RHEINNER GUIDE

Enterprise Document Management

Document Management, Imaging, and Workflow Technology Guide Series

This guide has been sponsored by

ALTRIS

Table of Contents

Introduction

Document-based information, such as electronic word-processed documents, spreadsheets, photographs, customer billing statements, digitized paper documents and correspondence received from a variety of sources, represents the overwhelming majority of information within today's corporations. Acting as the container for unstructured information, documents form the lifeblood of any organization and, in essence, represent an organization's ability to communicate and interact with other entities. And yet, despite its importance, the document remains an organization's least-understood, least-automated and most-dynamic strategic resource.

Today's organizations are, at once, drowning in and starved for information. Technological advances have dramatically increased the ability to capture, compile, report and, most of all, "create" information—to the point that most business organizations' ability to create document information exceeds their ability to access, manage and extract the information resident in that information base. Most corporations agree that effectively harnessing and managing this vital resource remains a formidable challenge.

The ability to control and manage business documents strikes at the core of competitive strategy. Efficiently capturing, managing, processing, storing and distributing documents enables a business to become more responsive to changing market conditions, simultaneously improving productivity and cost-competitiveness.

For a long time, the document management chore has fallen to a variety of disparate, specialized technologies like typewriters, printers, film, fax machines and photo copiers. Each performs a few specialized tasks; none is capable of managing documents. And the widespread use of computer and communication technology has only exacerbated the problem by creating more document information that must be managed.

Digital technologies, however, have also brought with them the ability to manage documents electronically, and individual technology solutions have emerged to deal with specific types of data. For example, COLD solutions manage computer-generated reports; document imaging solutions facilitate the handling of previously paper-based information; and desktop document management applications provide enhanced file management for computer-generated word processing files. Generally deployed as departmental solutions to meet specific document control problems, each solution has its own architecture and usage model. With a variety of document control problems to solve, organizations typically have deployed a variety of these solutions throughout the enterprise. The end result is that while individual departments have gained more control over documents, the enterprise has been unable to leverage this control across the entire organization.

This situation has changed dramatically during the last twelve to eighteen months. With today's sophisticated software environments, the ability to manage the wide variety of documents that exist in an enterprise with one architecture and usage model is becoming a reality. This Rheinner Group Technology Guide explores the concept of enterprise document management, details the necessary components of an enterprise document management platform, and outlines the key architectural considerations when implementing enterprise document management applications.

Document Information Defined

The term document is used in many different ways, and can be applied to numerous types of objects. Webster's Dictionary defines the word "document" as "original or official written or printed paper furnishing information or used as proof of something else." The advent of the digital age has clearly stretched the definition to include word processing data files, spreadsheets, computer-generated output and document images.

Obviously, computer technology and the traditional world of the paper document have crossed paths. At one time the two worlds existed separately: Analog devices were used to create and manage paper documents while digital devices manipulated data into the form of a document that would be converted to paper. Today, these worlds are inextricably linked.

The vast majority of paper documents produced and utilized by business organizations are generated electronically. While they could also be electronically communicated, most are not. In fact, the average US office worker consumes more than 50,000 sheets of 8.5 x 11-inch paper each year. And that figure is rising. Add to that the vast numbers of pre-printed forms, other pre-printed material, official corporate documentation and it's readily apparent that the world of paper presents a formidable challenge to anyone involved with managing that base of information.

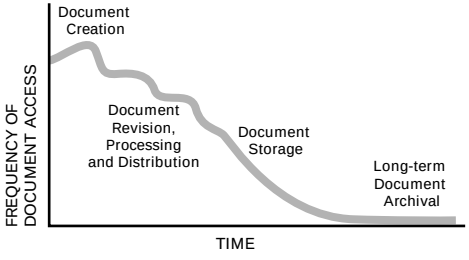
In the analog world, the distinction between a document and a work-in-progress is easy to determine. A document, in fact, became a document the instant it was created. In the digital world, the distinction is much more difficult to determine because a large number of digital documents exist and play the role of business documents before they ever hit paper. For this reason the corporate environment must view the challenge of document management on two levels. A large portion of the document load is on paper, an equally-sizable portion exists in its electronic data state. To leverage the full body of information requires management of all document information.

The Document Life Cycle

All documents have a life span that can be divided into stages of activity: document creation; active processing when the information is retrieved, routed, edited and processed; inactive processing and review when the information still must be available but is largely static; and document disposition when the information is discarded or archived (see Figure 1).

The overall duration of the life cycle, as well as each stage, varies greatly by type of document, by business process and by industry. In some industries, legal and other business practices require a document life cycle of over 100 years. While much of that life is spent in archive, the document may need to be accessible at any time during that period. In other cases, for example, the development of a product marketing campaign, document activity may be quite high for several months, after which time period the need to access the documents is minimal. Understanding the life cycle and usage patterns of documents within an enterprise is critical to determining system, application and subsystem requirements.

Figure 1: The Document Life Cycle



Just like documents, document management requirements are dynamic. As documents are created, revised, distributed, transacted, and archived, they move to different individuals within a department, and among various parts of the organizations. Some document management requirements, such as security and transaction integrity, remain constant through the life of the document; others, such as version control and flexible retrieval parameters, are more critical at certain stages of the document life cycle.

