

ECM Toolkit

Taking the First Step



Make informed purchase decisions based on your specific business needs, project goals, and technical requirements.



The ECM Association

1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910
Tel 301.587.8202 / 800.477.2446

www.aiim.org

Taking the First Step

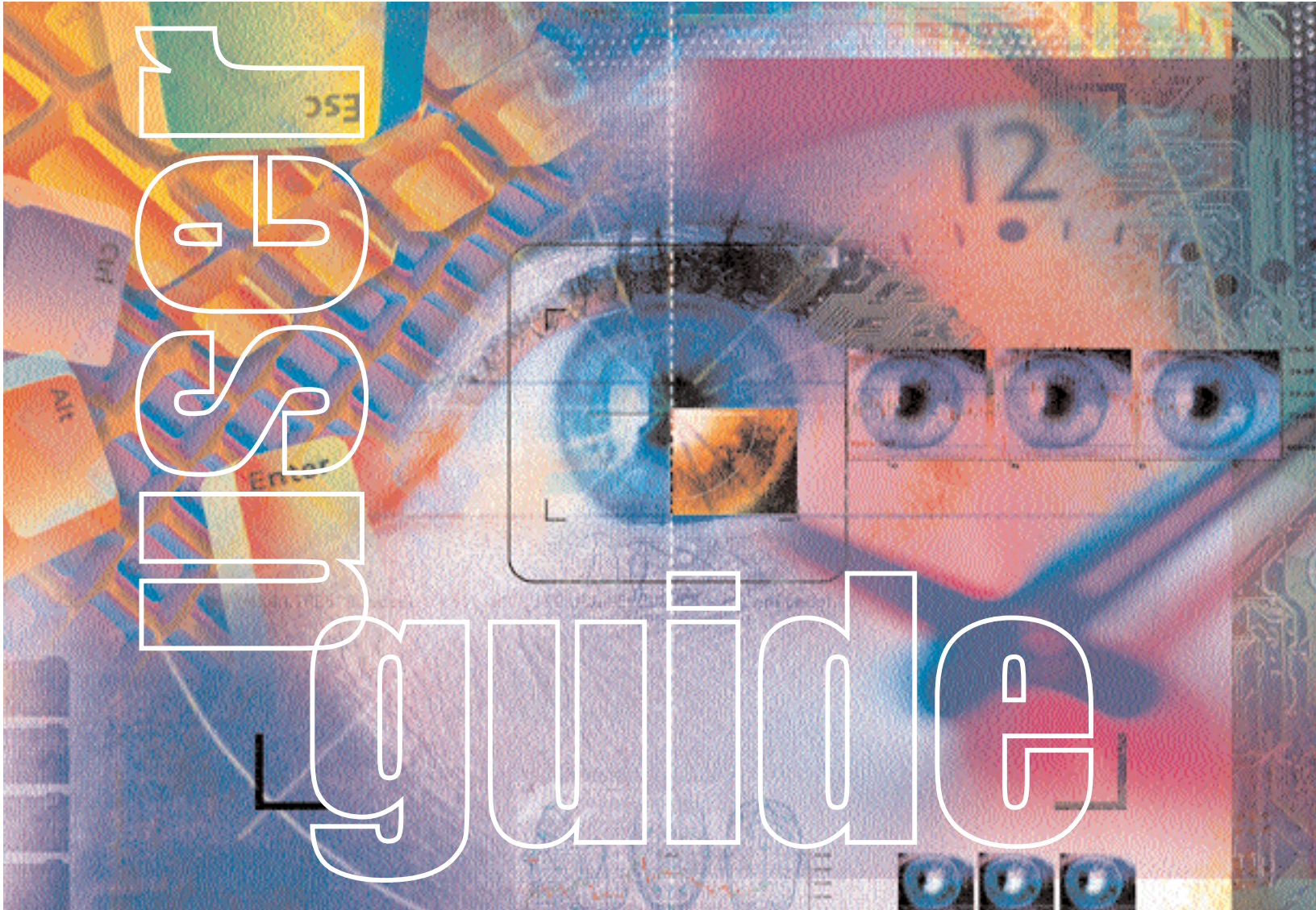
Make informed purchase decisions based on your specific business needs, project goals, and technical requirements.

Table of Contents

- Planning for an Enterprise Content Management System
- Developing RFP Guidelines for an Enterprise Content Management System
- Processing Unstructured Documents: Challenges and Solutions

AIIM *User Guide*

Planning for an Enterprise Content Management System



Authored by Cynthia Hodgson

Planning for an Enterprise Content Management System

*An AIIM User Guide
By Cynthia Hodgson*

*This is one in a series of User Guides from AIIM International.
They are intended to educate and inform readers on a variety of
enterprise content management topics.*

Copyright © 2004 by:
AIIM International
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
301-587-8202 / 800-477-2446

ISBN 0-89258-404-1

Published by:



The Enterprise Content Management Association

www.aiim.org

Planning for an Enterprise Content Management System

Enterprise Content Management (ECM) is defined by AIIM International as the technologies, tools, and methods used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. As illustrated below, the scope of a total ECM solution can be quite complex.

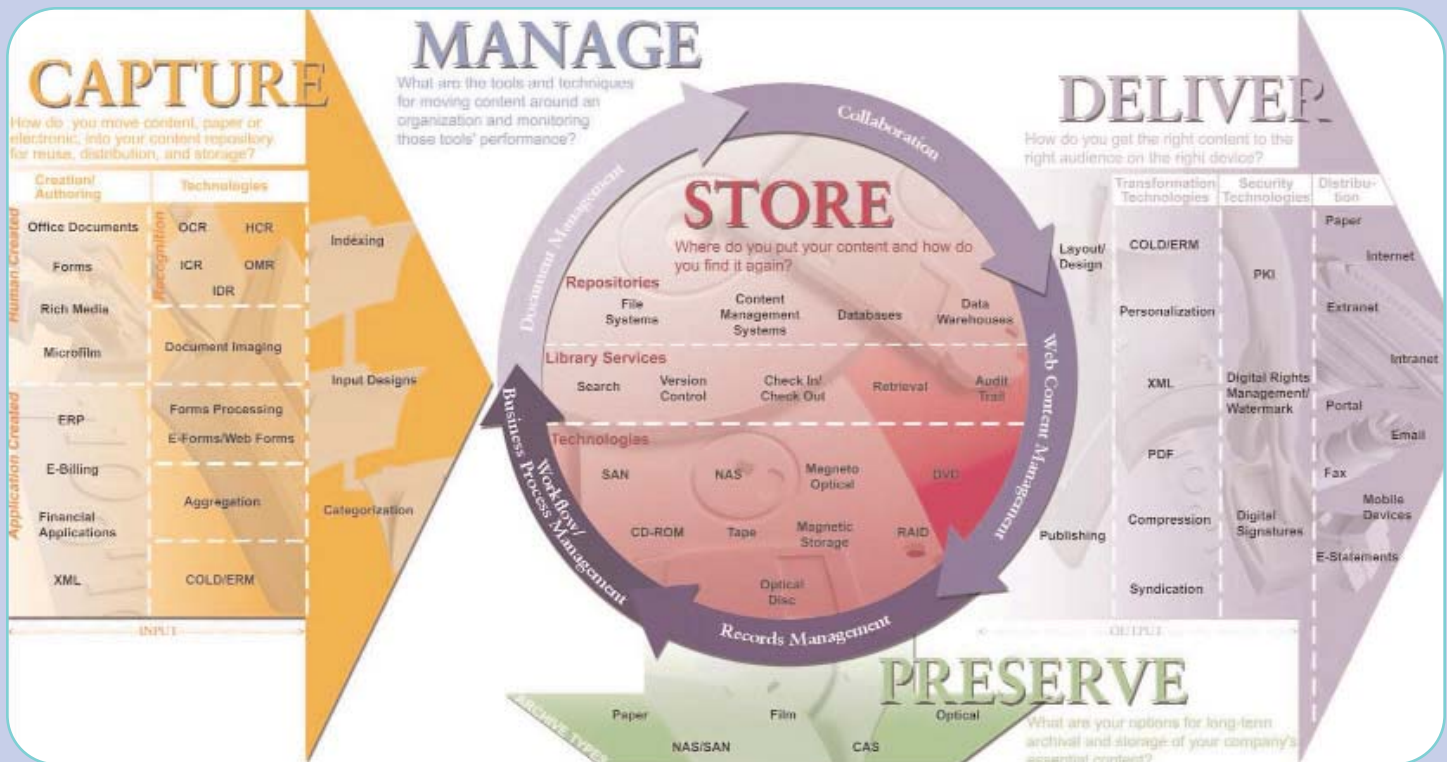


Figure 1: ECM 101 (Source: AIIM International and Doculabs)

Selecting and implementing the appropriate technologies and tools for ECM can be a daunting process. That process should start with the development of an effective strategy. This paper explores the need for an Enterprise Content Management (ECM) strategy and the issues to address in developing such a strategy.

The Need for an ECM Strategy

The huge amounts of information already in existence and the overwhelming increase of it every year have been widely reported in both industry and popular press. The volume of an organization's content alone could drive the creation of an ECM strategy, but there are additional forces making such a strategy even more critical.

Multimedia Content

Electronic documents are no longer a single entity; nor are databases made up of discrete records. Instead, a multitude of objects, including text, data, graphics, images, metadata, styles, formats, etc., are mixed and matched to create a collection of "documents." Improved technologies for streaming electronic sound and video have added to the complexity of object types to manage. Consider also that a single published "document" may have translations into multiple languages, multiple revisions, and versions regionalized with local information.

An organization must now find ways to manage all of these components of information as well as any combinations and versions of them that are created and "published" throughout the entire lifecycle of the information.

Wired (and Wireless) World

Employees, partners, customers, and suppliers have long expected to have electronic access to information. This expectation has grown to include remote access and increasingly wireless access. Wireless and thin client mobile devices often require information to be structured differently. Content may have to be specially packaged for wireless users and even “transformed on the fly” for mobile devices.

Online Content

Quality content is increasingly seen as an imperative for business success and even survival. Organizations that develop great-looking and features-rich websites find that without equally rich content, their customers don’t come back. Add the requirements that the content must be accurate, consistent, and timely, and organizations have quite a challenge, especially if they want to do it in a cost-effective way.

This content is not all created in-house. Customers, suppliers, and partners may be providing portions of the content on an ongoing basis. That requires an organization to manage content over which it has no control during the creation phase. It also means that these external entities may have to be involved or at least consulted in the development of the ECM strategy.

Need for Dynamic and Personalized Content

Prior to the days of the Internet and the World Wide Web, content mostly existed in the form of pre-formatted documents or database reports, mainly for internal use, and it changed on a regularly announced frequency, if at all. Now content needs to be made accessible beyond the walls of the organization, and users have come to expect content to match the capabilities of a Web browser type interface. Thus content must change dynamically, even instantaneously, at the initiation of the user and must often be personalized for each specific entity or individual user. Dynamic and personalized content can also be driven by the content-providing organization itself that needs to reduce its cost of providing customer service without reducing the services.

What Content Do You Manage?

In many organizations, content management is synonymous with “Web content management.” But even if you limit your content scope to what is accessible from a Web browser, that content is no longer isolated to separately created Web pages. The Web has become an access and fulfillment vehicle with the content derived from numerous back-office processes,

applications, and data stores. Enterprise Content Management processes and technologies become the connection between the front-end and back-office layers, as shown in the following illustration.

An ECM strategy must identify not only which content needs to be managed but also the priority in which ECM processes and technologies will be implemented with different content. Some organizations find it useful to do a formal content audit or inventory. Answering the following questions about the different categories of content should help set the priorities.

Who is the customer for the content?

Are the “customers” internal or external to the organization or both? How does the customer use this content? Are there ways to add value to the customer’s use of this content? Have the customers been asking for more, better, or different access to the content?

