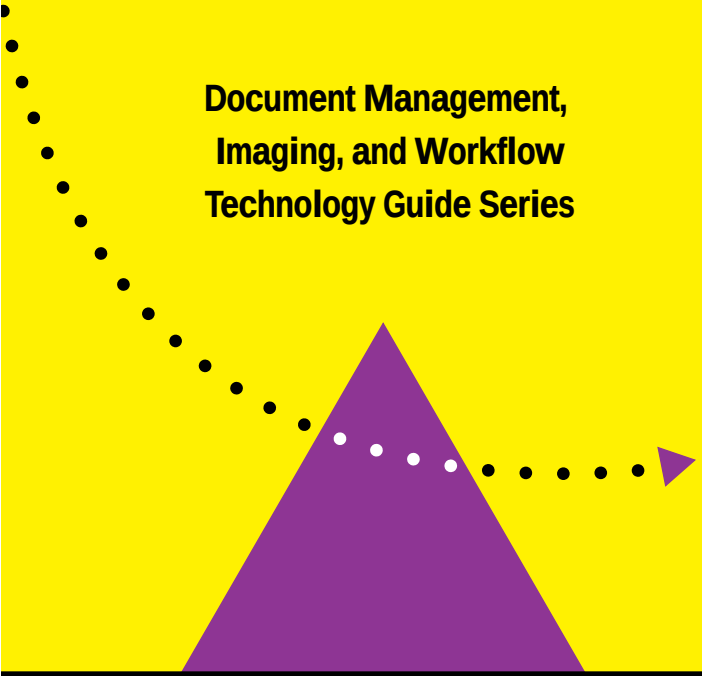
RHEINNER GUIDE

COLD Solutions

Document Management, Imaging, and Workflow Technology Guide Series



Visit ATG's Web Site
to read, download, and print
all the Technology Guides
in this series.

www.techguide.com

"The significant problems we face cannot be solved by the same level of thinking that created them."

Albert Einstein

Table of Contents

The Rheinner Group

The Rheinner Group is a leading research, consulting and education firm in the document imaging, management and workflow industry. Its Certified Document Imaging Architect Education Program, which covers many of the same issues addressed by The Rheinner Group's Technology Guides, is the most popular training program in the imaging industry.

For more information on The Rheinner Group, their education program course schedules, or to obtain help designing and implementing document imaging and workflow systems, please call 617-741-8100. Also, please visit their web site, The Rheinner Café, at www.rheinner.com.

The Rheinner Group is a proud sponsor of the Certified Document Imaging Architect (CDIA) certification program.

Introduction

Tremendous amounts of information, recording the details of credit card, check, ATM and other kinds of transactions in financial institutions and government organizations, are fed into mainframe computers every day. In addition, computers generate immense amounts of data, such as statements that are printed and mailed to customers. Since the expense of keeping this information on the mainframe forever is prohibitive, the information is archived and removed. It is either printed to paper or "printed" on a form of microfilm called microfiche, a process that is referred to as Computer Output to Microfilm (COM). Each year, companies spend over $1 million on COM products and services.

When data is archived using paper or fiche, it is inaccessible by computer. The data is returned to the pre-computer form essentially equivalent to the old ledger book. Retrievals of this information are also done in the old way, page by page. This may seem acceptable until you consider the numbers. With 15 million new credit transactions each day, Visa has to "roll off" an equal number of transactions to microfiche daily. Each month they put close to 500 million credit card charges on COM.

The cost of manually retrieving information from COM is high: One leading bank spends 30 minutes retrieving a canceled check record from microfilm, as contrasted with a couple of seconds if the information was on a computer. Though not ideal, COM retrieval is justified by the fact that as data on the mainframe ages, it is retrieved less frequently. The number of retrievals on any block of data eventually falls below an economic break point. And even though the cost of manual retrieval is extremely high, at this point the high cost of mainframe storage is not justified.

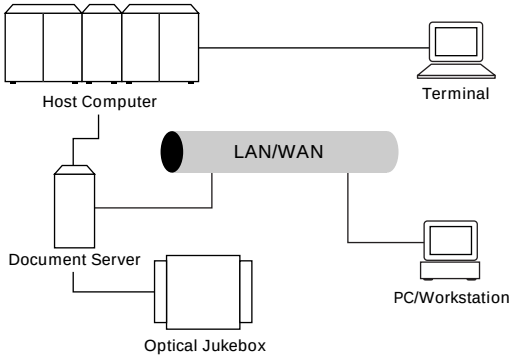
Using COM (and paper printouts) is like stepping back 40 years in terms of data retrieval. However, until recently, it has been the only acceptable solution because of its low cost. But the emergence of products, collectively referred to as COLD (Computer Output to Laser Disk), now provides an alternative to COM.

This Rheinner Group Technology Guide describes what COLD is, discusses the key benefits provided by COLD, and points out the major distinctions between COM and COLD. In addition, this guide explains the various ways COLD can be implemented, and outlines factors you'll need to consider when selecting a COLD system.

What is COLD?

COLD is an integrated set of hardware and software that takes computer-generated, formatted reports and, rather than printing them or sending them to COM, processes, indexes and stores them on computer media. This formatted output consists of point-in-time daily, weekly and monthly reports, such as general ledger or transaction history reports, as well as customer documents like statements and invoices. By taking report data that would otherwise have to be stored off-line and storing it on-line, COLD enables users to electronically search, view, print and process the information contained on any report.

Typical COLD Configuration



Most COLD systems operate in a similar way. Reports or statements are generated on the host computer, which may be a mainframe, minicomputer, or indeed any kind of computer which processes large amounts of data. At a certain predefined point in time, the reports are identified as ready for archiving and are transferred to the COLD system. The transfer may be accomplished by file transfer, by media transfer, or by the COLD system emulating a device attached to the host. After transfer, the report is identified, and lookups are performed to find the rules that determine how the report is to be processed. The report is then processed to extract all the index values that have been defined for it and an index file is built for the report. The report data is then compressed. After a number of reports have been processed, they may be further formatted with header information and transferred to the long term storage medium, which is most commonly optical disk, but may also be CD-ROM or tape.

The indexing and storage processing components of COLD are typically the most computationally challenging aspect of COLD, and generally determines the overall throughput requirements for the system.