Discrete Document Management Solutions

In an attempt to gain control over the various types of unstructured document information, organizations have implemented individual, discrete solutions built around the following technologies:

- *Extended Directory Services:* Designed to provide more control and better management of computer-generated data files (especially word processing data files), extended directory service technology adds enhanced file security, revision control, file descriptions, extended file names and user access privileges to the basic file directory management features of the computer operating system. While this is frequently referred to as document management, it is only a part, not the whole, of an enterprise document management system.
- *Document Imaging:* Document imaging technology enables the input, indexing, management, storage and retrieval of digital image files, which typically originated as scanned or faxed forms and letters. These image files require specialized viewing software, compression and decompression software, drivers for specialized subsystems, and support for a range of specialized storage environments.
- *COLD:* Computer output to laser disk technology allows the processing, indexing and storage of computer-generated, formatted reports on computer media. Previously, this data would be output to printed report pages and/or to microfiche. With COLD, users can electronically search, view, print and process the information contained in any report.
- *Workflow:* With workflow technology, organizations can define work processes in terms of participants, inputs, outputs and work flows, and automatically route work tasks and the information required to perform those tasks throughout the organization. Other workflow features include the ability to set rules and policies that govern the flow and fulfillment of work tasks, the ability to monitor workloads and reallocate resources accordingly, and the capability to revise the flow of work after identifying inefficiencies or bottlenecks.
- *Groupware:* Groupware technology permits the collaborative sharing of work and information among individuals and groups. Based on a messaging infrastructure and a document database, groupware provides coordination and collaboration features for group projects and activities.
- *Internet/Intranet Web Sites:* Companies are aggressively exploiting the platform-independent nature of the Internet to better manage the publication and distribution of unstructured data across the enterprise. Private Web sites make it easy for users throughout the organization to access information created with different applications. In addition, organizations are now beginning to Web-enable groupware and workflow applications for collaborative work management.

While each of these technologies offers compelling benefits, they are typically deployed as discrete solutions to specific departmental problems, not as a comprehensive solution to the overall problem of document management. And, as the individual technologies have evolved, each has developed unique and often proprietary access, retrieval and security approaches that make it impossible to sit at a single workstation and, at one time, access all the information relevant to a specific account, customer or project. In short, each system manages information objects that are generally inaccessible to the enterprise.

When implemented as discrete solutions, these technologies tend to stratify information assets, holding them captive in a single application instead of permitting access by other users in the enterprise. In addition, each system represents a new learning curve for the users, the internal application developers, and the IS staff charged with maintaining and supporting these systems. This is a costly situation, and one that inhibits the expansion of current applications and the development of new applications.

Enterprise Document Management (EDM)

A preferable alternative to the deployment of individual document management systems is an enterprise document management (EDM) system, a single software environment that can manage a variety of document types and be extended across the enterprise. With a consistent set of interfaces, development tools and support services, this fully integrated platform eliminates many of the drawbacks associated with the departmental approach to document management.

To provide enterprise document management, a software platform must include four basic capabilities:

- A core management environment with a built-in security and access model, as well as collaborative work processing features, such as document annotation, document distribution and rules-based routing.
- A flexible application development environment that allows the creation of customized document management applications to meet specific document control and processing requirements.
- The means to create a centralized or distributed document vault, which contains different types of document objects and which allows for controlled access to documents from all document management applications.
- A platform and document management infrastructure that can be utilized across the heterogeneous computing infrastructure of the enterprise.

Functional Features of the EDM Platform

To fully and effectively replace the discrete document management solutions that solve specialized document control problems, the EDM platform must be capable of managing multiple types of document objects throughout the entire life cycle, including:

- *Document Capture*—The EDM system must support direct importation of digital document content and the conversion of paper documents to document images.
- *Document Management*—The system must be capable of managing any form of digital document, including a variety of image formats, textual and compound documents, electronic files and digital audio/video formats. Users must be able to index and categorize documents, and document

storage locations must be referenced. In addition, the EDM system should track document access and retrieval activities.

- *Document Storage*—The system must support on-line, near-on-line and off-line document archiving to multiple storage mediums, including digital formats such as optical disk, CD-ROM, tape and magnetic disk as well as analog formats, such as microfilm and paper. The system must securely manage the use of media for document storage, support single or multiple document storage locations, migrate documents between storage media as required by document life cycle retention requirements and cache documents to fulfill retrieval requests from system users.
- *Document Access*—The system must present a unified search and retrieval paradigm, where documents can be accessed, retrieved and viewed under a consistent usage model regardless of content and storage location.
- *Document Retrieval*—The system must be capable of delivering selected documents to the desktop and include functions for personal manipulation of the document, such as annotation, copying and printing.
- *Document Exchange*—The EDM system must provide access to third-party applications, allowing documents to be extracted, contained by a standard file format and transported as required to support enterprise interoperability objectives.
- *Document Output*—The system must support personal and high-speed batch document printing, publication of document collections to CD format and exportation of documents for use by third-party applications.

EDM Applications

Document management challenges vary dramatically across the enterprise. The EDM platform, therefore, must include a core set of document management services that can easily and repeatedly be reconfigured to deliver a broad range of document management applications. These applications can be grouped into one of five categories:

- Creation and Revision Control
- Foldering
- Transaction Processing
- Document Enabling
- Storage and Retrieval

Creation and Revision Control

Existing word processing documents, spreadsheets, presentations, and other electronic documents are typically managed as file entities by the computer operating system. While this allows for the ability to save work and move older work to network servers or pre-defined directories for backup, it provides for no organizational control in a collaborative work environment. Individual users can easily edit the wrong document, accidentally erase a document, create a new version of a document that already exists, duplicate work that has already been done, and easily access documents not intended or appropriate for them. With the proliferation of PCs throughout the enterprise and the ability to store hundreds of thousands of documents, it is easy to see how this lack of control can become a serious challenge to an organization's productivity, security and competitiveness.