Which content has the most value?

The standard 80/20 rule is useful here—20% of the content drives 80% of the value. For a profit-making organization, value will most often be defined as that content which will contribute to increased revenue or reduced costs. For a non-profit organization, value is content that supports the mission or reduces costs. Organizations, profit or non-profit, that have a product or service orientation may give greater emphasis to the increased revenue and mission support aspects, while organizations with a process orientation would emphasize content management for its cost reductions.

Which content is most complex to manage?

Complexity may be derived from a number of content attributes, such as:

- The need to create the content from many different components of information
- The multimedia nature of the content components
- The large volume of existing information, possibly in legacy systems or formats
- The frequency of creation or acquisition of new information
- The frequency of update or revision
- The need to re-use the same content in many different processes
- The need to offer the same information in multiple languages
- The need for varied security controls to allow different user groups to access different portions of the content or different privileges of access, such as editing

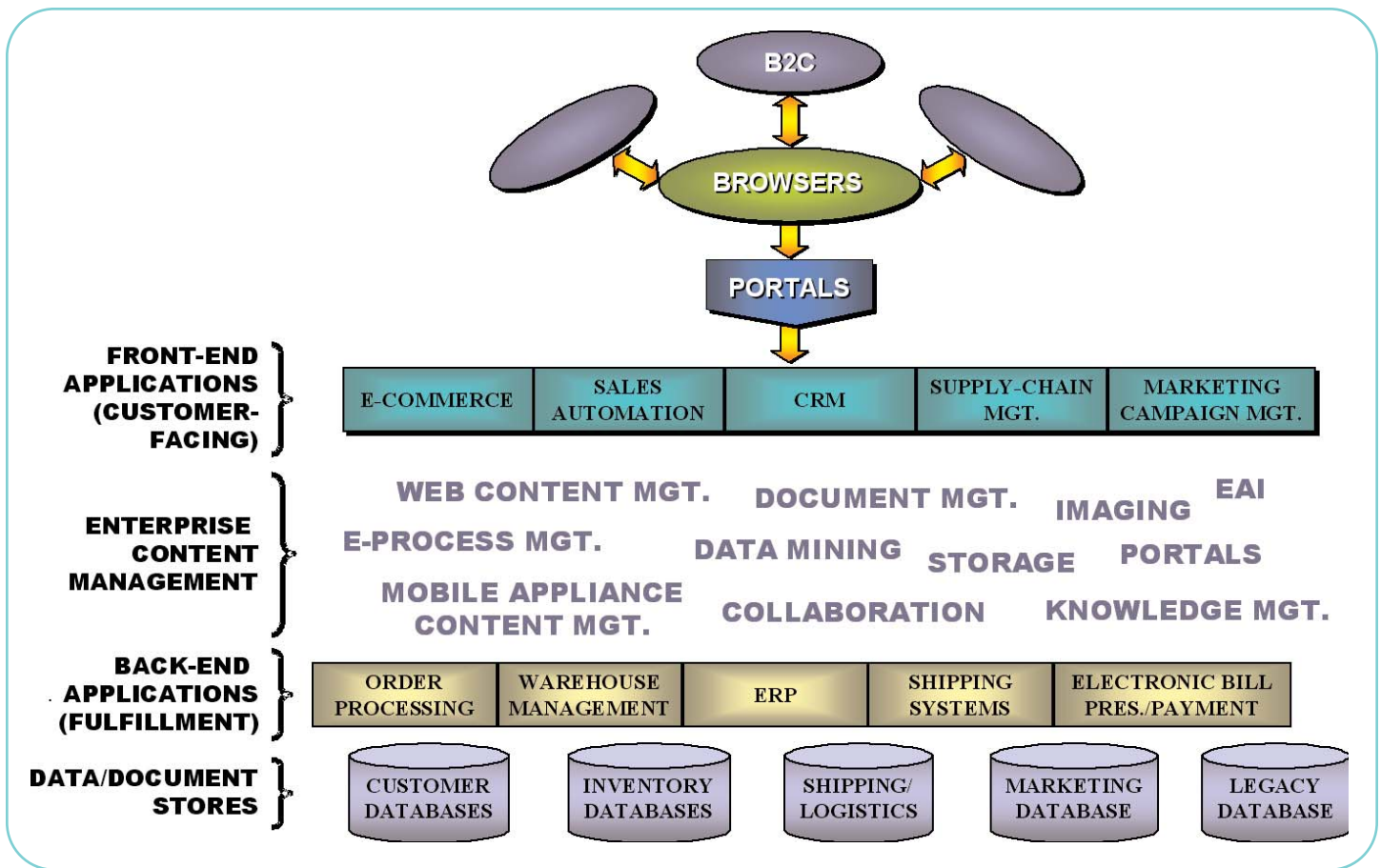


Figure 2: ECM provides the connection between back-end and front-end applications (Source: AIIM International)

Which content has legal or regulatory requirements regarding its retention or audit trail?

Legal or regulatory requirements, which generally relate to retention, add the complexity of time to content management. The lifecycle of the information could be very long, potentially outlasting several different storage technologies, and making a migration strategy critical.

Which content is received from external customers, suppliers, and partners and needs to be re-purposed for internal use or external access?

Applications frequently combine content from multiple sources. Often, the content has to be edited, reformatted, or mapped to internal content prior to making it available. Automation and standards compliance is usually necessary to make the repurposing process cost effective.

Beyond Technology

ECM is not just about technology. There are many decisions and process issues that need to be addressed prior to the implementation of any content management technology solutions.

Culture for Information Sharing

Organizations need to recognize and understand their current culture regarding ownership and sharing of information. New content management processes and technology can run counter to the existing culture causing the failure of the ECM initiative. The human behavior “change management” aspects of planning for ECM cannot be overemphasized.

Enterprise-wide Deployment

AIIM’s *Worldwide Enterprise Applications* study² found that “the most common scope of deployment (40% of respondents) for content management systems is across multiple departments and applications. While these are not true enterprise-wide implementations, they do cover a significant portion of the organization in many cases. And approximately one quarter of respondents have deployed content management applications either across their entire U.S. operations and/or globally.” Large enterprise deployments, whether global or domestic, require planning and up-front decisions on a number of fronts:

- What aspects of content management will be centralized and which will take a distributed/regional approach?

- Will the same technology be implemented everywhere even if content is managed regionally? If a common technology approach is taken, are there existing regional technologies in place that will have to be migrated?
- How do content providers in different locations collaborate on the creation of content?
- How do you ensure that global content, such as product specifications, is consistent across regions while still allowing for localization of related content such as marketing, sales, and pricing information?
- How do you accommodate the language issue so that critical content can be provided in the user's language of choice across the globe?
- How do you address countries' different legal requirements regarding restricted access to some information on a website?

Content Lifecycle

Management of content needs to begin at the time of creation and continue through archiving and final disposition.

Process decisions need to be made about:

- Which content requires approval prior to delivery and who needs to approve it?
- Which content requires audit trails when changes are made and what should that audit trail identify (what was changed, who changed it, when was it changed, etc.)?
- What are the retention requirements for each content type or category and how will the content be kept (in a production system, archived but online, archived offline, in magnetic or optical or micrographics media)?
- What is the migration strategy for content that must be retained beyond the expected life of the technology and/or media?

Metadata

Effective content management is no longer possible without the use of metadata. Metadata, the "data about data," is what ensures that users can locate the content they need when they need it. Searching on words in the full text of documents has been shown to be insufficient to effectively locate information. And full-text searching doesn't address retrieval of non-text objects such as photographs or images. AIIM's *Worldwide Enterprise Applications* study² makes several recommendations about how to provide metadata:

- The most effective kind of metadata is gathered from content creators, managers, editors, and users—those who are most familiar with the content. Metadata might

include themes in the document, contexts in which it can be applied and related terms.

- Depending on the context, a controlled vocabulary for index terms might be supplied to the user to use either on a mandatory or voluntary basis.
- Usage metadata can automatically be added each time the document is accessed, revised, or updated.
- Using document templates is an effective way to gather metadata, which can then be used to place the new content in an overall organizational scheme.
- Establishing document hierarchies and indexing are the most increasingly important steps in implementing a content management system.

Intellectual Property Concerns

Protection of an organization's intellectual property should be a key consideration when developing ECM processes. For example:

- Corporate branding and image needs to be maintained globally, a particular issue when there is regional decentralized deployments of content.
- Too early or inadvertent "publishing" of information can affect patentability.
- Information made available to the public is also available to competitors.
- Electronic content is especially easy to copy. On the other hand, Internet access makes it easier to find the unauthorized copying and use of an organization's content by others.

Likewise, care must be taken that others' intellectual property protections are not violated. Copyright laws are not well understood, and many people mistakenly believe that anything without the official © symbol is in the public domain, free for anyone to use. Anyone with responsibility for the creation side of content needs to be educated about copyright laws to avoid the inadvertent violation of others' property.

Technology Solutions

The complexity of effective ECM makes technology indispensable to any solution. But ECM is not a single technology. Figure 3 illustrates the key capabilities that are needed in each of the major ECM areas.

Given this diversity of capabilities, it is not unusual for an organization to need three or more vendors' products. Often one vendor's product becomes the focal point for ECM and other products are integrated with it to provide additional capabilities. The content analysis that should have been done, as discussed above, can determine the best approach for choosing

Capture	Templates & Forms	Imaging & Recognition	Aggregation	Categorization & Indexing
Manage	Collaboration	Approvals & Digital Signatures	Workflow	Security & Access Controls
Store	Databases & Repositories	Search & Retrieval	Version Control	Check-in & Check-out
Preserve	Retention Schedules	Media Types	Offline Accessibility	Migration
Deliver	Component Assembly	Personalization & Notification	Reporting	Format Conversion

Figure 3: Critical ECM Component Capabilities

the focal technology. Although consolidation in the industry is creating overlap and erasing distinctions, vendors can generally be placed in one of the following categories.

- Companies with a **document management heritage** have taken their traditional document management or imaging systems, which were focused on documents and unstructured text, and expanded the functionality to address integration with structured data and the delivery arena, especially Web-based delivery. They tend to be especially strong in the storage repository functions, workflow, and version control.
- **Web content management** companies focus on Web content formatting and delivery development. Content that is not intended for Web-based delivery is not usually addressed. They tend to be strong in the design and template creation areas and the personalization and customization aspects of delivery. Some of these vendors are expanding into the ECM area by adding capabilities to support transactional activities.
- **Supply Chain** vendors focus on managing the entire transaction utilizing a Web-based interface. They tend to be strong in product catalog management, databases and data warehousing, and security. Many of them have partnered with or acquired content management companies to expand their capabilities into documents and unstructured content.
- **Niche providers** offer solutions to any one of the subsets of capabilities required for content management. They are most common in the imaging and forms processing areas of the capture segment, the workflow area of the management segment, the reporting area of delivery, and the media / migration areas of preservation. Often they will

partner with vendors in one of the other categories to simplify the integration process with their product. While their solutions would not be an organization's focal technology, their products often have to be added on to provide the full capabilities needed for ECM.

In addition to the functional capabilities associated with capture, manage, store, preserve, and deliver, the choice of a technology solution should take into account typical IT considerations such as:

- **Infrastructure and platform compatibility**
Organizations typically want to leverage the investment in existing IT infrastructure. Utilizing existing technology platforms can also reduce the learning curve of IT personnel who will be implementing and supporting the technology.
- **Scalability**
Few organizations will attempt a global implementation out of the starting gate. But once a successful initial deployment has been done, easy and quick scale-ups will be needed to expand the scope. Scalability is also an issue if a deployment becomes so successful with the end users that its volume of use increases beyond the capabilities of the initial deployment platform.
- **Performance**
Downtime and poor response rate are problematic enough when encountered by an organization's own personnel, but they can be devastating when customers are directly accessing your information system.
- **Time required for deployment**
The vendor's past experience in deployment time should be determined for an initial prototype, a full deployment, and subsequent enterprise expansion deployments.