With COLD, users have interactive screens that enable them to view the catalog of reports. They may access reports via the catalog, or they may perform searches to help them select the reports of interest to them. Having selected one or more reports, the user can then ask to see part or all of a report. Depending on the format in which the report was stored, specialized viewers may be required to handle bit-maps, special print formats, form overlays, etc.

Paper vs. Microfilm vs. COLD

Before the advent of COM or COLD, the only way to store the large amounts of data generated from computers was on paper—big stacks of printed reports. With this method, which is still used by many companies, reports are generated and printed for the data that exceeds the established time limit for storage on the computer. The data is then sorted by various key fields, for predetermined time periods (day, week, quarter, etc.), and then output in report format. Multiple reports are generated and printed for each sort order required to manually retrieve the data. In many cases, these stacks of printed reports are distributed to various locations where they are manually stored and accessed.

Since retrieval is often based on a single data element, a typical inquiry might require reviewing several reports with the same data, but in different sort orders. For example, a customer service representative may have the transaction amount but not the account number. In this case a printed report sorted by account number would be of no use. The required report is one sorted by amount that also shows account number.

COM provides a more advanced storage and distribution method. Instead of, or in addition to, printing the reports, a machine exposes the report on microfiche, a type of microfilm. After exposing and developing the masters, the fiche is duplicated and distributed to the various locations that need the information.

COM technology serves two purposes. First of all, COM is used to archive data. A single master copy is produced so that the computer space can be freed for other purposes. Anyone who needs the data goes to the place where the master copy is stored, or requests that the archive location make a copy and send it. COM is also used to distribute information. Fiche copies (duplicates) are made of the master fiche and sent to all locations that need to access the information.

Both the paper report and COM methods share a common problem: the data is no longer available on computer screens. To access off-line data, users must leave the normal work area, go to a local collection of microfiche or paper reports and search for the required information manually. This time-consuming process often requires the user to look through a multitude of partially indexed pages to find the needed data, and then copy the

information to a piece of paper, photocopy, or, in the case of fiche, print the data. This process is often impeded by missing fiche or reports.

In increasing numbers of organizations, microfiche is maintained in a central library. In such cases, a request form is filled-out and sent to the library. The microfiche request must be specific as to the report type and the data required. Librarians receive the requests, locate the fiche, and print the required frames. The information is returned to the requester, usually in one to three days. There are many variations to off-line data requests, but all create long delays in resolving the inquiry.

The long delays and increased costs associated with retrieving off-line reports in some cases justify leaving the report information on-line on Direct Access Storage Devices (DASD).

COLD Technology Solves the Problem

At the root of the problem solved by COM is the expense of DASD. DASD is designed to read, update and write random blocks of gigabytes of data extremely fast, in only tens of milliseconds. The speed and random access of DASD permit large databases to be built with multiple indices to each data record. Many inserts, rewrites and reads can take place on multiple indices simultaneously. While the price/performance of DASD improves every year, it is still nowhere close (by roughly a factor of 100) to the price point that would enable all the data to be kept on DASD rather than being off-loaded to COM.

But DASD is not necessarily required to provide acceptable retrieval times for this historic transaction data. Most often this historical transaction data has very low access rates and is never updated. If an error is found, a new transaction is generated which corrects the error. A storage medium far slower than DASD that permits a single write but multiple, random reads can be just as effective. Many companies have recognized that WORM optical media meets this description. Unlike DASD, optical platters are removable from the drives that read them. Therefore, it is possible to build autochangers or jukeboxes that exchange platters between the shelves that hold one or more optical drives. In some cases, CD-ROM or tape may also provide suitable storage media.

COLD technology allows users to keep data on-line forever. Instead of taking the data off-line as is done with paper and COM, COLD changes the way the data is maintained on-line in order to achieve the cost requirements for on-line storage of such large amounts of data.

A Comparison of COM and COLD

CATEGORY	COM	COLD
REPORT GENERATION	Batch jobs sort old data and generate printed reports	Same, except redundant reports are eliminated
REPORT EXTRACTION	"Print" to COM machine	System-managed transfer to server
STORAGE	Expose, develop master microfiche	Automatically index, write to optical disk
DISTRIBUTION	Make copies, physically move, store in trays	Read data from optical and send to user when requested
ACCESS	Person goes to trays, finds fiche, uses viewer	View on-line on standard terminals like other computer data
SELECTION	Hunting through reports by search order	By on-line query
HARDWARE	COM machine, developer, duplicator, storage trays, viewers	Shared-use server, existing terminals and workstations, optical disk and jukebox
SECURITY	Only physical	Standard MIS: by user, by document and by field type

Benefits and Advantages

Improved access to information is the number one benefit provided by COLD. With paper or COM, you have to physically go to where it is stored, find the paper or fiche you need by some means of indexing, possibly

makes copies, and then return to your desk. With COLD, you find what you need by browsing a catalog or performing a search, and then retrieve the report to your screen, all within a small number of seconds. The multiple automated index methods make it easier to identify information and access to information is nearly immediate.

With COM, there is usually a time delay from sending the reports to fiche production until they are available for use. While some organizations pay a premium to minimize this time lag, the production costs are so high that many organizations accept long delays until the information is available on fiche.

Cost Savings with COLD

The immediate and improved access to information is beneficial enough to warrant switching from COM to COLD, even if COLD were more expensive. But, in fact, COLD normally costs less than the COM services that it replaces and can generally be justified on the basis of this savings alone. Many direct and indirect costs associated with COM are eliminated with COLD. For example:

- microfiche generation, duplication, and distribution costs are eliminated;
- valuable floor space is recovered by eliminating microfiche libraries and reader/printer devices;
- the costs of maintaining the microfiche libraries and devices are eliminated;
- the time limit for on-line transaction data can be decreased, making more DASD available;
- and delays in data availability are eliminated, reducing or eliminating customer service call-backs.

Substantial savings also result from decreased staffing requirements. With paper or fiche, labor is associated with moving data to the COM machine, producing the fiche, and transporting the fiche and duplicates to their storage location. And every time a new batch of fiche are produced, someone needs to integrate it with the existing store of fiche.

User Productivity Gains

Users need fast access to the data necessary to perform their jobs. With fiche, they either have to fill out a request form and transfer it to a central fiche department, or they have to get up and look themselves. In either case, someone is spending time looking for information. By having a computer do the looking, people spend less time and are therefore more productive.