By providing an environment that controls the creation, revision and disposition of documents,

document management applications avoid the common errors related to improper filing, duplication of work and security breaches, all of which enable an organization to regain lost productivity.

Foldering

In addition to managing the dynamic life of a document, organizations also have document management needs that are tied to a specific business process, such as customer service, contract management, or litigation support. Foldering features enable documents to be grouped into a folder, which can then be managed by the document management application.

For example, a single customer folder might contain all documents related to a specific customer, including correspondence, credit verifications and billing statements. The document management application controls distribution of and accessibility to that folder.

The use of workflow technology may be an important underpinning to this type of application, allowing for the automatic assembly of folders and the automatic forwarding of work to appropriate individuals or departments.

Transaction Processing

Most enterprises have departments that focus entirely on processing transactions, such as processing credit card applications or receiving payments from customers. These departments have entirely different business objectives than other departments within the enterprise, which, in turn, translate into very different requirements for the document management application.

Transaction-centered documents typically have a short and intense document life cycle. Documents are received, work is processed, and the documents are discarded, archived or forwarded to another department. Often, there are large volumes of documents and the ability to rapidly perform the transactions is critical to the operation of the enterprise. In this case, the core functionality of the document management application must be optimized for speed and efficiency.

Document Enabling

While some document management applications focus on controlling vital documents, others involve enhancing existing data applications by providing easy access to related documents. For example, an accounts payable process can benefit from having an electronic copy of the original invoice attached to a payable record. Should issues or questions about a specific invoice occur, a payables clerk would be able to retrieve the original invoice by accessing the data record in the accounting program.

In a document-enabling scenario, the existing data processing application is the user interface to documents. Documents are provided by a storage repository, which is linked to the specific line-of-business application. The document management system manages the link between the two.

Storage and Retrieval

Many of the documents created or received by an enterprise simply need to be stored. Hazardous material sheets, insurance policies, tax records, warranty cards, engineering drawings for original parts and a multitude of other documents must be retained and accessible for some required period of time.

For these documents, the emphasis is not on creation or transaction-processing, but on secure, cost-efficient storage and retrieval. The document management application provides the ability to store vast quantities of documents in a secure environment, while also enabling users to quickly identify and retrieve documents on an as-needed basis. The management of the

document repository, the index database and the access privileges of the individual user are provided by the document management platform.

EDM Architectural Considerations

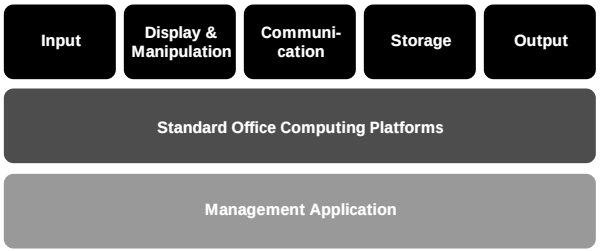
With the awesome task of integrating and managing all of the critical business information resident in unstructured document information, the EDM is navigating waters previously uncharted by technological solutions. That challenge brings with it unique architectural issues that must be addressed before the benefits of the EDM approach can be realized.

Managing Document Images

The architectural hallmark of an enterprise document management platform is not just its ability to manage the creation, revision and distribution of electronic documents, but the capability to manage all documents in a comprehensive manner. Inevitably this means providing comprehensive management of document images. A field-tested enterprise document management platform has as part of its internal design the ability to cross over into the paper world bringing paper-based documents into the same consistent electronic document management environment. While this technology is often trivialized as a simple scanning process, the ability to bring document images into a system and effectively manage them is a formidable challenge that can only be accomplished by serious software platform environments.

Every computing system consists of essentially five subsystems: input, display, communications, storage and output, each of which is integrated with the central processing unit (CPU) and specially configured to manage the type of data being processed. In the data processing environment, the keyboard is the input device directly supported by both the computing device and the computer operating system. Each of the other subsystems is similarly optimized to manage character and graphic data, supported by the computing platform, the operating system or some other intervening layer, such as the network operating system (See Figure 2).

Figure 2:
The Document Management Component Model



But image data is very different from other computer data. Most importantly, images of documents are simply a collection of dots that form legible characters to the human eye, but are not "usable" by computer systems. This means that one cannot search for the occurrence of a particular word or phrase in an image document; one cannot spell check it; one cannot edit the document–insert, delete or rearrange the elements. The second most significant difference is that image file sizes are much larger than other computer data files. For example, a word processing document that contains a letter might comprise a file size of two to three kilobytes; the scanned version of that letter might represent an entire megabyte.

Since the unique requirements of image data are not directly supported by the existing computing infrastructure, a comprehensive array of specialized hardware and software optimized to manage image,

rather than character, data must be integrated with the computing system. Similarly, to truly be enterprise-wide, an EDM system must be able to manage document images as easily as other electronic documents. The EDM platform must seamlessly integrate with the standard computing environment to perform the document management tasks required by both targeted business processes and by the enterprise. The EDM platform must also provide the subsystem support required by those applications.