- **Integration capabilities**

Integration potential and ease needs to be considered for linking in third party ECM products, legacy back-office applications, and for content-needy or content-creating applications, such as CRM (customer relationship management) and ERP (enterprise resource planning).

- **Training requirements**

The amount of training needed and the availability of training courses and materials should be identified for at least four different groups: the content creators, the content managers, the content users, and the system administrators.

- **Standards support**

Choosing a system with open standards support can make all the difference when integrating with other applications, migrating to a future system, and exchanging content with partners and suppliers. The appendix in this paper identifies some of the key ECM standards that AIIM and its members have developed or supported. AIIM's Recommended Practice, Implementation Guidelines and Standards Associated with Web-Based Document Management Technologies,³ discusses a variety of ECM standards and technologies in more detail.

Justification

Forrester Research's study⁴ of 53 enterprise-level CM deployments found that software license fees were typically \$100,000 to \$250,000. Professional services for implementation support and training added at least \$180,000, and some \$200,000 was needed for development of controlled vocabularies and content hierarchies. Average total deployment cost for the survey respondents was \$1.8 million.

That kind of cost requires a solid benefits analysis to justify the expenditure. The list below has examples of the tangible and non-tangible benefits that an organization can expect to see after implementing ECM.

Productivity Savings

Productivity savings, read: lower staffing costs, can come from many different aspects of the ECM implementation. Gartner Group predicted² that "if enterprises do not bring both internal and external content under control, the amount of time wasted by the average knowledge worker on document-related, non-value-added tasks will increase to between 30 percent and 40 percent of their time. After investing strategically in content management, enterprises usually can save at least half of the

time and money now spent on non-automated document [content] management."

These productivity savings can be accomplished through using ECM tools such as: templates and workflow—to streamline the process of content creation, review, and approval—or automated tagging to reduce errors and the time spent on doing manual mark-up. Substantial savings can be made in the time spent on searching for or waiting for the right information. Field personnel, for example, don't have to wait for documents to arrive from the central office. Content captured at the source reduces further handling time downstream, not to mention the reduction in paper and paper-related storage.

In many organizations, the IT staff is actively involved in the administration of content on their website in addition to the traditional IT systems administration. An effective content management system can remove IT personnel from having to manipulate and manage content and focus their skills on more value-added IT functions.

Reduced Customer Service Costs

Productivity savings can also come from customer service support. With real-time access to a customer's information, the customer service representative can reduce the duration of calls, eliminate the need for callbacks, and increase the call volume per person. Companies that have created online access to an individual's personal information, whether an employee's benefits, an investor's portfolio, or a consumer's credit card charges, can decrease the frequency of calls to the customer service center. And when the employee or customer is also given some transaction capability, such as bill payment or account transfers, the costs of processing these transactions is shifted away from the organization's staff.

Savings related to customer service can be more than just staff productivity. Phone and fax expenses can be reduced when customers are interacting more with the online support service and less with the call center. Information can be transmitted to the customer without the costs of printing and shipping.

Improved Customer Service

The impact on customer service goes beyond cost savings to actually improving the customer service experience. Customers can obtain many services online 24x7, accessing the services at their own convenience. Often, customers have access to valuable information that has never been available to them before. Personalization and notification services add to customer "delight" and can foster increased customer loyalty.

Increased Revenues

The customer loyalty created by improved services typically translates into increased revenues. Press stories abound of companies that have significantly increased their customer base and revenue from their Web presence. But this Web presence must contain quality, updated, and meaningful content to show such results. The increased business and use of the website can itself become the driver to implement more robust content management tools. With e-commerce, transaction purchases can be automated, so orders are automatically placed and products tagged for shipping based on an inventory level established by the customer.

Improved Quality of Content

While difficult to measure, quality of content should definitely improve, decreasing the amount of outdated, inaccurate information perpetuated throughout the organization, and ensuring consistency of content that is presented to the customer. The ability to break content into components not only improves the reusability of content but also allows content from multiple sources to be integrated in a way never possible before.

Feedback for Continuous Improvement

The ECM system can itself create valuable information by collecting data on which content is used, by whom, and how. Reports and statistical data gathered by the system can be used to identify process improvements and product/service enhancements.

Conclusion

The growth in the volumes and types of information, the emergence of wireless technology and the demand for more dynamic and personalized content are driving organizations to recognize the importance of developing a strategy for Enterprise Content Management.

The most critical success factor in implementing ECM is not implementation—it's planning. Developing the ECM strategic plan begins with understanding your current business processes, the existing content, and the customers of that content. Before jumping into technology solutions, non-technology issues around cultural behaviors, regionalization of content, metadata, retention requirements, and intellectual property have to be considered. The selection of a technology solution will be driven by the content analysis and the decisions made about non-technology issues. Organizations that have previously implemented ECM have identified tremendous benefits that can be used in justifying a new initiative.

Once you've planned big and globally, you want to start small and locally to achieve a quick success. Then you can expand both the functionality and the organizational scope of deployment. And remember to stay focused on your strategy and the goals you established in that strategy.

References

- ¹ John Mancini, *Enterprise Content Management: Critical Technologies for Business Applications*, AIIM International, 2001.
- ² Gartner Group, *Enterprise Applications: Adoption of E-Business and Document Technologies 2000-2001*, AIIM International, 2001.
- ³ AIIM Recommended Practice, *Implementation Guidelines and Standards Associated with Web-Based Document Management Technologies*, ARP1-2002.
- ⁴ John P. Dalton, *The Forrester Report: Managing Content Hypergrowth*, Forrester Research, January 2001.

Appendix

Standards Relevant to Enterprise Content Management

Capture Standards

ANSI/AIIM MS44-1988 (R1993), Recommended Practice for Quality Control of Image Scanners

Defines procedures for the ongoing control of quality within an electronic image management system.

ANSI/AIIM MS52-1991, Recommended Practice for the Requirements & Characteristics of Original Documents Intended for Optical Scanning

Describes the physical characteristics of original documents which will facilitate scanning of the documents.

ANSI/AIIM MS55-1994, Recommended Practice for the Identification and Indexing of Page Components (Zones) for Automated Processing in an EIM Environment

Identifies a media and application independent structure and indexing scheme of document pages and zones within a page.

ANSI/AIIM MS61-1996, Application Programming Interface (API) for Scanners in Document Imaging Systems

Provides a common programming interface between device dependent software and document image scanners.

ANSI/AIIM TR15-1997, Planning Considerations, Addressing Preparation of Documents for Image Capture

Provides planning information to organizations considering image capture as a means of converting an existing record collection.

ANSI/AIIM TR27-1996, Electronic Imaging Request for Proposal (RFP) Guidelines

Provides step-by-step procedures for analyzing system requirements, developing functional specifications, and evaluating configuration alternatives.

ANSI/AIIM TR32-1994, Paper Forms Design Optimization for Electronic Image Management (EIM)

Describes how to minimize the costs of EIM forms processing by choosing the appropriate scanning technology and form designs.

ANSI/AIIM TR40-1995, Suggested Index Fields for Documents in Electronic Image (EIM) Environments

Sample index fields are provided for processing and retrieving information captured for use with EIM systems.

Manage Standards

ANSI/AIIM TR35-1995, Human & Organizational Issues for Successful EIM System Implementation

Provides a fundamental framework for understanding the basic issues and concepts of organizational factors, human factors, and ergonomics for Electronic Image Management (EIM) systems

WfMC-TC-1003, v1.1 (Jan95), Workflow Reference Model

Describes the workflow system architecture.

WfMC-TC-1009, v2 (Jul98), Workflow Client API Specifications (WAPI)

Defines application programming interfaces for process, activity, worklist, and process definition workflow operations.

WfMC-TC-1012, v2 (Dec99), Workflow Interoperability—

Abstract Specification

Defines logical message sequences and contents for interoperating workflow with other systems.

Storage Standards

ANSI/AIIM MS53-1993, Recommended Practice; File Format for Storage & Exchange of Image; BiLevel Image File Format

Specifies a self-contained file format for bi-level image file transfer in environments other than facsimile.

AIIM Technical Report, Integration of Electronic Document Management Systems and Electronic Records Management Systems—Functional Requirements (in development)

Describes a framework and best practices for integrating EDM and ERM systems.

DMA 1.0 (1997), Document Management Alliance Specification

Defines an architecture and object model to enable interoperability between document management applications and systems.

ODMA 2.0 (1997), Open Document Management API

Specifies the application programming interface for desktop applications to use when integrating with document-management systems.

ANSI/AIIM TR21-1991, Recommendations for the Identifying Information to be Placed on Write-Once-Read-Many (WORM) and Rewritable Optical Disk (OD) Cartridge Label(S) and Optical Disk Cartridge Packaging (Shipping Containers)

Outlines recommended information that should be placed on optical disk cartridges and optical disk cartridge packaging (on a physical label or other printed surface) for the purpose of identifying the optical disks.

Preservation Standards

ANSI/AIIM TR25-1995, The Use of Optical Disks for Public Records

Provides guidelines for the planning, implementation, and operation of optical disk systems in applications involving long-term and permanent public records.

ANSI/AIIM TR31 (R1999), Performance Guideline for the Legal Acceptance of Records Produced by Information Technology Systems

Four part report describing legal acceptance criteria and providing a systematic approach for implementing compliant record-keeping practices.

ANSI/AIIM TR28-1991, The Expungement of Information Recorded on Optical Write-Once-Read-Many (WORM) Systems

Provides uniform practices for the removal of information recorded on WORM disk media in response to court orders.

ISO 15489, Information and documentation — Records management

Specifies international best practice in management of records in all formats and media. Accompanying technical report describes implementation methodology.

Document Management—Long-term electronic preservation—Use of PDF (PDF-A) (ISO standard in development)

A developing international standard that will define the use of the Portable Document Format (PDF) for archiving and preserving documents.

Delivery Standards

ANSI/AIIM TR33-1998, Selecting an Appropriate Image Compression Method to Match User Requirements

Provides practical methods for selecting an appropriate and optimal image compression scheme which matches user requirements.

See the World Wide Web Consortium website (www.w3c.org) for key Web-related standards such as XML.

About the Author

Cynthia Hodgson is an independent information consultant and writer with over 20 years of experience as a corporate librarian, library manager, and information technology manager at Westinghouse Electric Corp. and Alcoa, Inc. She has taught graduate level courses in the Schools of Library and Information Science at the University of Pittsburgh and the University of South Carolina and has held local and national offices in several library and information professional organizations. She has previously published articles in *Database*, *EContent*, *CD-ROM Professional*, *Library Management Quarterly* and numerous library / information newsletters. She is the author of the NISO publication, *The RFP Writer's Guide to Standards for Library Systems*. Cynthia Hodgson is currently located in Pittsburgh, PA and can be reached by email at cahodgson@earthlink.net.