Improved Customer Service

With COLD, companies are able to provide better customer service. Most importantly, a customer service representative is much more likely to be able to help a customer right away with a COLD system. If the customer service representative cannot leave his seat, find the fiche and get back in less than a minute, he will be unable to answer the customer's question while the customer is still on the phone. This means a call back and potentially a less satisfied customer. With COLD, the telephone operator is likely to be able to find the required information right away, discuss it with the customer, and make any required changes or updates to the on-line system in a single transaction.

Statements represent a very visible interface between an organization and its customers. Requesting new copies of old statements is a frequent customer service request. Since microfiche is a form of microfilm, optical/mechanical means are required to make copies from COM fiche. Using analog copiers, the image produced is not nearly as attractive as computer laser printer output. If the fiche is digitized using one of the latest digitizing fiche copiers, the result is improved, but resembles a fax copy rather than an original.

With a COLD system, on the other hand, statements

can be made to look as good as the originals. This is accomplished either by storing a computer-generated bit-map image of the printed page, or by storing the codes used to create the page. In any case, a data stream is generated and given to a production laser printer, which then prints copies for customers that look as good as the originals. Automated faxing of excellent copies can also be achieved using this technology.

Better Information Distribution

Geographically-dispersed sites often need access to report information. By the time fiche duplicates reach their destination, the information may be out-of-date. With COLD, remote sites can be given access to the COLD system just like any other computer system. The same communications media used to transport other computer-based transactions within the organization can provide distributed access to a centralized repository of COLD information. It may also be possible to distribute subsets of the COLD archive to remote servers, if the access to the reports is likely to be heavy.

COLD is Legally Acceptable

Legal acceptability has been a long-standing question in the world of optical disk. However, most experts agree that as long as the information is stored as a part of normal business practices, information stored in a COLD system on optical disk is just as "legal" as information stored by other means. In a number of high profile cases, the SEC and other administrative agencies in charge of an industry segment have issued specific regulations that permit the use of media other than microfilm for storing legally required information. Even in cases where no such ruling has been issued, COLD is now a widely used and accepted technology.

Data Security

Security for the confidential information stored on paper reports or on COM fiche is purely physical. Anyone who walks by the area where the fiche are stored can walk off with reams of sensitive information in his pocket. If an outside service is producing the fiche, they also have unsecured access to this confidential data. COLD eliminates this problem by putting the storage into the computer room. Access to the data is controlled by normal MIS security mechanisms.

COLD Offers a Compelling ROI

All of these benefits add up to a compelling cost justification case for COLD. The bottom line is very simple. Forget about "soft" savings. Forget about happy customers who think better of you because you can send them great looking material post haste. Simply look at the dollars. Most COLD systems have a payback period of less than two years. And many pay back the costs of implementation in a matter of months.

A COST COMPARISON—COLD vs. COM vs. DASD

PRODUCT ATTRIBUTES	COLD	COM	DASD
Processing and Distribution Time	Near-immediate	Days to weeks	Near-immediate
Cost of Copies	None	$.04 - $.10, plus labor and distribution	None
Cost of Master	$500 per 10 GB disk, about $.05 per megabyte	$.30 - $2.00 per sheet, roughly 1 megabyte	None
Capital Cost	Under $.40 per on-line megabyte	$200,000 or more per machine	$1.00 -$10.00 per megabyte
Labor per Retrieval	Seconds	1 to 30 minutes	Seconds
Elapsed Time per Retrieval	Seconds to Minutes	30 minutes to days	Seconds
3270 Access	Yes (some products)	No	Yes
Access by PCs & Workstations	Yes (some products)	No	Yes

Functional Elements

COLD systems perform a variety of tasks, including:

- Processing and generating reports with the aged computer data. The COLD system should provide tools that allow the administrators to easily identify the reports to be stored.
- Preprocessing report data for efficient storage on the selected storage medium. Software should process and stage the data, along with the indices. Reports also have to be filtered to identify and extract index fields during processing.
- Building the report index. The process of building the index is an important part of preparing the report for storage. For each report, the indices that have been specified have to be found and extracted from the report. There may be as few as a couple of index fields for a long report, or several fields could be extracted from each line of a lengthy report, depending on the retrieval requirements. Indices from multiple reports are often merged prior to being written to a tape or disk.
- Setting up reports. An administrative package typically permits report processing and access functions to be specified and performed without programming. There should be a master catalog of reports and an index to the contents of the reports on all volumes. Using interactive software, the administrator can establish all index categories for each report and define how the values are extracted from the reports. Access controls are established on the reports after storage.
- Transferring report data to the storage subsystem. Typical transfer methods are either tape or network communications. The ability to automate this process, along with logging and error recovery, are key features of production COLD systems.
- Storing the report data on the selected media. COLD systems must include all the software required to compress and store the data and index information on the selected media. The compression function is an important factor in using disk space efficiently. Storage must be performed at rates sufficient to handle the volumes produced by the largest users. All the data should be staged on server magnetic disk prior to being written on the actual storage media. It should be possible to schedule the processing and writing to optimize the server during normal working hours. Finally, there should be controls on exactly what is stored, so that the user can control whether the text of the report is stored, the image (bit-map) of the report stored to enable exact reproductions of the report to be generated, or both.

  The actual storage devices supported by the COLD system are important selection criteria. Sharing the optical storage with other applications may be very important for smaller systems; for larger systems that are likely to fill the available storage quickly, sharing is not as important.
- Intelligent printing. The "classic" COLD situation involves printing reports to line printers that support a single, mono-spaced font. However, many reports today are sent to printers which are loaded with pre-printed forms. Additionally, important classes of output such as customer statements are sent to intelligent printing devices such as laser printers. The data stream that is generated contains control characters, font changes, position changes, references to overlays that have been downloaded to the printer, and other complicating factors. The COLD package should be able to accept reports in such complex formats and still filter out the index values and other text. In addition, the COLD system should be able to store formatted reports (typically in IBM AFP or Xerox Metacode format)

in such a way that they can be reprinted and retain the look and feel of the original.