In particular, three subsystems—capture, storage and communications—provide the core services required by the document management application. If these subsystems are not properly designed, configured and implemented, they can be points of EDM system failure as applications expand across the enterprise.

Document Capture

Capture is an essential ingredient of a document management solution. It is the capture subsystem that is responsible for the delivery of image data to the management application. The capture system must be optimized to reliably, quickly and optimally deliver images to the system. To accomplish this, the EDM system must provide direct support for the document capture hardware and software. The capture system also may be required to work in tandem with other software in order to process images. The most common of these are forms processing, OCR/ICR and raster-to-vector conversion for engineering drawings.

A flexible, scalable capture architecture is best suited to the varied and, sometimes, unpredictable nature of capture requirements across an entire organization. An EDM system must be able to accommodate the department that deals with millions of paper documents, along with the department that utilizes a few thousand pages.

Document Storage

Not only are document images larger than their computer-generated counterparts, but they typically must be retained for a longer period of time. The EDM platform must include the facilities to manage a storage architecture that includes both long-term archival, as well as features that enable rapid retrieval of active documents.

When designing a storage subsystem, the storage hierarchy should guide the selection of devices to achieve an appropriate balance between capacity, performance and cost (See Figure 3).

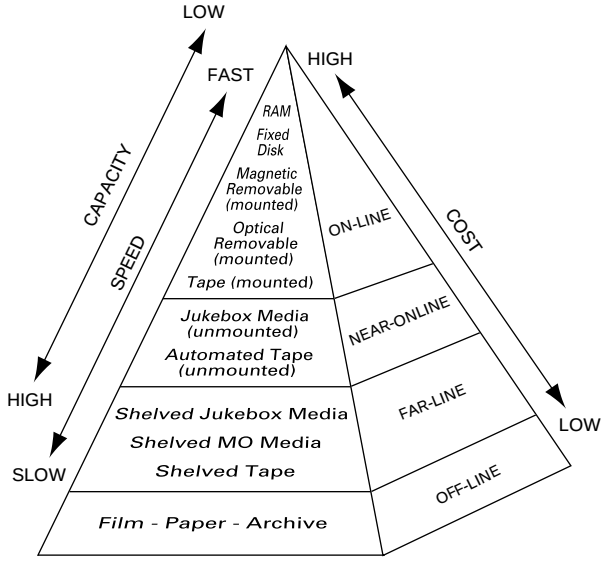


Figure 3: The Storage Hierarchy

The storage architecture in an EDM environment generally requires the use of multiple devices, including WORM (write once read many) optical disks, optical jukeboxes, RAID (redundant array of independent disks) and/or automated tape storage systems. The management application may require that rarely-used files be automatically moved downline in the storage

hierarchy. Here again, the EDM platform should make provisions for this type of storage management. By allowing data to be moved across a variety of devices, the software platform provides the system designer with the flexibility needed to balance low-cost archival storage with high-performance retrieval of active documents.

Communications

Due to their nature, EDM applications place substantial demands on the communications computing infrastructure. First, enterprise document management, as implied by its definition, involves collaborative work activity for small workgroups, departments, multi-departments and the entire enterprise. Work items and document objects must be moved across these groups in an efficient manner. In addition, managing non-traditional data types, such as document images, is an input/output (I/O) intensive operation. Both the size of these document files and the processing performed on these files can tax the communications subsystem. Unlike traditional data processing activities, in which reasonably-sized electronic documents are requested from a server, worked on locally for an extended stretch of time, and then written back to the server, very large image files are constantly requested, viewed, moved back or routed to others.

With these issues in mind, users should look for advanced levels of control in the communications infrastructure when selecting an EDM system. For example, targeted image extraction technology, which selects a subset of data for transfer according to the resolution and depth required for viewing, can greatly reduce the network impact of managing document images. Because each document image is typically stored at a much higher resolution than can be seen on the screen, this technology built into the enterprise document platform minimizes processing speeds and memory requirements for display. This also reduces network requirements by as much as 95%, delivering dramatic improvements in both speed and capacity.

Utilizing Workflow Technology

Another core EDM technology is workflow, which can be utilized to automatically distribute work items on the basis of system-wide rules. The ability to route, manage and balance the flow of document objects across an enterprise is a core competency of the EDM platform.

Workflow software is typically built around a workflow engine that can either be e-mail- or server-based. E-mail based workflow uses the electronic mail system to distribute documents according to a limited set of rules. Individual users can invoke the workflow to send documents to other e-mail recipients and define a set of procedures for the document. Useful for ad hoc applications, e-mail-based workflow provides individuals with greatly enhanced personal control over work-related fulfillment requests sent to others in the enterprise.

Server-based workflow achieves similar results, albeit in a considerably more powerful and comprehensive manner. With server-based workflow, organizations can define a complex and interdependent set of work procedures that are automatically invoked by the system, rather than on an as-needed basis by an individual user. While most server-based workflow applications now allow the individual to customize procedures to some extent, the primary objective of server-based workflow is to push work items through a specific process on a predetermined basis. With advanced rules-based routing, reporting and work balancing features, server-based workflow can automate transaction-oriented work processes throughout the enterprise.