AIIM International

1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
800-477-2446 / 301-587-8202
aiim@aiim.org
www.aiim.org

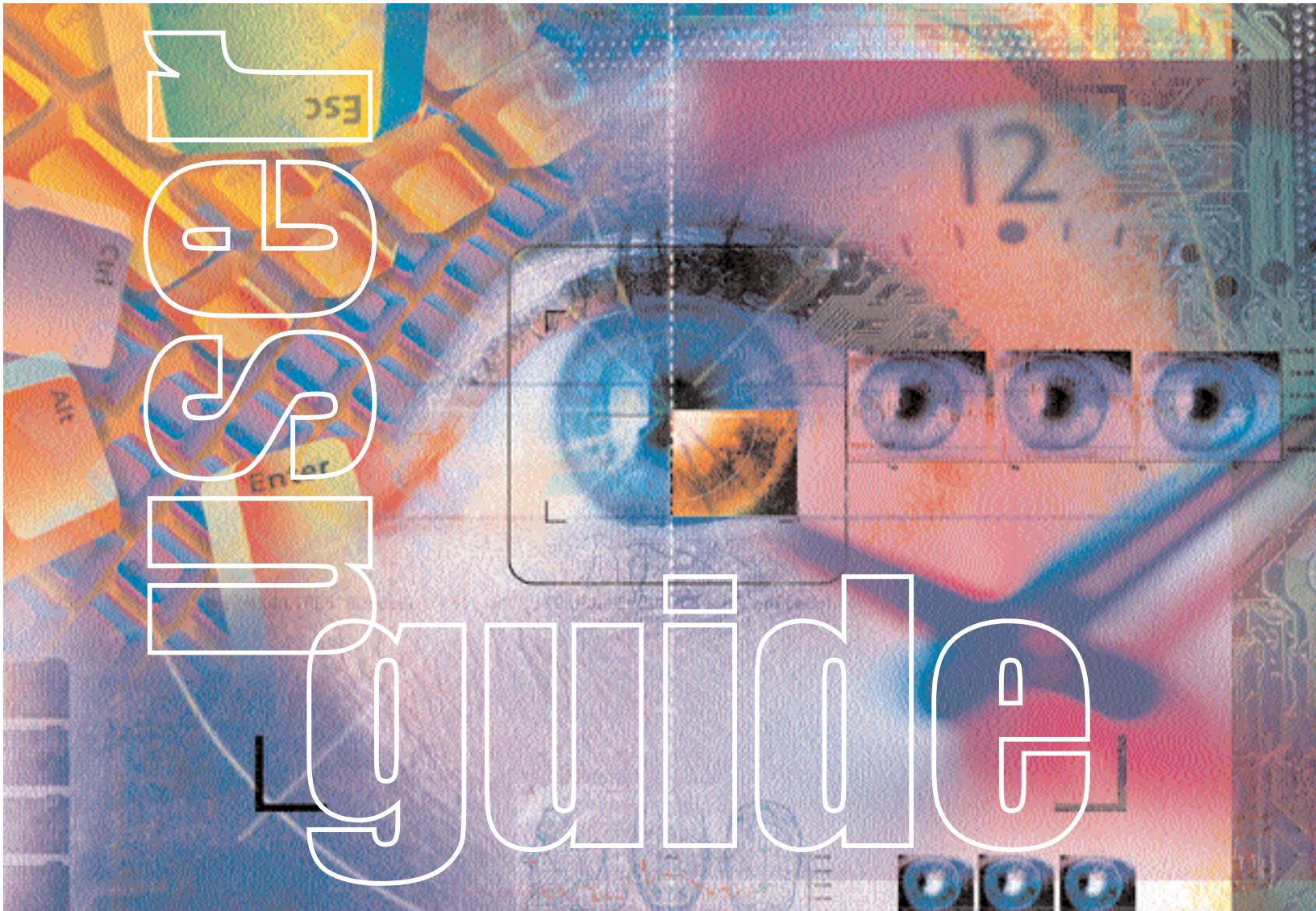
For over 60 years, AIIM has been the leading international organization focused on helping users understand the challenges associated with managing documents, content, and business processes. Today, AIIM is the leading international authority on Enterprise Content Management (ECM). ECM is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and technologies provide solutions to help users with the four C's of business: CONTINUITY, COLLABORATION, regulatory COMPLIANCE, and reduced COSTS.

AIIM provides:

- **Market Education**—AIIM provides unbiased information through *AIIM E-DOC Magazine* and *mID* (Managing Information and Documents), the leading industry magazines in, respectively, North America and the UK; its 20-city Content Management Solutions Seminar in the U.S. and Canada; the IM Expo event held throughout the UK, and Inf Ireland.
- **Professional Development**—This industry education roadmap provides a variety of opportunities. IM University is a multi-faceted program offered in Europe. The Web-based Fundamentals of ECM Certificate Program familiarizes users with the core concepts and technologies related to ECM. The AIIM Webinars round out user education on key issues.
- **Peer Networking**—Through chapters, networking groups, programs, partnerships, and the Web, AIIM creates opportunities that allow, users, suppliers, consultants, and the channel to engage and connect with one another.
- **Industry Advocacy**—AIIM, as an ANSI (American National Standards Institute)-accredited standards development organization, acts as the voice of the ECM industry in key standards organizations, with the media, and with government decision-makers.

AIIM *User Guide*

RFP Guidelines for an Enterprise Content Management System



Authored by Bud Porter-Roth

RFP Guidelines for an Enterprise Content Management System

An AIIM User Guide

By Bud Porter-Roth

*This is one in a series of User Guides from AIIM International.
They are intended to educate and inform readers on a variety of
enterprise content management topics.*

Copyright © 2004 by:
AIIM International
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
301-587-8202 / 800-477-2446
ISBN 0-89258-403-3

Published by:



The Enterprise Content Management Association

www.aiim.org

RFP Guidelines for an Enterprise Content Management System

Enterprise content management systems can range from simple scan, store, and retrieve systems to complex highly integrated systems that manage documents, records, and other data types throughout the corporation. AIIM defines enterprise content management (ECM) as "...the technologies used to capture, manage, store, preserve, and deliver content across the enterprise in support of business processes." ECM systems are used around the world and across commercial industry, federal and state governments, and county and city governments.

Driving the ECM industry is the overwhelming need to control, manage, collect, consolidate, and maintain business records—be they paper, word processor files, data elements in a database, email, or the ephemeral content of a catalogue Web page. Not only is the world's usage of records and information growing, but our need to control and make these records available is also growing. Instant and accurate data is the lingua franca among buyers/vendors, buyers/customers, and governments/citizens.

Pushing this industry along, as if adding an explanation point to the whole industry, is legislation that requires companies to be able to maintain corporate records according to government policies and requirements. Legislation is driving many companies to adopt their initial records management policies and ECM systems. Legislation is also driving companies with existing systems and solid record keeping policies to re-examine everything from the ground up: re-emphasizing the need for records management at all levels of the corporation.

This user guide is about how the buyer assembles the requirements for an ECM system into a formal request for proposal (RFP). An RFP is a document that defines a business problem or need and provides enough information to allow potential vendors to propose a solution. Sent to potential vendors, the RFP seeks to establish a competitive environment that allows the buyer to select the best solution at the best price. The winning proposal becomes the foundation for the contract, which defines project tasks and performance goals.

The RFP and proposal form a contractual bond between the buyer and vendor. The contract provides both buyer and vendor a common agreed-upon vehicle from which the project can move forward. The contract provides both parties with an understanding of the performance measurements, key project milestones, payment options, and a means for settling contract disputes should they arise.

RFPs are used to purchase ECM systems when:

- Multiple vendors can provide a solution
- Not all requirements and specifications are available
- The project requires different skills, expertise, and technical capabilities from vendors
- Pricing is negotiated with the vendor based on final schedules, products, implementation, training, etc.
- System functionality is more important than price

When a buyer writes an RFP, the buyer may be purchasing technology for which he or she is not familiar and may not understand all of the requirements and implementation issues associated with the technology. The vendor, on the other hand, may be aware that not all requirements are correctly stated and that additional work may be needed in order to adequately define the intended application. The RFP provides a common ground for both parties to understand the project needs, agree on a solution, and form a partnership to complete the project.

To this end, the RFP may be considered a "starting place" for the project. Vendors may ask questions about any of the RFP requirements and may challenge some of the requirements as not possible with currently available technologies. These questions may cause you to change the RFP requirements or to adjust the project timeline to be in line with vendor feedback. On the other hand, during evaluation of proposals, you may question statements in vendors' proposals and request they reevaluate their position. The final RFP and final winning proposal will typically be different from the original documents.

The RFP process can be time consuming, costly, and may not always produce the required results even after extensive vendor evaluations, demonstrations, reference checks, and modification to both the RFP and proposal. The reasons for a failed project are many and varied, but having the original RFP and the original proposal allows both buyer and vendor to review the documentation trail and determine where the project went

astray. Without this type of documentation as an audit trail, it would be nearly impossible to understand what happened and why. *Because ECM systems can be complex projects, it is suggested that an RFP always be used when contemplating the purchase of a system.*

RFP Planning and Development

An ECM RFP can be a significant undertaking that may require resources from different departments. As the purpose of an ECM system is to manage content across the enterprise, the data may originate in one department but be used by many departments, including the corporate intranet and Internet. Writing an ECM RFP requires the participation of departments that are responsible for the original data or work, support departments such as IT, and key stakeholders that have corporate budget responsibility.

An RFP will typically begin when a need is identified within a department or the corporation (or a government agency). For example, the Information Management (IM) department may determine that information presented on the corporate website is often out-of-date, which could make the corporation liable for the misinformation. In addition, the IM department may determine that corporate website pages are not being archived according to the records retention schedule. Based on the initial need identification, the IM department would begin to further study and identify where the problem originates, what departments share responsibility, and what the risks involved with outdated documents and potential solutions might be. The following steps outline the basic tasks that lead to writing an RFP.

Initial Need Identification

When a problem or need is identified, the first step is to do a preliminary analysis and determine whether this need is sufficiently credible to warrant further examination. The initial study may be done by the department that originally discovered the need, such as the IM department in the above example. The study may be a high-level review of the need that lists the basic issues, what risks are involved by not taking further action, what benefits may be gained as a result of solving the need, and what resources will be needed to move forward. The result of this task is to more formally identify the need and provide enough information to either proceed with the next steps, suggest alternate means of solving the problem, or to recommend that the team stop any additional work.

If the project is approved, the next step is to identify resources and form an RFP project team. The team is responsible for further

detailed analysis, identification of vendors, writing the RFP, and evaluating proposals.

Formalize the RFP Team

If the initial analysis determines that a real need exists, the next step is to gather a team of subject matter experts and technical resources. RFPs are a group effort. While each RFP team will be different, according to the need being reviewed, the basic team consists of the following:

- **Business Operations.** Operations is the department that conducts the business for which a technology will apply and may be the sponsoring department for the RFP. (Operations may be HR, loan servicing, accounting, or corporate communications, for example.) Operations staff understands the basic day-to-day work of the department, knows how they use information, and what the current problems are in a business process. In planning for a new technology or system, operations is the group that determines what work steps the new system should provide, how the information will be presented, and how a new system would improve upon old work methods. Business operations typically provide the following information:
 - Define the current work process
 - Develop the product's functional requirements
 - Show data flows into and out of the department
 - Operational problems that limit the current department functions
 - Operational processes that may not be changed (usually guided by state or federal laws)
 - Operational processes that can be changed or eliminated
 - Analysis of work volume and throughput requirements
 - Description of internal resources required (mainframe access)
 - Description of external resources required (regular outsourcing of work to a specialized company)
- **Information Technology (IT).** IT may be responsible for a variety of functions, and may also be the originating department for the need. IT is generally responsible for providing project management support for the RFP project and may provide consulting resources for the technical analysis of the issue being studied. In terms of contributing to the RFP, IT is responsible for such areas as:
 - Development of RFP technical requirements (based on the functional work description supplied by the operations staff)

- A complete description of the current technical environment
- Description of any technologies with which the new system must co-exist or communicate
- Gathering standards for the RFP, such as the communication's network, the GUI/workstation interface, programming and development standards
- Identification of potential technologies for the project
- Technical qualification of potential vendors
- Development of RFP project requirements

• **Purchasing/Procurement.** The third group in the planning process is the purchasing department. Purchasing is normally associated with post-RFP work such as negotiating contracts and developing payment schedules. However, purchasing can and should provide a valuable service during the planning of the system.