- Media management. The software should maintain a comprehensive catalog of all volumes that contain information. This permits information to be found automatically regardless of the volume or volumes on which it is contained.
- Distribution of reports. People at different offices may need to view a report. In addition, the business process may require that people work on certain reports as soon as they are available. Therefore, COLD software often manages the report distribution function, maintaining lists that match report types, people who need them, and the means best used for transferring the report to the recipient.
- Retrieving and displaying data. The people who need to access report data must be provided a means for doing so. The appropriate methods depend very much on the hardware and software that people already use. If people are connected to mainframes, access programs that can be loaded into the users' CICS (mainframe) region must be provided. Similarly, functionality should be available for PCs and/or UNIX workstations. The access program interacts with the user to determine the information the user wishes to see. Making a request should resemble a normal database retrieval operation: the user fills out a screen with the known variables such as date and account number. The information request should then be sent to the storage server or host computer. The MIS department should be able to customize the retrieval application to meet various user needs.

  When requested by the user, one or more pages of computer output is sent by the server or host to the client workstation or terminal. The user should be able to view any requested pages of information at any time, search for text strings, print, store as a local file, etc. The user should be able to search within or across reports based on the indices that were built during the storage process. The user should also be able to search for information that was not indexed in advance, at the cost of longer search times. The software should be able to adapt to whether the user's terminal is able to display images, and whether the report is stored in text format, image format, or both.

Implementing COLD Solutions

From an architectural perspective, there are several ways to build COLD systems. Since all the data in question generally comes from a mainframe and the people who need to access the data have access to the mainframe, one of the earliest COLD approaches was also a mainframe approach. These systems accept mainframe-generated data directly, write it to optical disk, and deliver the data to the terminals attached to the mainframe. However, just as in other areas, COLD buyers are interested in moving away from mainframe-based COLD solutions to those based on a client/server architecture.

Client/server systems are ones in which central functions are performed at a server, and user-specific functions are performed on a client workstation. Client/server COLD systems are scaleable and generally suitable for enterprise use. These systems include a client application that communicates via remote procedure calls (RPC) to a server component. The server component, which generally runs on a UNIX or Microsoft Windows NT platform, provides security, optimal network utilization, scalablity and central administration. Application Programming Interfaces (APIs) are used to allow other applications to access the information and control the server component. Ideally, a client/server COLD system will support rich connectivity to the legacy systems environment, both for download, and report viewing and retrieval purposes.

Document Imaging COLD Solutions

Many document imaging vendors have realized that their document imaging products could be adapted to store computer-generated report data in addition to document images. The computer data is transferred to the imaging system and stored on optical platters. A record for each report is entered into a DBMS that stores the index of all documents in the system. The imaging system then permits browsing through the catalog of all documents and recalling and displaying any page of any report.

The benefit of this approach is that it makes use of equipment and software that may already be installed for document imaging purposes. In addition, image and COLD data are integrated into a consistent electronic file folder metaphor, making it easier for users to interact with the system.

This approach is attractive for storing low-to-medium volumes of paper that are likely to be used within the context of the imaging system. However, the COLD information can only be indexed to the page level since indexing down to the record level would result in the database growing to a massive size. This can result in more time-consuming retrievals. Further, only people with workstations suitable for document imaging can easily access the COLD data that is treated along with document imaging. In many cases, a larger population of users needs access to this information.

Stand-alone COLD

Some vendors have put together the components to store and index the data without the overhead and constraints of document imaging COLD. Most of these are PC-based, although there are a couple of UNIX-based products on the market. The products are sometimes able to extract indexing information from the reports and automatically generate a record-level index during a preprocessing step. The report and the index are then both written to the optical disk. Software is provided which permits key fields to be entered. Both the index look-up and the data are then extracted from the optical platter and displayed on the connected terminal.

The benefits of this approach are that it provides both detail-level indexing and storage of everything on optical platters. Additionally, the cost is low. The disadvantage is that these systems are scaled for small volumes and are usually purchased by organizations with small output volumes that don't even warrant a COM-based solution. Also, the systems are often isolated from the rest of the organization's computer systems and networks, which limits the number of users who can retrieve information.

Integration Issues

The basic functionality of a COLD system can be implemented in a number of ways, depending on the operational context and the other computer systems with which it must be integrated.

Many COLD systems are implemented within the department that requires access to the information, essentially in isolation from other systems. After the COLD system has been installed, the members of the department go to their computer screens, access the COLD retrieval module, and use it to meet their report accessing needs. The COLD interacts with other systems in the sense that it shares a computer and network with them, but nothing else.

There is growing demand for integrating COLD with document imaging because people want to have one place to go where they can find all information about a subject, such as a particular customer, regardless of whether the information was generated by a computer or received in paper form and converted to an image document. The ideal thing is to have a document cataloging system that is able to access reports stored in a COLD system, as well as images stored in an imaging system so that related COLD data and document image data can be retrieved.

COLD systems can also be integrated with workflow systems. One predominant method of integration is to use COLD as an archive of report documents needed to process a specific case, in the same way that an imaging system is used as an archive of document images that may be needed to resolve or process a particular case. Another more powerful method of integrating COLD with workflow is to feed certain reports into the workflow system for action and processing. A good example is a report on overdue accounts. This report can be fed into the workflow system, and each overdue account can be turned into a case, integrated with all the other information relevant to collecting that overdue balance. The workflow system can then route the cases, assign them to collectors and track the results.

Operational Issues

Before implementing a COLD system there are a few significant operational issues that must be considered and resolved. The first is the choice of a report storage format. There are five options:

1) The raw character stream generated by the report program, suitable for printing on a line printer. The typical data stream on a mainframe consists of 133 characters per line, with a control character in the first position to control line advance. Coding is EBCDIC. The typical data stream on other computers consists of variable length line terminated by LF and/or CR codes, with line advance and page breaks determined by control codes mixed in the data stream.

2) The characters and format that comprise the character stream generated by the document formatter suitable for printing on a bit-map printer. On mainframes, this is typically AFP or Metacode, depending on whether the target printer is IBM or Xerox. There may be various fonts loaded in the printer, and there may be various page formatting codes at various points in the data stream. On other computers, the typical code sequences are HPCL or Postscript.

3) The character stream, along with a reference to one or more bit-map overlays, which serve the purpose of a pre-printed form. The overlay may be created from the formatted data stream, or it may be created separately in an off-line operation. This method is often used in document imaging-based COLD systems.

4) The compressed bit-map that was generated by the printer's formatter, suitable for reproducing the exact appearance of the original printed page. This is normally generated by software which emulates the operation of the printer's formatting engine, and is compressed into a sequence of TIFF files or something similar.