Organizations typically have the need for both types of workflow, as well as the need to move

document objects between both systems. In some departments, the flexible and ad hoc nature of e-mail-based workflow facilitates the completion of unpredictable and iterative projects. In other departments with specific and repeated work procedures, server-based workflow completely automates the fulfillment of these procedures.

Document and Content Management on the Web

The rapid emergence of Internet infrastructures and products with browser-based interfaces are affecting nearly all aspects of information system architectures, including networking services, operating system services, DBMS services and front-end development tools. Perhaps more important than the architectural implications is that Internet computing is fueling the proliferation of network-based digital content, including text, documents, multimedia files and other complex object types.

Document management activities are fundamentally affected by the changes that are occurring around the Internet, the Web and corporate Intranets. The implications are even more dramatic for an EDM platform. The ability to capture, access and store information from the Web, as well as the ability to use standard Web browsers to access document information, are key requirements for a document management platform that is truly enterprise-wide.

Internet technology lends itself to many of the aspects of a document management solution, particularly when it is being deployed internally where it is not victimized by the slow downs and uneven performance of the World Wide Web. Intranets, which can be organized according to context and content, can be used as a document distribution vehicle for the EDM platform. In this way, the simplicity of access becomes a convenient entry point to retrieve vital corporate documents.

Summary

Commonly thought of as a singular application, document management is more accurately defined as a general concept that can have many different implementations. The document management needs of departments across an enterprise vary greatly: some require comprehensive document creation and revision control features, some are transaction-oriented, others folder-oriented, and still other departments may need all of these functions. And yet, while document management requirements are diverse, the need to share document information across the enterprise is universal. Meeting the unique document management needs of the department along with the universal document sharing requirements of the enterprise has long been a challenge to organizations.

The answer is an enterprise document management platform that provides a consistent set of tools, interfaces and core services that can be configured to deliver different user applications. It is the platform that is extended throughout the enterprise; not necessarily each type of application. With streamlined application development tools that enable the rapid and repeated creation of specific applications, and common document vaults, object management services and user interface tools to facilitate document exchange, a true enterprise document management platform meets the individual needs of departments and the collective requirements of the enterprise.

CASE STUDY 1: Alcatel Network Systems Redefines the Rules of Competition

Companies today are struggling to gain better control over a vast array of documentation. Invoices, procedures, maintenance records, training videos, facility layouts, photos, schematics, spreadsheets, product drawings, checks, specifications and claims are forcing companies across all industries to rethink how they work with these documents in an efficient, competitive manner.

Consider the demands on today's enterprise. It is no longer acceptable to operate on a local basis. The emphasis is on being departmentally and geographically global. Communication between offices, across departments and around the world must be as quick and effective as a hallway conversation.

A simple database of documents no longer meets a company's needs. Since most information is used by many departments companywide, the challenge is to ensure seamless access of accurate data from one source. An enterprise document management system from Altris contributes services that embrace each specialty within an organization system and reaches across geographic and departmental boundaries. The system improves processes by linking traditional structured data and documents, creating powerful new methods for review and change. And it can incorporate a wide variety of technologies in a standard, supportable, scaleable, consistent manner. A good example of the system's ability to grow in many directions is Alcatel Network Systems, Inc., which stands out as a company that has leveraged enterprise document management to secure the number-one spot in the highly competitive telecommunications manufacturing industry.

Alcatel designs, manufactures and markets a full line of voice, data, and image transmission products to customers such as BellSouth, Ameritech and Sprint, who in turn compete vigorously in the telecommunications services market. The company's goal is to ensure that Alcatel is the number-one name in the telecommunications manufacturing industry across the United States and in key international markets. For Alcatel, success is directly related to their ability to attract new customers and provide existing customers with superior service. They measure success by their ability to achieve long-term growth and strong financial performance.

Under pressure to live up to its goals and objectives, Alcatel has long focused on delivering products on, or preferably ahead of, time. Despite the scrutiny placed on process improvement, only recently was there a review of the documentation distribution and archive process. The results were surprising. According to Mikal Mattix, Project Manager for Alcatel Network Systems, "At one time, we estimated that our 4,000-person company used more than 4.7 million pieces of paper in a year, and that is just engineering reference documentation."

Dealing with the volume and size of documents and the fact that the documents had to be interchangeable among Alcatel's four manufacturing centers across the southern United States and two others in Mexico and Canada, proved to be a bottleneck in the Alcatel documentation distribution system. In fact, Mattix quips that there almost came a time when they could manufacture a item quicker than they could identify and deliver the documentation.

Alcatel Network Systems document enabled line of business and PDM applications using an Altris document management component to help maintain their number-one ranking in this competitive industry.

An Electronic Document Retrieval (EDR) system has replaced the traditional hardcopy distribution method, allowing released engineering documentation

to be accessed from any workstation or PC at any one of Alcatel's four manufacturing centers. The EDR controls the product flow and allows on-line viewing, printing, and distributing of aperture cards, microfiche, and scanned documents.

"Before EDR, second shift manufacturing employees had no way of getting their hands on necessary documentation, holding back the manufacturing flow for hours or even days," explains Jim Rutelonis, Director of the Document and Data Management Department. "Today, 300 users at our North American sites have round-the-clock access to the master document vault, which is located at headquarters in Rchardson, Texas."

Alcatel was able to reduce nearly 50% of the 4.7 million pieces of documentation generated annually by the engineering department. Costs associated with maintaining paper archives are being reduced as is the expense of shipping documents among the various locations. And, Alcatel expects even further improvements in the months to come.