- Developing and maintaining a list of potential vendors
- Developing and maintaining an estimated pricing list for budget development and providing financial analysis of vendors and their pricing
- Participating with IT and business operations in the project cost justification by providing financial data and support
- Developing basic purchase agreements and additional amendments for the project
- Formalizing the RFP for vendors (the requirements and responsibilities) with IT and business operations
- Participating in contract negotiations with selected vendors
- Managing the on-going contract after award

There may be other business units that participate, depending on the project. For example, the Information Management (IM) department may be asked to consult on the records retention guidelines and policies for a project, Legal may be asked to review and provide an opinion on some aspect of the project, and HR may be asked to provide input on the potential issues with layoffs, outsourcing, or reassignments.

• **Product/Systems Education.** If the decision is to move forward with writing an RFP, the "team" may need to become educated on the various types of technologies that can potentially resolve the problem. As a first step, the team identifies potential vendors and begins inviting them to give presentations and demonstrations. The team may also attend industry conferences, participate in user groups, hire consultants to give presentations, and visit other companies using the technology.

As part of the education process, the team may also decide to benchmark the issue in question with other companies. A benchmark study helps the team decide what other companies have implemented, what problems they experienced, and whether the application of a technology solution was successful. Benchmarking is a very useful tool when contemplating an ECM system and helps the team establish guidelines and boundaries for the project in addition to gaining further insight into the vendor community.

Education is imperative to understanding the problem and how technology can or cannot resolve it. Without adequate education, the team may consider technologies that are not adequate, which results in eventual project failure.

Educating the team is also necessary to properly identify which vendors receive the RFP.

• **Business Process Analysis.** The first step in identifying the business problem/need is to perform an analysis of the business process. This analysis can take many different forms depending on the project, but generally the analysis will be focused on the documents or business content (electronic, paper, data, etc.) and how they are used to perform the needed business operation. If we revisit the original example of outdated corporate Web content causing customers to view potentially obsolete material, the business process analysis begins with understanding how and where the original data was created, how it is updated, and how the updated data is distributed. For example, new data may be pushed to the content manager, or pulled, or it may automatically update a Web page.

The analysis then proceeds in documenting what other departments use the data and how staff receive and distribute it to still other areas of the company. Included in the analysis are the functional needs for the data, such as converting Microsoft Word files to Adobe's portable document format (PDF) or converting data files to XML and assigning metatags or indexing information to the data.

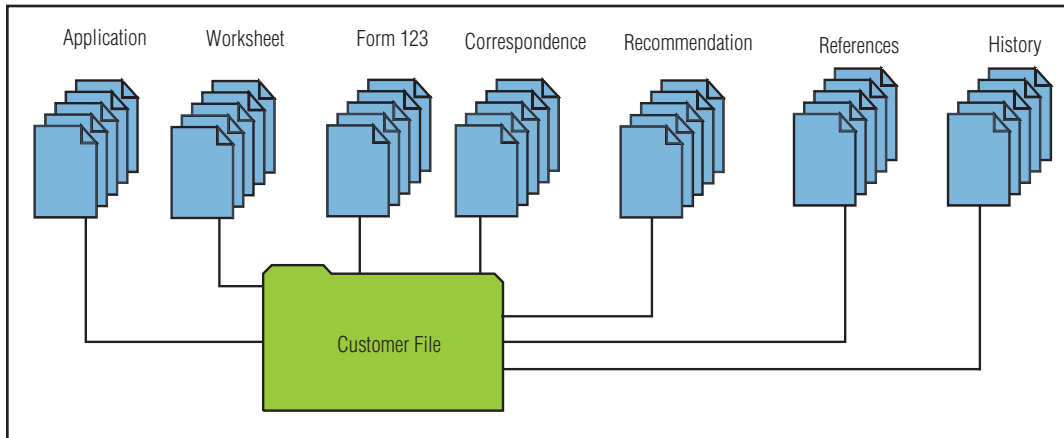
Once the "current work process" is documented, a baseline of information is created that can be used to develop the functional requirements for the RFP. *In other words, we have to completely document the current process(es) before we begin to reengineer them using new technology and processes.* The results of this study are then included in a written report documenting the process that may include tables of document types and volumes, workflow diagrams, and a collection of physical documents that are used in support of the work.

The illustration below is an example of a diagram used to illustrate how a customer folder is organized. This would be used in the RFP.

The table below is an example of the type of data needed to quantify the volume of work processed in a business operation and used in the RFP to provide the total number of docu-

cally convert on-line MS Word documents to PDF format for viewing by a user.”

Requirements development is the listing of system attributes based on functional work requirements. In the PDF example above, the current system may require a person to manually convert all MS Word files to PDF, manually index or assign



ments used in an application. This is only an example as there could also be columns for indexing requirements, processing time limits, etc.

The workflow diagram below is an example of a process flow that can be developed to show vendors the details of a work process. Vendors use this process flow to evaluate how

metatags, and manually upload it to the corporate Intranet. The RFP team understands that technology exists that can do this work automatically and makes this automatic conversion and indexing work a requirement in the RFP.

As a general rule, requirements should remain, as much as possible, functionally stated

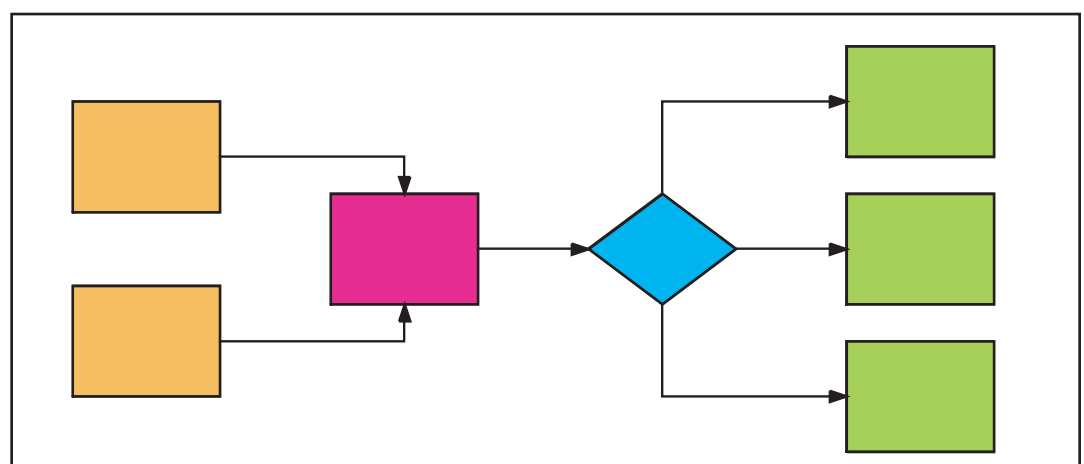
and should not “mandate” a solution to the vendor. For example, the functional need is to present website information to the viewer in a form other than Microsoft Word. The RFP should not dictate that PDF be used but allow the vendors to select the best technology to fit the need. One vendor may propose PDF while another proposes conversion to

Item	Document Name	Type	Quantity	No. of Pages	Condition
1	Application form	Paper	500,000	1,500,000	Well worn
2	Contract	Paper	500,000	2,500,000	Excellent
3	Application Notes	Word	500,000	500,000	Electronic
4	Credit check	Fax	250,000	400,000	Good

HTML. Either format may be acceptable for this particular project. By specifying the solution in the RFP (for example, PDF format), you will lose vendors that may have other acceptable solutions.

they are able to reengineer the process using their application software.

- Requirements Development.** Once a baseline is created, the next step is to review what changes can be made to automate a manual process or what changes can make a process more efficient through reengineering. Requirements define how a system or product operates to fulfill expectations. For example, an ECM system requirement may “allow for the manual addition of indexing data” or “automati-



Managing the RFP

As with any project, an RFP needs to be managed. The RFP is a project by itself and should have a project manager, plan, and schedule. The RFP manager will also need to acquire the facilities, resources, and people needed to successfully complete the RFP. An ECM RFP team would typically be comprised of the following:

- **Users (departmental).** Users are the people that work with the data in a business application. Users can be customer service representatives (CSRs) that process claims or respond to customer requests, HR personnel who are responsible for posting corporate policies, or accounting personnel who process invoices. These are the people that use the needed data for their daily work and who understand how their individual business process works. Users may be responsible for initiating the project—i.e., Accounts Payable may want to automate the AP process by capturing shipment data at the warehouse door, posting that data to a vendor accessible area on an extranet, and allowing vendors to link to a payment system.
- **Information Technology (IT).** IT will typically support a user-initiated RFP process and may initiate an RFP within their own department. When IT supports a user RFP, such as the AP automation process mentioned above, IT may provide the project management expertise, help map the current process by serving as in-house workflow consultants, and provide parts of the RFP requirements that a user may not have the technical capability to provide, such as the development and testing program requirements.
- **Purchasing.** Purchasing plays a direct role in any RFP that is being written. Purchasing may be able to identify the appropriate vendors for the project, may be able to provide financial reviews (such as a D&B report) of the vendors that are on the short list (the final two or three vendors), and would either handle or help in the negotiations and contract signing when a vendor has been selected.

The RFP project manager may also enlist the services of Legal to review a vendor's purchasing agreement, Human Resources to determine the impact of a workforce reduction as part of the project, or request the services of an outside consulting firm to provide technical support for evaluating responses.

The typical ECM RFP requires additional project resources and more extensive planning because it involves different business units and approvals from many different corporate stakeholders. It is the responsibility of the RFP project manager to

secure the necessary resources and people to complete the RFP. It should be remembered that an RFP can take as many as six calendar months (or more) to complete, and the team is required to participate at varying levels over that time period.

RFP Organization

Instituting a logical organizational process that vendors can easily follow is one of the keys to success when planning an RFP. Below is a suggested outline that covers the basic sections of an RFP. This suggested outline may be modified to fit your particular project. For example, you may add a section called "Contractual Information and Response" if your contractual requirements are particularly complex, and you want to ensure that vendors respond directly to your contractual needs. If your company is particularly security-conscious, you may separate the security section from the technical requirements and make it a separate section that vendors need to address in their proposals.

While an RFP can be assembled in many different ways, the following is a suggested outline for the major RFP sections:

1. **Cover Letter**
2. **Administrative Section**
3. **Technical Section (system requirements)**
4. **Management Section (project management requirements)**
5. **Pricing Section**
6. **Contracts Section**
7. **Appendices**

These sections are reviewed below.

Cover Letter

The cover letter is an important first page to your RFP. The cover letter introduces the RFP project and provides some of the most vital dates. The cover letter may also include anything special about the RFP that should be noted by the vendors.

Special dates often include:

- Proposal due date and time
- Bidder's Conference
- Bidder's Intent to Respond form due date

Special notices and comments include such items as:

- Brief introduction and description of the project
- Bidder's conference is mandatory
- Information in the RFP is highly confidential and bidders may not share the RFP
- Who to contact about the RFP

- Warning that vendors may not contact other people on the RFP team

The cover letter reinforces information that is found in the administrative requirements or other places in the RFP.

Information in the cover letter can help vendors determine quickly what the RFP is about, who should receive the RFP and be responsible for the proposal, and which are important dates. Below is an example of a cover letter.

Vendors:

The ACME Corporation is seeking a vendor who will provide a comprehensive packaged solution for our ECM needs. ACME requires a system or systems to manage electronic data in a variety of formats at a variety of locations. The system must be flexible in regards to ease of use for authorized users, system administration, integration with existing systems, changes in business rules, and routing of information to people inside and outside the department.

The attached RFP contains all instructions, requirements, and relevant dates for submitting your proposal. After your initial review, please sign and return your intent to bid form. If not returned, you may be disqualified from further participation. This RFP also requires that finalists will be required to provide a demonstration of their system. Failure to participate in the demonstration will eliminate you from further participation.

