at all times.

The interface for viewing batch information is the Taskmaster Job Monitor (see Figure 5), which also offers complete visibility on every station and logged-in browser and can generate a wide variety of reports. If they want to view activity on a certain station or user, administrators can simply filter out the other users or stations.

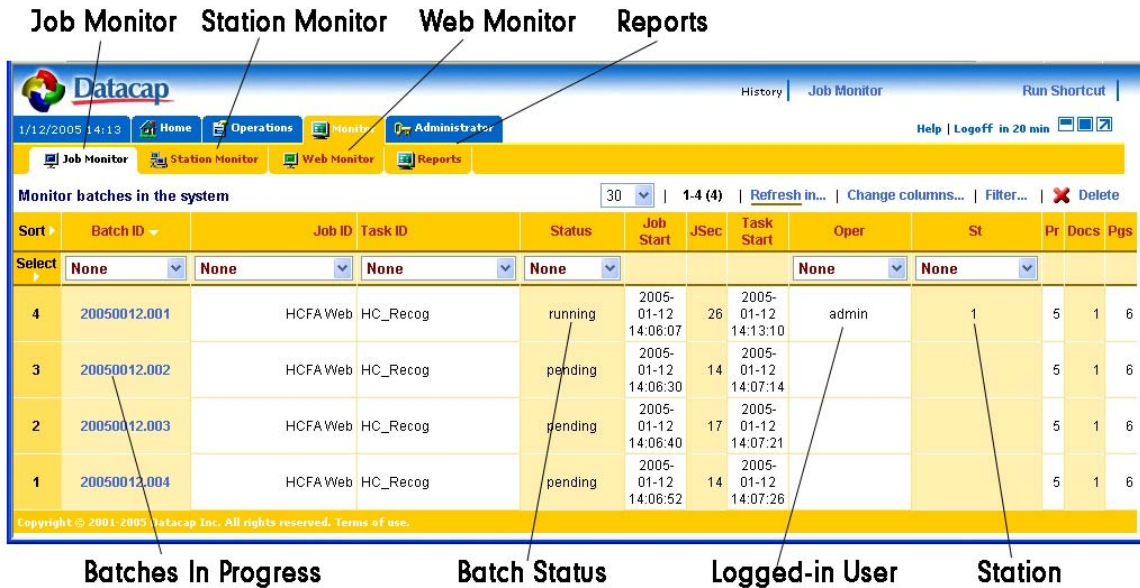


Figure 5 - Taskmaster Web Job Monitor

Another useful tool provided by Datacap for extra control over distributed capture is the Taskmaster Enhanced Notification Utility (ENU). This feature enables administrators to monitor Taskmaster for certain conditions, such as when a user deletes a batch, or if a batch is pending too long, and it notifies the administrator with an e-mail message. By using the Taskmaster Job Monitor, combined with ENU, those charged with managing a distributed capture application can maintain complete control, with full reporting and tracking capabilities, no matter where users are located and what tasks they're performing.

## Conclusions

Distributed capture is a way to reduce or eliminate the costs associated with the transportation of business paper documents. It also allows for faster processing of documents, better customer service, and reduced local storage. The indexing of data capture occurs most effectively where the office workers are located or where the centralized records and databases are maintained. The method chosen will depend on the types of records and operations in place.

The Internet has become the primary method to transmit the images, replacing FAX – the advantages being better quality images and faster transmission. Automated recognition tools have the ability to further reduce indexing and categorization costs for improved retrieval, and will be increasingly employed in the future.

**Distributed Capture Benefits:**

- Improved business processes
- Improved customer service
- Less space required for filing
- Less copying
- Improved, more rapid access to documents

## **Will Distributed Capture Work for You?**

### **A Checklist**

#### **You probably have a business that can benefit from distributed capture if:**

- You have to wait for paper to be received before you can work on the transaction.
- You receive paper at remote sites and pay to ship it to somewhere for centralized scanning.
- Any of your remote staff need to make photocopies in order to work on the papers while they are being shipped/ scanned.
- Any of your partners / agents deliver paper documents to you.

#### **What sort of distributed capture device should you use?**

You may be able to use existing equipment such as MFP digital copiers, but if more than 20% of your papers are double-sided, it is probably preferable to install specific duplex document scanners

- Are digital copiers installed and if so are they networked?
- What proportion of your paper is double sided?
- Look at the quality and weight of the paper that is to be scanned -- if it is poor quality or very light weight, plan on using a scanner and not an MFP.
- Volumes: if scanning a few hundred pages a day, an MFP may be perfectly adequate, but if scanning a thousand or more pages, a scanner is better
- Can you further improve processes by enabling mobile scanning, capturing documents while traveling, on a laptop, and delivering over mobile Internet connections?

#### **Where and how should the indexes be captured?**

Localized indexing works best if:

- Office staff have difficulty finding the papers they need and/or have to add indexes.
- Office staff get called by centralized indexing or data capture operators to get assistance

Centralized indexing (that can be automated with batch controls and OCR) works best if:

- The capturing location is a different company (e.g. an agent)
- Paper is delivered to remote sites due to geographic convenience
- Capture is via a mobile device

#### **Key elements in selecting a vendor**

- Select a vendor who has experience in distributing the capture and integrating with any centralized document management or content management system.

- Select a vendor who has understanding of your business processes. They will be able to understand your needs and manage the integration more effectively.
- Select a vendor who has the ability to automate indexes. This will reduce the need for extra effort by office workers who do not want to become scanner operators or key data entry operators.

### **About Datacap Inc.**

*Since 1988, Datacap Inc. has provided leading document capture and forms processing software solutions to organizations worldwide. Datacap Taskmaster software efficiently transforms paper into information, reducing data entry costs and improving information accuracy. A client/server, rules-based document capture workflow platform, Taskmaster provides highly flexible solutions for both image indexing and data entry automation. Taskmaster also enables scanning, indexing and administration from a browser and integrates with all leading document management solutions. <http://www.datacap.com/>*

### **About Harvey Spencer Associates**

*Harvey Spencer Associates (HSA) was founded in 1989 to provide consulting and analyst services to companies involved in the scanning, recognition and capture markets. Since then, many of the leading companies in the capture market have been clients of HSA. Harvey Spencer is a well known author and leading expert in the capture and forms processing markets.*

*Harvey Spencer Associates  
2 Penfield Drive  
East Northport, NY 11731  
631-368-8393*

[www.hsassoc.com](http://www.hsassoc.com)