"One of the most exciting aspects coming up within the system is how we have linked with Alcatel's Intranet using the WISDOM product," explained Mattix. "We were fortunate enough to be a tester of this product for Altris and we've found that we have many benefits. We have even tested the product from Richardson to Stuttgart, Germany and they have retrieved documents in seconds."

WISDOM is an exciting supplemental interface to Altris document management systems. By enabling access to these systems through Internet and Intranet browsers, WISDOM enables rapid integrated document management deployment through an enterprise. This approach optimizes the investment in on-line document management while minimizing rollout time. And by allowing access to view tools within the browser or as helper applications, WISDOM provides a path to full enterprise document management and view functionality.

Customer service is also benefiting from the new document management system. The EDR system allows employees to work with customers in "real time."

"Everyone in our company depends on a speedy network, so we didn't want to invest in an enterprise document management system that would drain that resource. At the same time, we couldn't arbitrarily decide whose application had top network priority," said Rutelonis. "The solution was multiple servers: the main document archive server in Richardson and a second "cache" server in Raleigh, North Carolina." The Raleigh and Clinton, North Carolina, operations rely on the EDR system significantly more than the other sites combined.

Alcatel's research found that once requested, documents continue to be in demand over a certain time period. The multi-server architecture system circumvents a predictable data traffic jam. Upon an initial request, documents are cached locally: Raleigh and Clinton requests are cached on the Raleigh server; all other document files are cached on the Richardson server. Subsequent request for that same document become local, keeping the wide area network free to quickly process other transactions. Over time, preset algorithms determine when the cached documents should be returned to the archive.

"Not only were we able to save money by integrating document-enabled PDM with our existing hardware and software, but we have an innovative architecture in place that ensures the data is always accessible," said Rutelonis. Because of Alcatel Networks Systems' foresight to base its new document management system on a multi-server architecture with Intranet document distribution, the EDR can grow for many years with minimal impact to vital enterprise-wide network traffic.

CASE STUDY 2: Good IDEAS Enhance Utility

Altris has integrated document management systems at more than 25 utilities around the world. By adding view, markup, edit, and distribution functionality to existing applications, utilities gain the upper hand in information and records management.

Entergy Corporation, headquartered in New Orleans, Louisiana, is an energy company with worldwide power production and electric service operations.

Its regulated subsidiaries deliver electricity and related services to 2.4 million retail customers in Arkansas, Louisiana, Mississippi, and Texas. Other subsidiaries provide wholesale electricity to other utilities, invest in, and mange domestic and overseas power systems and sell energy-efficient lighting and related systems and services.

Entergy has made a strong commitment o using technology to improve its work processes. Microsoft Chairman and CEO Bill Gates noticed and said, "I think one of the key elements in a company being very competitive will be how it uses information tools in the years ahead, and Microsoft has been very impressed with how Entergy is applying the latest tools in a way that makes it more effective in its business."

An impressive example of effective application of technology is Entergy's IDEAS-Integrated Documents Electronic Access System.

The IDEAS project involves long-term implementation of an electronic document management system encompassing all controlled documents in 30 different operational classifications, each with unique life cycle and revision control procedures. Integrating images, documents, applications, and users through a multifunctional software set will contribute to Entergy's competitive advantage well into the future.

Currently, 950 users at four nuclear locations—Arkansas Nuclear One; Grand Gulf Nuclear Station; River Bend Station; and Waterford 3—as well as the Jackson, Mississippi, nuclear headquarters and at Entergy's Corporate headquarters in New Orleans—have access from their desktop computers to engineering drawings, nuclear regulatory procedures, safety procedures, QA records and correspondence. The document database reflects 10 years of information collected from a combination of scanned and electronic files. Entergy's goal is to ultimately roll out the system to approximately 2,000 users.

"We now have the ability to pick up information without having to go to a document control area and without waiting for a physical search, scan, or copy," said Bruce McCall, IDEAS Project Manager at Entergy's Grand Gulf Nuclear Station in Port Gibson, Mississippi. "The benefits to the end-users will grow considerably as the database grows and additional functionality is added."

The IDEAS database contains approximately 2-3 million records at each of the nuclear sites. Even so, that is only a portion of the system's expected population.

Managing this database is a Pro EDM system supplied by Altris and integrated with EPOCH software driving two 120-Gb Hewlett-Packard jukeboxes. One small server, located at corporate headquarters in New Orleans, is used to develop and test all system changes before live site installation. A Sun Sparc Centre 2000 server resides at each of the participating power plants to support concurrent users at each site.

The system also employs CAD-Connect technology, which manages documents in image form and in native file formats, providing a transparent link to Intergraph Microstation. Workflow and full-text search functionality will be added, increasing the system's utility. One unique application of the IDEAS system is to maintain Health

Physics dosimetry data, which is stored on a mainframe computer. When data is requested, it is downloaded complete with images to IDEAS.

"Since the data is archived electronically, the application has replaced microfilm and paper storage of a very large database containing files that must be retained for the life of the plant," explained Margit Triggs, IDEAS Project Manager at the Waterford 3 plant in Killona, Louisiana. "Now, dosimetry data can be referenced from any PC, from any of the IDEAS-capable locations, with the assurance that requested information will be current and available quickly."