Key RFP dates are:

- July 30, 200X – Proposals are due by 4:00 P.M.
- July 15, 200X – Final questions are due
- June 15, 200X – Bidder's Conference
- June 10, 200X – Return Intent to Bid form

Please review the Administrative Section of this RFP for full details concerning all submission requirements for this RFP. All contact with ACME must be through:

Mike Smith
 RFP Team Leader
 msmith@acme.com
 415-555-1212

Example Cover Letter

The cover letter may also contain instructions to the vendors that they must return the "Intent to Bid" form or a Non-disclosure Agreement (NDA) prior to receiving the RFP itself. In some cases, the RFP may contain highly confidential and proprietary data, and you do not want to release it to the general vendor community.

Administrative Section

The administrative requirements section allows you to establish the rules and set the requirements for responding to the RFP. The administrative section is an important section for the RFP because it allows you to establish how vendors will contact you, how they will format their proposals, what rules must be adhered to during the RFP competition, and other important items that you want to specify. Some typical items in the administrative section include:

- Introduction (basic overview of the RFP)

- Schedule of events
- Contact names and addresses
- How questions will be handled
- Information about the bidder's conference
- Information about oral presentation and demonstrations
- How proposals will be evaluated
- The required format for vendor proposals
- General proposal submission information, such as number of copies, printed or electronic, etc.
- How alternate proposals will be handled
- How to include subcontractors in the proposal

While each of the instructions is important, there are three key instructions that will make your life easier when you are reading and evaluating vendor proposals. The first is direct all vendors to a single contact for any questions they may have. Furthermore, ensure that all questions are written and sent to the designated person. Do not accept any telephone questions, and do not allow vendors to contact anyone other than the designated person. The designated person can then review and distribute questions to the appropriate resources. With everything in writing, there is no opportunity for a misunderstanding.

The second key instruction is to explain how a vendor can identify a requirement in the RFP. Quite often, RFPs are written by a team of people who have different writing styles and different ways of specifying requirements. This results in a confusing RFP for the vendors, and they may not be sure of what is or is not a requirement.

In order to ensure that vendors can spot your requirements, there should be a consistent phrase that tells the vendor what follows is a requirement. For example, in Federal Government RFPs, the word "shall" indicates a requirement—"The vendor shall provide a detailed project plan by phase," or "Vendors shall describe what training is provided with the software." If you are using the word "shall" to denote a requirement, ensure that team members do not lapse into other methods of denoting a requirement such as the most common one, "Vendors are required to provide annual reports for three years," or "Vendors should provide annual reports..." Without some standardization, vendors will miss requirements, and an otherwise capable vendor may receive poor marks due to an unclearly written RFP.

The third is to specify how vendors must organize their proposals. One example of a proposal format may be:

- Cover letter
- Executive Summary
- Technical Response

- Management Response
- Pricing
- Appendices

This outline requires the vendors to provide proposals in a consistent format, which will facilitate your evaluation process.

The administrative information and requirements can be very helpful to both you and vendor. If this section is weak, vendors will ask more questions, and their proposals may not be as consistently formatted, which causes evaluation problems. The administrative section is also the first “test” of the vendor’s reading and compliance skills.

Introduction

The introduction provides the vendors with a very high level overview of the issues that are driving the RFP. Think of the introduction as an executive summary for the RFP. This summary will give the vendors a brief overview of the company and then move into a synopsis of the problems or issues that are the driving force behind the RFP. The introduction will help the vendor to better understand your company and the project. Below is an example of an introduction.

ACME Bank is a federally chartered, publicly traded bank that ranks among the nation’s top five banks and internationally among the top ten banks. Our main financial products range from private and personal banking to multi-national loans for large construction projects. ACME leads the field for large construction loans in North America and is among the top banks in Europe participating in multi-national and multi-bank construction loans.

ACME Bank is releasing this request for proposal (RFP) for the implementation of an enterprise content management (ECM) system for loan servicing and processing. ACME is seeking to implement a global solution with a potential user base of approximately 20,000 users worldwide and expects a solution of an appropriate scale.

Currently, loan documents and portfolios are managed in their paper document form, kept in secure vaults, and copies are faxed or couriered nationally and internationally when needed for review or signatures. Electronic document management technologies will eliminate the need for faxing and courier services while providing secure transfer, secure storage, faster access, and disaster recovery capabilities.

Example RFP Introduction

The above is an example of the type of information that can be found in an introduction. The introduction for your RFP will be longer, providing more detail about the company and the issues behind the RFP. The RFP may direct readers to an appendix (or a website) for additional detailed information about the company, locations, and products.

Technical Requirements

The technical requirements section is the heart of the RFP, and it is where vendors should spend the most time initially. This section provides the details for the amount and type of work that has to be accomplished by a product, a service, or a person. The requirements are the foundation for a vendor’s technical proposal, but also drive other sections such as project management and pricing.

The technical section can be somewhat difficult to write because it is a balance between describing what the current needs are versus describing what solution you expect vendors to provide. In some respects you have already described the solution with your choice of vendors and the technologies that they represent—for example, by selecting companies that specialize in scanned documents, electronic documents, workflow, etc., instead of companies that specialize in managing Web content or databases, you have chosen the type of technology you believe is correct for your application. It is appropriate to provide information on how you expect the future system to operate (in a benefits-oriented fashion) but not specify features that are unique to one vendor that other vendors cannot satisfy.

In some cases, it is acceptable to provide minimal requirements in an RFP in order to receive the most comprehensive and wide-ranging proposals. The down side to casting such a wide net is that you may receive a bewildering number of proposals with solutions that are only marginally acceptable, making it almost impossible to evaluate them. Thus, instead of spending quality time on the vendors with good solutions, you will find yourself spending hours upon hours weeding out the non-compliant vendors.

For example, if your project is to scan and store paper documents and make them available electronically for workflow processing, the requirements and solutions are fairly straightforward and the vendor community is manageable. Thus, you may have a tightly structured set of requirements, a limited number of vendors, and a good idea as to the final solution. On the other hand, if your project is to implement an enterprise-wide intranet that contains a variety of information from all departments, the solution may not be as clear-cut, and you may want to formulate requirements that are broader in nature to attract a more diverse set of vendors.

One last consideration is that by providing requirements that are so tightly focused that only one or two vendors can respond, you limit your choices in terms of products, prices, and vendors (Remember that an RFP promotes competition

among vendors.). *Tightly focused and overly restrictive requirements may be a signal to other vendors that a solution has already been chosen and that the RFP is merely an exercise to justify that solution.*

If possible, the requirements should reflect a reasonable understanding of the products and services that are being requested in the RFP. Requirements should include the following three characteristics:

- A requirement identifies a capability, characteristic, or quality factor of a system in order for it to have value and utility to a user.
- A requirement must be measurable in some manner.
- A product or service must exist in order to satisfy the requirement.

Requirements are the key components of an RFP. It does you and the vendor no good to have requirements that are difficult to spot or understand. If a requirement is in an RFP, it must represent something that is needed within the project.

Requirements can range from the project specifications to the product specifications. In order to write an effective technical requirements section, the RFP team must have the following knowledge:

- The team must know in detail how the current system operates (if there is one) and must be able to communicate that knowledge to the vendors.
- If there is not a current system and this is a new application, the team must be able to “design the system from scratch” since there is no existing system to use as a basis for the new system. This means that the RFP team may have to do considerable requirements analysis with the users in addition to researching and comparing requirements to potential vendor solutions.
- The team must know what is expected of the new system and provide vendors with direction.
- The team must know what constitutes acceptable solutions, given that they have several valid solutions from which to choose.
- The team must be able to intelligently evaluate the differences between the acceptable solutions.

The technical requirements section may contain the following information:

Introduction to the basic issues that are driving the RFP

Description and overview of the current work processes or business environment (This can be done in overview fashion with detailed workflow processes in an appendix.)

- Description of the current technical environment (This can also be completed in overview fashion with technical detail such as network diagrams in an appendix.)
- Description of the proposed work environment
- System technical requirements
 - Description of the work application
 - Description of communications/network needs
 - Description of security requirements
 - List of constraints (i.e., Windows xx environment)
 - Table or matrix of documents by numbers, sizes, condition, etc.
 - Table or matrix of number and levels of users

Requirements can be written in a variety of ways depending on what information you are trying to obtain from the vendor. A requirement may ask for information; it may be a question; or it may be a specification. Below are some example requirements.

- Vendors shall describe how your system supports and facilitates metatagging of documents.
- Vendors shall describe how you import existing websites and data into your software application.
- The software application shall support a variety of viewable formats including PDF, HTML, and Microsoft Office viewers.
- The software application shall allow users to personalize their Web page.
- The application shall provide individual user security options for personal Web pages.

The above examples are direct and show an understanding of the intended application and application software. The requirements also request specific information to which a vendor can directly respond. Try to avoid “requirements” in which a vendor will “dump” a complete data sheet in the proposal as a response to your requirements. This happens when a requirement is too broadly stated such as, “Does your application provide for website security?” This question invites vendors to either respond with a “Yes” or to dump several pages that give general information about the vendor’s security offering.

A second common problem with writing requirements is subjective and/or ambiguous language, as is the case with the following phrases:

- But not limited to
- Shall support at least 10
- System shall be flexible
- Shall have sufficient disk space
- Training shall be appropriate for at least
- Instructions shall be easy to read and user friendly

- System shall be intuitive to use by the user

These types of requirements beg for either the vendor to purposely misunderstand them or to supply a response that is not supportable. For example, “System shall be intuitive to use by the user.” A vendor could simply respond, “Our system is intuitive to use,” and the response would be valid since there is no definition or discussion of what “intuitive” means in this system context.

Since the technical requirements define the need in an RFP, they should be as clearly stated as possible while requiring as short an answer as possible. A quote from George Bernard Shaw says it all, “I’m sorry this letter is so long, I didn’t have time to write a short one.”

Management Requirements

The management section provides information on how the project will be organized and who will be responsible for certain tasks. This section may request the following items be proposed by the vendors based on unique project and requirements:

- Project management and implementation plan
- Project schedule (in a software application that both customer and vendor share)
- Site preparation requirements
- Application development plan
- Testing
- User testing
- Acceptance testing
- System acceptance test plan
- Maintenance description and schedules
- References
- Training (description of class and schedule)
- Documentation (paper manuals, CDs, online)
- Staffing requirements (do you need to add additional personnel?)

Testing, for example, is not a popular topic with vendors because it slows the overall project down and introduces an element of risk—“I’ve never seen that error message before!” Testing requires much work on both sides as you, the buyer, must fully support the testing plan by providing input, testing data, a testing environment, and people who can physically exercise the application(s).

An enterprise content management system can be a complex undertaking that potentially changes how you operate a department or a corporate infrastructure application like your

intranet. For example, the implementation of a corporate-wide electronic document management system to manage all electronic documents such as MS Word files can have a profound change on the company and all personnel. The addition of this type of technology can require significant development and testing activities with the vendor, require extensive training development (and training of the users), require major business operations changes for the users, and may significantly change the IT department to accommodate the new system. These types of changes to a corporation cannot be successfully implemented without a detailed project management plan.

The project management plan specified in the RFP and written by the vendor will be a starting point for the project. The initial project plan in the proposal will be the vendor’s “best guess” at the project based on the requirements in the RFP. Once the contract has been awarded, both the buyer and vendor will work together to produce a final project plan and schedule.

Pricing Requirements

Pricing for an ECM system can be a complex matrix of products and services. How do you compare apples to apples when you may have oranges and bananas in the mix? The key is to give the vendors clear guidelines for developing their pricing section. This means that vendors are instructed to break down their pricing into component areas such as application development, system software, project management, maintenance, etc. For a complex ECM system, the software may encompass several products from the same vendor or products from several vendors. For example, ACME ECM Systems, Inc. may provide the basic application EDMS software but team with other vendors for forms management, OCR, and input capture software.