5) A universal format supported by a major company, such as Acrobat by Adobe, which provides a hybrid solution. The characters and formatted data are converted into Acrobat PDF format at time of storage, and interpreted by Adobe viewer software at time of reading.

Your choice of format is to a great extent dictated by your choice of COLD supplier. All COLD systems support at least one of the above formats; some support more than one. The ideal format is one that requires a minimum of storage space, provides a robust ability to search the report, and allows the re-creation of the original document. Your choice should be driven by the nature of searching and retrieval done on report information, as well as the need to reproduce the original document, for example if customers frequently request duplicate copies of statements.

Report Setup

The other major operational consideration is setting up each of your reports for processing. Many users underestimate the time and steps involved in setting up and administering the COLD report process. Report setup consists of:

- Providing for the transfer of source files to the COLD system, including error recovery;
- Identifying various report types;
- Identifying the method of report-level indexing and pagination for each report type, as well as the method of extracting the indexes from the report. This must also be done for page and line-level indexing;
- Providing for report storage management, index setup and migration.

Determining the appropriate level of indexing for COLD reports is one of the more challenging aspects to report setup. Users generally want a fine level of indexing that extends to each line of a report. However, this amount of indexing can be expensive, both in terms of compute time and space requirements.

Indexing criteria should be based on the age of documents and should be applied conservatively. If possible, index recently-generated documents to the page or report level, rather than to the line level. Older documents may only require report-level indexing depending on the frequency and nature of the retrieval requirements.

Summary

Companies in virtually every industry from banking and financial services to restaurant chains are realizing the near-term productivity, cost-savings and enhanced customer service benefits that result from implementing COLD as a replacement for paper or COM output of computer-formatted report information. With a compelling cost justification case and mature client/server solutions that support broad retrieval, storage and output alternatives, as well as access to other relevant documents and images, it is not a question of whether COLD makes sense; it's how quickly you can implement the technology.

Case Studies

Acordia of Colorado, MIS Department

Using FilePower® Report Management System, FPreport™ to streamline business processes and boost efficiency and quality of internal accounting procedures.

According to Walter F. Ahern, Acordia of Colorado's CEO, Acordia values its information technology systems second only to the customers it serves. And to provide those customers with the highest quality property/casualty insurance, employee benefits and individual insurance, and financial services programs, Acordia turned to Optika for a state-of-the-art document image and information management system.

The insurance agency business requires handling a great deal of paper, such as policies, master plan documents, correspondence, accounting reports and various other documents. This paper can bog down commun-ications with customers and overall organizational effectiveness. Acordia believes that innovative computing systems technology will allow it to better serve its customers and outshine its competitors. And to help it win the race to stay ahead, the insurance agency is implementing an information management system to provide on-line access to critical business documents and data. The solution is based on Optika's FilePower family of products, including FPreport, Optika's on-line report management system.

Out With the Old

Before installation of the system, end-of-month accounting functions were performed manually, requiring accounting staff to manage and sift through paper versions of reports generated by the company's IBM® AS/400® systems. The reports required many boxes of greenbar paper for each month-end cycle.

Acordia's accounting function utilized approximately 25 different varieties of production reports and more than 20 specific end-of-month reports, such as an accounts payable report, accounts payable register, accounts receivable register, zero balance register, and customer statements.

Accounting staff would review the report information and manually re-enter report data into Microsoft® Excel®. Data was extracted by "producer" (sales representatives), by month and by business company, a time-consuming, manual process that would take approx-imately one and a half weeks to complete. After extracting the report data, the Excel spreadsheets were then distributed to more than 20 Acordia producers for review.

...And In With the New

With Optika's new FPreport COLD system in place, the month-end accounting process has been automated and simplified and a multitude of users have immediate access to these reports and other accounting information.

All internal reports are now written to optical disk and indexed on the FPreport system. Printing on paper has been eliminated, saving computer time and labor associated with printing, packaging and physical distribution. The automatic "data extraction" feature of the Optika COLD system allows users to extract sales data from various reports and consolidate it into a Microsoft Excel spreadsheet. Automatically creating this single sales report saved more than a week in what had been a slow re-keying process. It goes without saying that this critical information is now available one week sooner.

In addition, the manual filing of accounting reports was eliminated, saving labor, errors and printing costs in the process. The reports that were produced for corporate management and sales staff previously took three days to complete now take advantage of an automated process and are available in minutes. Walter Ahern's goal of improved service at reduced cost through the use of information technology has been achieved.

Acordia of Colorado's system utilizes Intel® Pentium®-based servers running Microsoft Windows NT™ and Microsoft SQL Server™ for database management with Intel 486-based workstations. Report data and images are stored on a Hewlett-Packard® 20XT jukebox and are printed and faxed using Optika FPprint™ and FPfax™ servers. This configuration will allow Acordia to store five years of historical computer output and make it available electronically to Windows-based workstations.

Observations

Kathy Francis, MIS Supervisor for Acordia of Colorado, observes, "With the FPreport system, we can cut this week and a half long procedure down to two or three painless days. We've also eliminated keying errors and can now perform on-line research. In addition, Optika's integrated product suite will allow us to marry report information with scanned images and other documents and use workflow to coordinate and manage our business processes."

The Optika system was initially installed in the benefits department to handle the extremely heavy paper volume requirements. "Although we have only a partial implementation and the time period has been short, our first quarter 1995 results showed a 36% gain in revenue per employee over the same period last year," said Ryan Hettich, Acordia of Colorado's CFO. "I expect similar results as we bring the rest of the organization on-line and continue to improve how we serve our customers."

Future Plans

The next step in the evolution of Acordia's system is to use Optika's FPmulti™ document imaging software to scan and index documents that are critical to the accounting process. This implementation allows all documents, computer-generated and scanned, to be retrieved from a single database, allowing the company to have a true "central information file" for client documents and internal reports. After archiving these documents, Acordia will use Optika's PowerFlow™ workflow management software system to automate the flow of critical business processes for even greater productivity enhancements.

Janney Montgomery Scott

Using FilePower® Image Management System, FPimage™ and FilePower® Report Management System, FPreport™ for a document storage and retrieval application.