According to Triggs, "IDEAS was built with an open architecture. We are working carefully at this critical stage to ensure growth and expansion to new users, sites, and applications in the future."

Glossary

Algorithm—Prescribed set of mathematical steps which is used to solve a problem or conduct an operation.

American National Standards Institute (ANSI)—A standard-setting, non-governmental organization, which develops and publishes standards for "voluntary" use in the United States.

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

Annotation—The ability to attach notes to graphics or images by tying them in, using a light pen or digitizing tablet. Useful for clarifying documents or editing images.

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—Application frameworks are sets 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.

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.

Authorization code—Identifying code, often a password, that allows a user access to a system. Used mainly for privacy and security. Also used to divide up a computer's capacity among departments and/or hierarchies; different "grades of service" are given authorization codes.

Backup—A duplicate copy of data placed in a separate, safe "place"—electronic storage, on a tape, on a disk in a vault—to guard against total loss in the event the original data somehow becomes inaccessible. Generally for short-term safety.

Bandwidth—(1) Number of hertz expressing the difference between the lower and upper limiting frequencies of a frequency band. (2) Width of a band of frequencies. (3) Maximum number of information units (bits, characters) capable of traversing a path per second.

Boolean search—Search strategy for selected information that uses AND, OR, NOT functions.

Business Process Reengineering (BPR)—The radical restructuring of the business processes, organizational boundaries, and management systems of an organization. Business process redesign and business process automation are components of BPR.

Cache (Memory & Magnetic)—Small portion of high-speed memory used for temporary storage of frequently used data. Reduces the time it would take to access the data, since it no longer has to be retrieved from the disk.

Class hierarchy—A class hierarchy is a group of superclasses that are related through an inheritance tree. In object-oriented programming, class hierarchies are used to identify common data and procedures once as a superclass, and then to enable the superclass to act as a template for related object types.

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.

Common Object Model—The Common Object Model is the functional equivalent of the Component Object Model for UNIX-based platforms that today include SunOS, OBM, AIX, HP-UX, ULTRIX, OSF/1 and OpenVMS. The Common Object Model defines a common DCE RPC-based protocol and a subset of core OLE functions that Digital and other interested companies plan to support within their products.

Compact Disc (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.

Component software—Component software is an application that contains one or more component objects that can freely interact with other component software through OLE capabilities.

Compound documents—Compound documents are documents which contain multiple data types. Often, the different types of data have been created by different applications, and embedded into the document.

Compression—A software or hardware process that "shrinks" images so they occupy less storage space, and van 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.

Consultative Committee for International Telegraph and Telephone (CCITT)—International organization that develops international combinations standards.

Contention—When two or more users try to access the same device at the same time. There are "collision prevention" techniques to solve contention problems in LANs.

CORBA—Common Object Request Broker Architecture. A standard which defines the manner in which software "objects" created in one program can be used in another.

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

Dynamic Data Exchange (DDE)—Allows one free standing program to give commands, take requests, and give and receive data from another free standing program in a Microsoft Windows environment.

Dynamic Link Library (DLL)—The Microsoft Windows specification for linking program subroutines from one application with the subroutines of another.

Electronic Data Interchange (EDI)—An electronic communications standard which connects business trading partners for conducting contract negotiations, sales, invoicing and collections.

Electronic forms—Graphics that are merged electronically with data to appear one each frame. Can be as simple as a borderbox, or a logo or running header.

Electronic imaging—Electronic techniques for capturing, recording, processing, storing, transferring and using images.

Electronic mail—A means of connecting computers in order to send messages to one or more individuals or groups.

Encapsulation—A technique used to isolate some of the decisions made in writing a program. To encapsulate decisions a program is organized into an interface, such as a set of procedures, and an internal part. All access to the programs services are available only through the interface. The internal part of the program can be used, but not accessed.

Encryption—Security measure involving scrambling or transforming digitized data into a coded form by means of an algorithm.

Ethernet—Particular implementation of a bus-type local area network that communicates at 10 megabits per second.

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.

Filters—These are criteria that allow you easily to exclude specific files or groups of files from migration. They might include file type, owners, size, file attributes, etc. They can be used singly or in combination.

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.

Full text search—The ability to search text files for occurrences of certain words digits, sentences or patterns of characters. Generally, a scanned document cannot be full text searched. To do that, the document would have to be retyped or scanned with an OCR to create a text file.

Graphical User Interface (GUI)—Computer control system that allows the user to command the computer by "pointing-and-clicking," usually with a mouse, to pictures or "icons," rather than typing in commands.

Hierarchical File System (HFS)—In DOS, the file management system that allows directories to have sub-directories, and sub-subdirectories. In Macintoshes files may be placed into folders, and folders to be placed within other folders.

Imaging—Recording "human-readable" image—pictures, images, motion, text, etc., into "machine-readable" formats, i.e. microfilm, computer data, videotape, OCR output, ASCII code, etc.

Jukebox—A device that holds multiple optical disks and one or more disk drives, and can swap disks in and out of the drive as needed. Same as an autochanger.

Keyword—A word associated with a document or document image to aid in its retrieval from storage.

Laser—Source that produces light that is nearly monochromatic (of only one wavelength) and highly coherent (with waves in phase both temporally and spatially). Acronym for Light Amplification by Stimulated Emission of Radiation.

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.

Management Information System (MIS)—Management information system that is provided by computer processing.