Depending on your RFP and its application, you may develop a pricing spreadsheet and require vendors to use your spreadsheet to enter their pricing data. On the next page is a brief example of a pricing table that you would ask vendors to use.

Contracts Section

While not mandatory, it is advisable to include a draft contract or purchase agreement with your RFP. This allows vendors to review the draft and to determine whether they can comply or not. If vendors have a problem with the contract, they can review the issue in their proposal and perhaps suggest what is acceptable. This will alert and prepare both sides for potential issues, which is much better than awarding the contract to a

Item	Description	Unit Price	Extended Price
1	Application Development	\$150/hr	\$500,000
2	Application Software	\$225,000	\$225,000
3	Hardware	\$00.00	\$00.00
4	Maintenance	19%/year	\$137,750/year
5	Training	\$500/class	\$1,500
6	Documentation	Included	Included
Totals			

Sample pricing guide

vendor only to find that you cannot reach agreement on contractual issues.

Appendices

As part of an ECM project, you have most likely done extensive work in some or all of the following areas:

- Workflow analysis of the existing business process with 150 Visio workflow diagrams
- Pulled all of your network diagrams for the whole company
- Cataloged all of your current servers, desktops, workstations, etc.
- Completed a 100-page return on investment analysis complete with five Excel charts
- Asked HR for an organizational chart of the department and senior management
- Had the corporate librarian research all about taxonomies and metatags and develop some suggested rules for your project
- Had Legal research all about Sarbanes-Oxley and provide you with information as to how vendor's software must comply
- Information Management supplied you with the corporate records retention schedule and policy
- Information Management supplied you with a list of the corporate records that are stored in various facilities and their status, such as, active, inactive, on hold

You want the vendors to be able to review this information so that their proposals are fine-tuned to your unique needs.

The RFP appendix is where all of this information should reside. An appendix allows you to provide voluminous information that vendors should have in order to understand your applications and needs; however, placing 150 pages of workflow diagrams in the middle of the technical requirements section is neither desirable nor needed. This information can be referred to in the text, "For an example of the Loan Approval process,

see Appendix F." Or, "Appendix G contains a copy of our current records retention guidelines including forms for approving destruction of documents."

By placing material in an appendix, you can provide the vendor with a wealth of information and therefore lessen the number of questions

received by vendors while increasing the information that they can use to write their proposals.

Evaluation Criteria

As part of the RFP development process, and in concert with writing the requirements and specifications, the RFP team is also responsible for developing the proposal evaluation criteria. Evaluating proposals must be accomplished in a fair and meaningful manner; otherwise, it will appear that the RFP was "already rigged" for a pre-selected vendor.

Evaluation of proposals encompasses many different areas, including but not limited to:

- Basic adherence and compliance with the RFP administrative requirements. For example, was the proposal submitted on time? Was the suggested vendor proposal format followed? Did the vendor acknowledge and incorporate changes to the RFP as requested? Was the proposal readable?
- Overall understanding of the RFP issues. Was it evident that the vendor understood the basic issues driving the RFP? Did the vendor's proposal respond to those issues with a logical and understandable solution?
- Technical requirements. Did the vendor respond to each of the technical requirements and were the responses adequate?
- Management requirements. Was a reasonable and acceptable project and implementation plan submitted with the proposal? Did the plan demonstrate an understanding of the RFP needs?
- Pricing. Was the pricing reasonable compared to the estimated budget and other proposals that were submitted? Was the pricing broken into the component parts as requested, or was pricing presented as a single total?

The basic steps in evaluating proposals are:

- Perform a quick review of all proposals submitted, and elimi-

nate any proposals that do not comply with basic criteria, have obvious holes in the responses to requirements, or are haphazardly and sloppily assembled.

- For the remaining proposals, read them in-depth and score them against the evaluation criteria. The objective of this second evaluation is to reduce the remaining vendors to two or three—called the short list.
- The short-listed vendors have a roughly equal chance to be selected as the contractor. The last step in the evaluation is to evaluate the remaining vendors according to their references, demonstrations, and presentations. At this point, pricing can be the determining factor between two vendors with equal evaluation scores, given that both have good references.

There are several schools of thought on how to score proposals. They range from assigning a simple “meets/does not meet” scoring to a complex scoring system using points and weights for each section. A middle-of-the-road scoring system assigns numerical values to the different RFP sections, and the point values are divided among the requirements within a section. An example is shown below:

		Vendor 1	Vendor 2	Vendor 3	Vendor 4
Requirement	Point Value				
1 Hardware Solution	200	200	175	190	180
2 Software Solution	200	150	175	200	150
3 System Design	400	350	325	375	300
4 Management Plan	200	180	175	195	140
Total	1,000	880	850	960	770
Vendor Ranking		2	3	1	4

Notice in the example that pricing is not given a point value and is not included in the overall scoring structure. Price, in this case, is used a final determination between the final two vendors, if needed.

Below is an example of the point distribution within a given section for Vendor 1:

RFP Reference	Requirement Description	Point Value	Total Points
	Project Planning	200	
3.2	Project Plan	25	20
3.2.1	Project Schedule	25	15
3.2.2	Team Resumes	25	25
3.2.3	Training Plan	25	25
3.2.4	Documentation	25	20
3.3	Site Preparation Plan	75	75
	Total		180

Evaluation may include going on a site visit to the vendor’s headquarters, visiting one or two of its reference accounts, and asking the vendor to set up a demonstration system that you can use for a period of time—generally 30 days. If you ask for a demonstration system, there should be specific activities planned that test the system based on criteria developed for the demonstration. For example, you may ask for a basic system to be installed, so you can test the indexing, searching, and retrieval of information in a specially prepared test database. The testing must be fair and tests that would go beyond the normal operation of a test system should be avoided.

As a side note, during the review of each proposal, you may find excellent nuggets of information that can be used by you whether the particular vendor wins or not. Ask each reviewer to keep track of ideas or suggested procedures that may be useful for the project. When all proposals have been reviewed and a winner selected, you may consider reviewing these nuggets of information and determining if they can be incorporated into your project. For example, a losing proposal may have an excellent point about developing and maintaining a corporate taxonomy that you would like to consider once the

project starts. There is no reason why this information should not be considered and if valid, incorporated into your project.

The final step in the evaluation process is to recommend a vendor as the winner. The recommendation report reviews why the chosen vendor

was selected and why the second place vendor was not selected. This report is given to management and purchasing, which will begin contract negotiations with the vendor.

Evaluation can become quite complex and take a considerable amount of time to complete. Evaluations are completed at the end of a potentially long project when the proposal team

members may be getting tired and frustrated because they also have their “real” jobs to do. *It is important to have the evaluation criteria completed in concert with the RFP and certainly before the proposals are returned.* Otherwise, the

evaluation will be less than objective, and personal agendas will begin to appear in the absence of a structured plan.

Post-RFP Activities

Once evaluations are completed, several steps must be completed prior to notifying the winning vendor. These steps will help you to organize and close the RFP phase of the project.

- **Evaluation Report.** An evaluation report is completed when the team has selected a winning vendor. The report reviews what vendors were considered, how they were evaluated, and why the winning vendor was selected over other vendors on the short list. This report is then delivered to senior management in the business unit, information technology, and purchasing. If needed, or desired, the RFP team may be asked to provide a presentation of the evaluation results.

The evaluation report is beneficial for a number of reasons:

- It allows the RFP team to review all events leading up to the selection and demonstrate that a fair and objective decision was made.
- By having the business unit, IT, and purchasing in agreement with the decision, you ensure their participation in the project when it is started.
- It provides the business unit with an “audit trail” of the project if there are changes within the company that necessitate a review of the project.
- If it is needed, you can defend your decision should a losing vendor file a protest.
- The report finalizes the decision and closes the RFP phase of the project.

The report should contain, but not be limited to, the following:

- Summary of why the RFP was initiated and of the project goals and objectives
- List of participants on the RFP Team and their roles
- How vendors were selected to participate
- A list of the vendors that submitted proposals and their scores
- Review of the evaluation criteria, including which were most important
- Which vendors made the short list and why they were not selected
- Review of why the winning vendor was selected

The next step, if the evaluation report is accepted, is to notify the winning vendor and schedule a meeting to begin the contract negotiations, and get the contract signed so the project can begin. If not previously done, the losing vendors should be

notified in writing that you have selected another company. As a cautionary note, it is possible that negotiations can break down and the winning vendor is deselected. Reasons for this can be varied, but they may revolve around contract issues, such as liquidated damages, ownership of custom-developed programming, and payment terms. To mitigate these types of contract issues, ensure that you include your purchasing contract with the RFP and request that vendors review it; and, as part of their proposal, point out specific contractual issues that they may have trouble with should they be selected. This allows you to prepare for these issues prior to making a final selection, but it also prevents a contractual surprise at the most inopportune time.

While you may think your part is finished and that purchasing will take over the negotiations, it is advisable for you to become part of the purchasing/negotiation team and participate in the final negotiations. As this is a “technology” purchase that is not for common off-the-shelf products, there will be some “fine points” that the purchasing person may not understand or consider during the contract review. You will be able to help and facilitate the contract negotiations and perhaps avoid a stalemate should a potential dispute arise.

Other closing activities you may consider:

- At least one copy of each of the vendor’s proposals and the evaluation criteria should be filed away. These should be kept at least until the project begins or classified according to your records retention criteria.
- Offer to “debrief” vendors that were not chosen. *Remember that these vendors, especially the ones on the short list, spent time and resources on this effort and should be given a reasonable explanation as to why they were not selected.*
- Review the proposals that lost for potential information that can be incorporated into your project.

Distributing Your RFP

Getting your RFP out and into the vendor’s hands is a crucial step in the overall process. In today’s world, publishing can be a simple hard-copy printing of the RFP or a combination of hard-copy printing, electronic printing to a CD, or placement of the RFP on a website. Hardcopy printing makes the RFP available to vendors in paper form, and the document is typically mailed. Electronic printing provides a variety of benefits and allows the document to be delivered to vendors in a variety of methods.

Electronic printing can originate with the word processor file, such as MS Word, or it can be converted to another file format, such as Adobe's Portable Document Format (PDF). The file is then placed on a CD-ROM and mailed to prospective vendors. All other communication with the vendors, and from the vendors, is generally written and mailed or emailed.

A second option is to place the RFP and all other documentation on a website. Vendors can receive email (or written) notice of the RFP and the URL for the website. If placed on the website, the RFP can be in the original file format (MS Word) or converted to a PDF file or HTML file. The website can also contain other materials such as workflow studies, equipment and facility lists, and department organization charts. Ongoing communication with the vendors can be placed on the website and may include such items as RFP addenda, questions and answers, and general communication.

The advantage of the website as a medium is that it makes it very easy to post changes and information to the RFP on the website. Vendors are responsible for periodically checking to see if there is new material and downloading it. This also eliminates the printing and copying of numerous materials, mailing costs, and delays in receipt due to postal mail.

Each of the physical file formats have advantages and disadvantages. If you use the original file format (i.e., MS Word), vendors will be able to read the file and cut and paste parts of the RFP into their proposals, which saves them time and makes their proposal easier to read. Vendors will also be able to directly use tables or spreadsheets provided in the RFP for such things as pricing or a product compliance matrix. However, having the original file allows the vendor to mistakenly change the file (their problem, not yours, but still a problem) and base their proposal on incorrect information. Original Microsoft Word files may also be problematic because of the many versions that are in use and the potential difference between different operating systems such as PCs using Microsoft or Apple Macintoshes.