In the normally conservative world of high finance, one wouldn't necessarily expect to find a 162-year-old firm aggressively embracing new technology. But two years ago, Philadelphia's oldest and largest full-service securities broker-dealer, Janney Montgomery Scott (JMS), did just that, adding a state-of-the-art document imaging system to its operations.

Founded in 1832, Janney Montgomery Scott employs nearly 1,200 people, and has a tradition of providing exceptional financial services to customers in the New England and mid-Atlantic regions. Recognizing that the use of imaging in the securities industry was inevitable, Jim Wolitarsky, JMS' chief financial officer, began evaluating systems long before most other broker-dealers did. "Our manual records got the job done, so we didn't really have a problem to solve, but it was clear we had an opportunity to become more efficient," Wolitarsky explained. Like most innovators, JMS recognized that the long-term benefits of computer-based imaging far outweighed the potential risk of being an early adopter of imaging technology.

What makes Janney Montgomery Scott's commitment to imaging even more impressive is that when it undertook the project, few other broker-dealers had employed the new technology, and the Securities and Exchange Commission had not yet officially recognized the use of optical technology for records' storage. Due in part to involvement by the Securities Industry Association, the SEC issued a letter in June of 1993 stating that it would "take no action" against brokerage firms which stored records on optical media, provided they complied with requirements regarding security, auditing, backup and other matters. Although not an official endorsement, this action by the SEC allowed broker-dealers to decide for themselves if imaging would meet their document retention needs. JMS' system, already in use for nearly 1½ years at that time, required no changes to meet the SEC's guidelines.

New Approach Leads to Change

To stay in compliance with Securities and Exchange Commission regulations, broker-dealers like JMS are required to maintain complete customer account files and to generate and maintain a paper trail for each financial transaction. The guidelines for paperwork creation, filing and archiving are strict. In addition, JMS produces a variety of forms and reports for their own internal use. Because documentation is generated whenever a new account is opened or a change is made to an existing account, thousands of pages are processed each day.

In the past, after each transaction, a JMS employee would provide completed forms to the central records room, where the documents were microfiched and filed as necessary. Although the volume was manageable, JMS was faced with the space consideration inherent to any paper intensive business, and the poor performance of a fiche-based approach. Shuffling paper between depart-

ments, filing cabinets and off-site storage was slow, inconvenient and expensive.

JMS turned to imaging, utilizing a solution centered around Optika Imaging Systems' FilePower software. The system was configured and installed by Moorestown, New Jersey-based integrator Judge Imaging Systems. A Novell 3.11 network supports over fifty users including about a dozen imaging workstations using a 10-Base T implementation of Ethernet. Scanning is accomplished using two Fujitsu scanners, one 3096 and one 3093. Images are stored on a Plasmon RF50JM fifty-platter jukebox.

Learning By Doing

The project was divided into multiple phases to provide for gradual implementation. The first phase involved record retention and retrieval, and although somewhat simplistic, it allowed JMS to learn by doing. Electronic folders were created for customer documents using Optika's FilePower management system. JMS began scanning historic documents for the previous two years' transactions. Eventually, new client information and daily transactions were added to the mix to keep the records current. Today, JMS is adding an average of 2,500-3,000 documents per day to the over 600,000 already available on-line.

Even during this introductory period, substantial benefits were realized. Because client records are accessed daily by individual brokers, accounting, and many other operational departments, the use of paper and microfiche resulted in delays of a few minutes to a few hours if the records were on-site, and several days if stored in a warehouse. Today, information is available in seconds at retrieval stations located in individual departments.

The second phase involves the use of Computer Output to Laser Disk (COLD). Currently, thousands of pages of reports are produced by JMS' outside service bureau on a nightly basis. Some reports may only be used for a matter of days before being discarded. By linking to an off-site mainframe and "printing" the reports directly to optical disk, JMS will be saving substantial printing and fiche costs. Users will have instant access and searching capabilities from their desk. And because the Optika system will operate on the users' existing PCs, the primary additional hardware costs are only those needed to expand the network.

"During phase one, it was difficult to quantify the benefits in terms of dollars and cents," Wolitarsky explained, "but during phase two, the payback is readily quantified in terms of cost savings for printing and fiche production."

One immediate payback for JMS management was how quickly everyone involved with the imaging project realized that a good decision had been made. Even the people who had no prior experience with computers were able to readily learn Optika's Windows-based system.

With phase two of the imaging technology project now under way, Janney Montgomery Scott has already begun planning for phase three. The details? JMS won't say. But you don't get to be 162 years old without taking intelligent risks designed to meet the needs of your customers in an ever more efficient manner.

Glossary

Access method—The technique or the program code in the operating system that provides input/output services. It defines where a group of data will be stored on a medium. By including the access method in the basic operating system, computer makers have made the programmer's job much simpler.

Address—Disks and other storage devices have numbers that identify locations by sector and by byte, like addresses on a city street. Retrieval software searches for the address assigned to the desired information in order to locate it.

Addressability—The ability to place information at a certain chosen area in an image.

Addressable capacity—The number of locations on an image that are addressable. To calculate, multiply the addressable vertical positions (row) by the addressable horizontal positions (column). Think of a matrix of dots, eight across by 16 down. The addressable capacity of the matrix is 128.

Allocate—To reserve the required amounts of a resource, such as disk space.

Alphanumeric—Set of characters composed of letters and numbers; may or may not include punctuation marks and other symbols; excludes printer control characters such as Carriage Return and flow control characters such as XON and XOFF.

Alphanumeric COM—Computer output microfilm which is limited to receiving and recording letters, digits and punctuation characters. Cannot handle raster or vector graphics.

American National Standards Institute (ANSI)—A standard setting, non-governmental organization, which develops and publishes standards for "voluntary" use in the United States. Standards set by national organizations are accepted by vendors in that country.

American Standard Code for Information Interchange (ASCII)—The most popular coding method used by small computers for converting letters, numbers, punctuation and control codes into digital form. Once defined, ASCII characters can be recognized and understood by other computers and by communications devices. ASCII represents characters, numbers, punctuation marks or signals in seven on-off bits. A capital "C", for example, is 1000011 while a "3" is 0110011.

Analog—Comes from the word analogous, meaning "similar to." Analog devices record or monitor real world happenings, motion and sound, for instance, and convert them into "analogous" electronic representations, i.e. film or audio tape. Analog means recreating the continuous nature of the original thing. It is the opposite of digital, which translates the original happening into ones and zeros—an "unanalogous" representation. Analog information cannot be interpreted by a computer.