MAPI (Messaging API)—A Microsoft published API that separated the client from the server functionality, allowing various clients, like mail front ends, word processors, spreadsheets, etc., to access the messaging capabilities of back-end mail servers, such as Microsoft Exchange Server.

MAPI Workflow Framework—The MAPI Workflow Framework defines custom message classes and associated collections of properties to enable the interoperability of different workflow systems. By sending and processing MAPI Workflow command messages, workflow systems can interact and share information.

Multimedia—Combining more than one media for the dissemination of information, i.e., using text, audio, graphics, animation and full-motion video all together. Requires enormous amounts of bandwidth and processing power.

Object technology—Object technology is a broad term that refers to the use of "objects" to (1) analyze; (2) model or design; and/or (3) implement some aspect of a computer system.

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.

OLE controls—OLE Controls are a special form of component Automation Object. OLE Controls are similar to Visual Basic custom controls (VBXes), but their architecture is based on OLE. This means that OLE Controls can be freely plugged into any OLE-enabled development tool or application.

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

Optical Character Recognition Reader (OCR)—The ability of a scanner with the proper software to capture, recognize and translate printed alphanumeric characters into machine readable text. Most OCRs work by using either Pattern Matching or Feature Extraction.

Optical drive—Machine for reading or writing a data storage medium (disk, tape, card, etc.) that uses light for examining patterns.

PostScript—A software published by Adobe Systems that translates graphics created in a computer to language a (PostScript-compatible) printer can understand. It is called a page description language. Postscript-compatible printers have interpreters in them that create the proper dot patterns to recreate the screen image—text and graphics—to a page of paper.

Protocol—Formal set of conventions governing the orderly exchange of information between communicating devices by defining such things as connection establishment, security provision, data sequencing, error control, etc. Protocols achieve efficient line use by reducing the amount of information transferred by distinguishing between device control information and data.

Proximity Search—A feature of full-text searching, in which every occurrence of a word within a certain distance of another word is found, i.e. finding every time the word "budget" is mentioned within 20 words of the word "Congress."

Query—To ask or inquire about something within a database.

RAID—Redundant array of independent or inexpensive disks. A storage device that uses several optical discs working in tandem to increase bandwidth output and to provide redundant backup.

Relational database—A database built and operated in accordance with the relational model of data which holds that all data be organized as a set of two dimensional arrays or tables which have a relation to each other.

Remote Procedure Call (RPC)—A mechanism through which applications can invoke procedures and object methods remotely across a network. Using RPC, an application on one machine can call a routine or invoke a method belonging to an application running on another machine.

Scaling—Technique using an algorithm to convert a bit-map of one density into a bit map of another proportional density. Scaling usually involves enlarging or contracting an image.

Scan—To convert human-readable documents into bit-mapped images.

Source-document capture—Conversion of documents, usually paper, to microimages or digital images.

Tagged Image File Format (TIFF)—A bit map file format for describing and storing color and gray scale images.

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.

Tiling—Reproducing oversized artwork or documents by breaking the image area into parts (called tiles). Adjacent tiles repeat a small portion of the image, and they may contain crop marks as well. The repeated portion of the image (the overlap) and the crop marks aid in reconstructing the overall image from the tiles.

Transmission Control Protocol/Internet Program (TCP/IP)—A set of protocols developed by the Department of Defense to link dissimilar computers across networks.

Value Added Reseller/Value Added Dealer (VAR/VAD)—Companies that buy equipment from computer or telephone manufacturers, add some of their own software and possibly some peripheral hardware to it, and then resell the whole computer or telephone system to end users.

Vaulted—Computer files that have been literally locked away in a vault for long-term safekeeping. These files are no longer available on-line or near-line.

Virtual storage—Your total storage capacity equals your fixed-drive capacity, combined with the capacity of removable storage devices. This gives an illusion of unlimited storage capacity.

Visual Basic Extensions (VBXs)—Add on Programs (written in C or C++) to the Microsoft's Visual Basic development environment that perform specific functions, such as putting a menu on a screen, or deskewing images). VBXs can be combined to build customized document imaging application modules.

WAPI—The application programming interface/interchange for client workflow applications and tools in order to be able to interface to the Wordflow Enactment System. WAPI is an acronym for Workflow Application Programming Interface/Interchange.

Wide Area Network (WAN)—Communications network that links broad geographic areas.

Wildcard—A character in a text search that stands for other characters. For instance, a search for GEO* (with the asterisk being the wildcard) would find all occurrences of words starting with the letters GEO—geography, geostationary, geology, etc.

Workflow—In imaging software, a program that queues, tracks and otherwise manages documents and collections of documents as they progress from entry into the system, through the various departments in the organization, to their final destination.

Workflow process—The computerized facilitation or automated component of a process.

Workflow process definition—The component of a process definition that can be automated using a workflow management system.

Workflow process monitoring—The ability to track workflow process events during workflow process execution.

Workstation—A single-user microcomputer or terminal.

Write Once Read Many (WORM)—Optical storage device on which data is permanently recorded. Data can be erased, but not altered, and no additional data can be added.

X.25—A standards recommendation from CCITT. The term X.25 has come to represent a common reference point by which mainframe computers, word processors, mini-computers, micro-computers and a wide variety of specialized terminal equipment from many manufacturers can be made to work together over a type of data communications network called a packet switched network.

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