Application—A broad and generic term for any software program that carries out a useful task. Word processors and graphics programs are applications.

Application framework—A set of objects that provide packaged functionality and programming interfaces to accomplish specific tasks.

Application Program Interface (API)—Generic term for any language and format used by one program to help it communicate with another program. Specifically, an imaging vendor can provide an API that enables programmers to repackage or recombine parts of the vendor's imaging system, or integrate the imaging systems with other applications, or to customize the user interface to the imaging system.

Architecture—Refers to the way a system is designed and how the components are connected with each other. There are computer architectures, network architectures and software architectures.

Archival quality—The extent to which a reproduced image will (or won't) last "forever."

Archive—A copy of data on disks, CD-ROM, mag tape, etc., for long-term storage and later possible access. Archived files are often compressed to save storage space.

Attribute—In graphics, the condition a font is in—i.e. boldface, italic, underlined, reverse video, etc. is its attribute. In MS-DOS, files can be assigned attributes that define how accessible it is, i.e., "read-only" is a file's attribute. In a document retrieval system, an attribute of a file is one of the keys by which the document has been stored and indexed.

Audit trail—Record of activity that has occurred in a certain file, or on a certain computer.

Automated retrieval—Using a computer to identify and locate a stored image of some kind. Generally requires the use of key words or codes in an indexing scheme.

Bit—Contraction for Binary digit. The smallest unit of data a computer can process. Represents one of two conditions: on or off; 1 or 0, mark or space; something or nothing. Bits are arranged into groups of eight called bytes. A byte is the equivalent of one character.

Bit map—Representation of characters or graphics by individual pixels, or points of light, dark, or color, arranged in row (horizontal) and column (vertical) order. Each pixel is represented by either one bit (simple black and white) or up to 32 bits (fancy high definition color).

Bit-mapped image—Representation of image data where each pixel has a corresponding memory element. See also digital image, image, memory and pixel.

Bit mapping—Creating rectangles over documents, mostly white=zeros and black=ones, about one million spots per page.

Bit-mapped font—A set of dot patterns that represent all the letters, characters and digits in a type font at a particular size.

Bit-mapped graphics—Images which are created with sets of pixels, or dots. Also called raster graphics. Contrast with vector graphics.

Case—An individual instance of work to be performed for a business process; it can consist of one or more folders, documents and forms.

Client/Server—The relationship between machines in a communications network. The client is the requesting machine, the server the supplying machine. Also used to describe the information management relationship between software components in a processing system

Compact Disk (CD)—A standard medium for storage of digital data in a machine-readable form, accessible with a laser-based reader. CDs are 4¾-inches in diameter. CDs are faster and more accurate than magnetic tape for data storage. Faster, because even though data is generally written on a CD contiguously within each track, the tracks themselves are directly accessible. This means the tracks can be accessed and played back in any order. More accurate, because data is recorded directly into binary code; mag tape requires data to be translated into analog form. Also, extraneous noise (tape hiss) associated with mag tape is absent from CDs.

Compact Disk Read Only Memory (CD-ROM)—A data storage system using CDs as the medium. CD-ROMs hold more than 600 megabytes of data.

Compression—A software or hardware process that "shrinks" images so they occupy less storage space, and can be transmitted faster and easier. Generally accomplished by removing the bits that define blank spaces and other redundant data, and replacing them with a smaller algorithm that represents the removed bits.

Computer Output to Laser Disk (COLD)—Technique used to transfer computer-generated output to optical disk.

Computer Output Microfilm (COM)—The process of converting data (having been input by a number of means) to microfilm or microfiche.

Computer readable—Data which is in a format, such as ASCII, or on a medium, such as disks, tapes, optical discs or punched cards, that a computer can understand.

Data compression—Reducing the amount of electronic "space" data takes up. Methods include replacing blank spaces with a character count, or replacing redundant data with shorter stand-in "codes." No matter how data is compressed, it must be decompressed before it can be used.

Data decompression—The regeneration of a bit-map from a compressed representation.

Data set—A collection of related data. Usually refers to the part of the data to be viewed, but can also include indexing information, commands, printing parameters, etc.

Database Management System (DBMS)—Set of programs designed to organize, store and retrieve machine-readable information from a computer-maintained database or data bank.

Decompress—To reverse the procedure conducted by compression software, and thereby return compressed data to its original size and condition.

Descriptor—The key word, code or phrase that an automated document retrieval system uses to identify and locate the document. Descriptors sometimes "summarize" the most relevant data in the document, so that reading the descriptors—rather than retrieving the entire document—is sometimes sufficient for the purposes of the search.

Device drivers—Programs that tell the computer how to communicate with particular peripheral devices.

Digital—The use of binary code to record information. "Information" can be text in a binary code like ASCII, or scanned images in a bit mapped form, or sound in a sampled digital form, or video. Recording information digitally has many advantages over its analog counterpart, mainly ease in manipulation and accuracy in transmission.

Direct Access Storage Device (DASD)—Any on-line data storage device. A disc, drive or CD-ROM player that can be addressed is a DASD.

Field—The smallest logically distinguished unit of data in a record, as in "There are 12 fields in that record." In a database, the individual items of related information, for example, policyholder's name, address, social security number, etc. "Logically distinguished" means that there are similar units of data in other records that have something in common. For example, "last name" is a field, an entire mailing address is a record. All of the address records comprise a database.

Field separator—The prearranged code, typically a comma, that separates fields in a record. Also called a delimiter: "The records in that database are comma-delimited."

File—All the data that describes one document or image, maintained under a single naming code and stored in a computer or in a storage medium

File server—Local Area Network (LANs) were invented to allow users on the LAN to share and thereby conserve the cost of peripherals (printers, modems, scanners) and to likewise share software. The file server is the machine on the LAN where the shared software is stored.

Folder—(1) A term for the basic element in its file management scheme. A folder holds sets of files. A folder can hold other folders. It is basically a hierarchical tree-directory scheme, just like DOS's directories and sub-directories. (2) A logical collection of electronic documents stored on the document management system.

Formatted data—Data which has been processed with software to attach the necessary titling, indexing, and job separation instructions.

Gigabyte (GB)—A million (actually more) bytes of data, or a thousand megabytes. Imaging applications commonly take up huge amounts of data. For example, it only takes tenb 8½ by 11-inch color pictures, scanned at 600 dpi, to fill a gigabyte.

Gray scale—The spectrum, or range of shades of black an image has. Scanners' and terminals' gray scales are determined by the number of gray shades, or steps, they can recognize and reproduce. A scanner that can only see a gray sale of 16 will not produce as accurate an image as one that distinguishes a gray scale of 256.

Index—A descriptive set of data associated with a document for locating the document's storage location. In a more complex and demanding role, indexing can be used to consolidate documents that may not be, at first glance, related, or that may be stored in different locations, or on different media. Indexing stored documents is the great intellectual challenge in document retrieval. Anyone can scan a piece of paper to microfilm. The hard part is devising an indexing scheme that describes every possible parameter of each document for later searches, comparisons and processing.

Indexing—A method by which a series of attributes are used to uniquely define an imaged document so that it may later be identified and retrieved.

Interface—An interface is simply a mechanism for different pieces of software to interact. For instance, application programminng interfaces (APIs) are provided with operating systems to access system-level services from programming languages; database management systems to access SQL database services; and any number of other types of applications and system software.

Local Area Network (LAN)—Data communication network of connected devices within a small area, such as a building or group of buildings. High-speed transmissions over twisted pair, coax, or fiber optic cables that connect terminals, personal computers, mainframe computers, and peripherals together at distances of about one mile or less.

Megabyte (MB)—Approximately one million bytes. Precisely 1,024 kilobytes, or 1,048,576 bytes.

Microfiche—Microform in the shape of a rectangular sheet having one or more microimages usually arranged in a grid pattern, with a heading area across the top.

Microfiche scanner—Device for scanning microfiche.

Microfilm—A film medium, in tape-like roll, for recording reduced pages of documents sequentially.

Micrographics—The branch of science and technology concerned with the methods and technique for recording information on, and retrieving it from, microform. Those methods include reducing and recording images by photographic means, or directly onto film by computer (computer output microform, or COM); the location and retrieval of documents through indexing and mechanical means; and the display and magnification on display screens or paper output.

Migrate—To move files from one storage medium to another, from on-line to near-line or near-line to off-line. Usually files are migrated when they match parameters set by network managers. These parameters include age, time since last access and size.

Near-line—Data that is available on a secondary storage device that the user can access, but at a slower rate than the on-line data is accessed.

Off-line—Data that is not physically stored on an accessible drive, such as removable tapes or disks.

OLE—OLE is a set of system services that provides a means for applications to interact and interoperate. Through OLE Automation, an application can dynamically identify and use the services of other applications. Applications that accept objects from other applications are called containers, while the application providing the object is called a server. Through OLE object linking, objects created in one application can be linked into container applications. As the linked object is changed or revised by the server application, it is automatically updated in any container applications. Through OLE object embedding the container application does not maintain a link to the object's data source, so updates to an embedded object must be made from within the document itself. Through OLE Visual Editing, embedded and linked objects can be directly edited within the container application without switching to the server applications.

On-line—Data that is available on a primary storage device so that it is readily accessible to the user.

Operating system—Collection of programs that, taken together, manage the hardware and software; it is the operating system that makes the hardware usable, providing the mechanisms that application programs use to interact with the computer.

Optical disc—A storage device that is written and read by laser light. Certain optical discs are considered Write Once Read Many (WORM), because data is permanently engraved in the disc's surface either by gouging pits (ablation); or by causing the nonimage area to bubble, reflecting light away from the reading head. Erasable optical drives use technologies such as the magneto-optic technique, which electrically alters the bias of grains of material after they have been heated by a laser. Compact discs (CDs) and laser (or video) discs are optical discs.

Output device—Any device by which a computer transforms its information to the "outside world." In general, you can think of an output device as a machine that translates machine-readable data into human-readable information. Examples: printers, microform devices, video screens.

Rewritable optical—Optical media from which data can be erased and new data added. Magneto-optical and phase change are the two main types of rewritable optical discs.

Read cache—The cache is used to accelerate read operations by retaining data which has been previously read, written, or erased, based on prediction that it will be reread.

Retrieval key—A word, number or phrase associated with a document to aid in its retrieval from storage. Sometimes called descriptors. There are often many retrieval keys used together to fully locate a document; together they are called an index.

Rewritable optical disk—Optical disk on which data is recorded. The data in specified areas an subsequently be deleted and other data can be recorded.

Server—A computer which is dedicated to one task. A database or directory server would be responsible for responding to a user's search request, returning the list of stored documents that meets with the parameters of the request.

Storage media—The physical device itself, onto which data is recorded. Mag tape, optical discs, floppy disks are all storage media.

Text based—Representation of images that requires the use of pre-existing characters rather than vector or raster graphic techniques.

Text management—All the techniques and technologies involved in creating, storing and retrieving text files in an organized and logical manner.

Text search—A technique for examining text files for occurrences of specific sets of characters, either in a string (a word or sentence) or in proximity (a certain word in the vicinity of another word). A "contextual search" involves finding entire documents based on a string of characters that appears in it.

Windows—A Microsoft operating system that features multiple screens and a graphical user interface (GUI).

Workflow—A program that queues, tracks and otherwise manages documents, work items, and collections of documents and work items as they progress from entry into the system, through the various individuals or departments in the organization until a business process is completed.

Workflow application—A software program(s) that will either completely or partially support the processing of work items in order to accomplish the objective of a workflow process activity instance or instances.

Zoom—To enlarge a portion of an image in order to see it more clearly or make it easier to alter.

RHEINNER GUIDE

This Technology Guide is one of a series of guides, written by The Rheinner Group and published by ATG, designed to put complex document management, imaging, and workflow concepts into practical and understandable terms. Each guide provides objective, non-biased information to assist in the internal education, evaluation and decision making process. This Technology Guide, as well as the other Document Management, Imaging, and Workflow Technology Guides in the series, are available on ATG's Web Site.

www.techguide.com

A joint venture

ATG
The Applied Technologies Group

Tel: (508) 651-1155
Fax: (508) 651-1171
E-mail: info@techguide.com

The Rheinner Group

Tel: (617) 741-8100
Fax: (617) 741-5885
E-mail: info@rheinner.com