A PDF file is an exact electronic copy of the original file and uses a standard reader to view the file, so there should not be any "version" problems. PDF also allows vendors to deconstruct the file and get ASCII or RTF text from the file, so they can still cut and paste parts of your RFP into their proposal.

NOTE: A common practice is for the vendor to retype a requirement and place the requirement before the proposal response. For example:

- **Requirement:** Printers shall be capable of handling a work group of 25 people who print an average of 35 pages per

hour per person. No user shall wait more than five minutes to have their job completed.

- **Response:** We are proposing two 17 page per minute printers for redundancy instead of one high-speed printer. Assuming most print requests are not serial but may be multiple simultaneous requests, all print requests should be printed within a three-minute period.

Placing the requirement above the response allows you to read the proposal faster with fewer references to the RFP document.

Once you have decided on the file format, there is a number of ways to deliver the document to the vendors:

- **CD-ROM.** The CD is mailed to vendors with a cover letter but not with a printout of the complete document. If you use this method, make sure the most important dates *are in the cover letter*.
- **Email.** It is possible to email your RFP to all vendors. However, you will have to manage a potentially larger number of email addresses and ensure they have been "tested" prior to sending the RFP. For large RFPs that are over two megabytes, email may not be the most efficient way as some recipients may have trouble with large email attachments. Depending on your RFP, you may have more than one file and may have to use a file compression product (Zip files).
- **RFP website.** Using your website to hold the RFP is becoming a common practice. It has many advantages such as you do not have to print anything; large appendices can be available for downloading; questions and answers can be posted to the site; and any RFP changes can be quickly updated by posting the change to the website. It is also possible to password-protect the site, so that only designated vendors can view your proprietary information.
- **Print and mail.** Ensure that you have a name and address of a person at the vendor's company. If you send your RFP to ACME Software, Sales Department, chances are that your RFP will float for days or weeks before it finds a home.

Typical Pitfalls

Below is a discussion of the most typical problems that befall companies when writing RFPs and selecting vendors.

- **Not enough vendor/technology education.** The ECM vendor community is large and crowded. There are many good solutions to the ECM problem, but each vendor will have a slightly different twist to its product and may have features and functions that other vendors do not have. The RFP team

should invest as much time as possible in education as it will save them time when writing requirements and evaluating proposals.

- **Poorly defined requirements.** This is typically due to two basic reasons. First, see item one above. Second, not enough time is spent understanding and documenting the internal requirements. This results in requirements that are so broadly stated that they are meaningless to a vendor. An RFP requested, for example, that the ECM support output to different formats and devices including any future formats that may be developed within the industry. This type of requirement invites the vendor to either say they can't support unknown formats (most vendors would not say this) or to say they will "...support future formats," which may not be true.
- **Poor coordination among key players.** For example, an RFP described the content developers, the content administrators, and the IT group, but never described the actual users of the system—the people who would use the system to obtain the information they needed. When vendors questioned the RFP team about the "users of the system," the RFP team could not adequately define who a user was, what a user would do on the system, how many users there were, how many hits were expected, what the average length of time spent on the site would be, etc. In their haste to completely define the "solution," the RFP team forgot their audience. One RFP team member even ventured that since there was not a current system in place, she could not determine who would use the system and how.
- **Providing requirements that differ from one section to another.** A common mistake is not checking for consistency within a section or between sections. One requirement in the technical section may state that there are 250 total users for the system, but in the management section, the total number of users is 325. Another example may be that the system must support metatags, and another section says that metatags will not be used. This type of problem will be caught by vendors, and they will ask you to clarify the issues. These types of questions require you to research the problem, write a response, and potentially cause other requirements to be changed—which may delay the overall effort.
- **Insufficient Budget.** Many RFP teams are "surprised" when they see vendors' pricing for their RFP. The RFP team may have estimated a budget based on old or inaccurate data, or they completely misjudged some aspect of the intended sys-

tem. For example, one of the most costly areas of today's ECM projects is the time needed to develop (i.e., program, code, or customize) the application, and yet, this is the most misunderstood area of the project. How can the RFP team estimate whether the programming will take one man month or six man months? Some RFP projects are undertaken without a budget and when proposal pricing is reviewed, the project is simply cancelled because the project costs are much more than senior management will allow.

- **Not enough time allocated for the RFP.** Many RFPs suffer because there is not enough time provided for the overall RFP writing effort—from initial research and analysis to system implementation. The amount of time typically needed for an RFP is underestimated due to inexperience with the RFP process. The result of having less time than expected is that evaluations are truncated, contract negotiations are sped through, project development is hurried, and the project in general starts to have more than the expected number of problems, which slows things down.

Given ambiguous or unclear information in the RFP, most vendors will simply agree with the RFP, and if an issue surfaces later in the project, they will plead innocent. This is in the spirit of, "Better to beg forgiveness than to ask permission" because once a vendor has been selected, and the project is underway, it is very difficult to change vendors.

These issues also increase the risk of having a problem contract or no contract. If serious enough, these issues lead to vendors not bidding on the project, or proposals that are overly conservative as vendors bump up their pricing to cover any contingencies.

Conclusion

An RFP is a document that defines a business problem or need and provides enough information to allow potential vendors to propose a solution and a price. A well-written RFP promotes competition among vendors and encourages them to provide unique solutions based on their products, services, and interpretation of the RFP requirements. Proposals received allow the buyer to select the best solution at the most competitive price. Together, the RFP and winning proposal become the foundation for the contract, which defines the solution, the project implementation tasks, and establishes performance goals.

RFP requirements should not be too tightly constrained because fewer vendors will be able to bid on the project. Because the ECM vendor community is large and diverse, the RFP team should be open to a variety of solutions (within reason) and teaming arrangements. Writing an RFP helps the buyer to understand their unique need in a level of detail that would not otherwise be available. This level of detail ensures that, as much as possible, the RFP represents a fair and objective assessment of the problem. Writing an RFP also provides the following benefits:

- Gives the “project” definition
- Establishes a real need for the system
- Defines basic and specific technical requirements
- Establishes a budget
- Establishes project timelines and schedules
- Consolidates and scrubs requirements
- Provides early identification of potential problems
- Identifies appropriate vendors
- Identifies the correct technologies
- Establishes project goals, objectives, and critical success factors
- Promotes competition among vendors
- Allows you to choose a solution that provides the best value

The RFP process helps you make the right decision by minimizing the risk of selecting a vendor and solution that is not right for your application.

'Note: The Federal Acquisition Regulations (FAR) mandate an RFP format for federal government projects. Federal Government readers of this guide must format their RFPs according to the FAR, but can still use the information contained in this guide for developing the content of each section.

Bud Porter-Roth, Porter-Roth Associates

Mr. Porter-Roth is the founder and principal consultant for Porter-Roth Associates (PRA). He is a recognized expert in the RFP field and consults with business and government on RFP strategies and tactics and has managed and written many RFPs. PRA specializes in RFP development and associated tasks such as:

- Project organization
- Feasibility studies
- Requirements definition and analysis
- RFP and proposal development
- RFP and proposal evaluation
- Vendor selection
- Project implementation, management, and oversight

Mr. Porter-Roth has been an independent consultant in the technology industry for over 12 years. Prior to forming PRA, he worked as a Senior Consultant for TRW Financial Systems (TFS), Director of Proposal Development for Plexus Computers, and Manager of Proposal Development for Tandem Computers.

He is a frequent speaker and writer for associations and publications on the topics of writing RFPs, procuring technology systems, and writing proposals. He is the author of the following books: *Request for Proposal: A Guide to Effective RFP Development*, *RFP Guidelines (AIIM TR-27)*, and *Proposal Development: How to Respond & Win the Bid*.

Mr. Porter-Roth received a BA and teaching credential from U.C. Berkeley. He is an AIIM Master of Information Technologies (MIT) and a Certified Document Imaging Architect (CDIA). He has chaired various standards committees at AIIM including the RFP Guidelines (TR-27) committee.

Contact him at budpr@erms.com



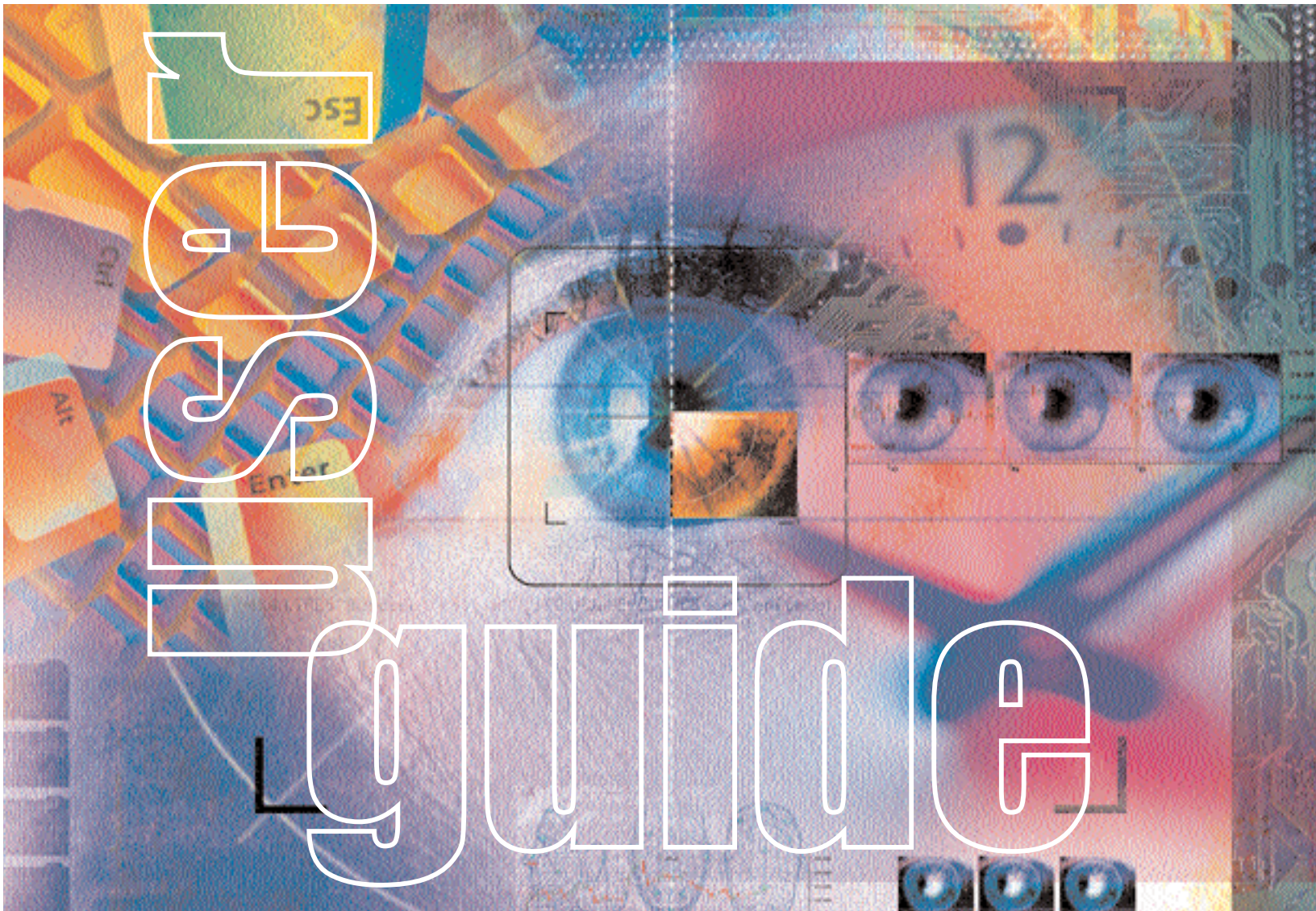
AIIM International

1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
800-477-2446 / 301-587-8202
aiim@aiim.org
www.aiim.org

For over 60 years, AIIM has been the leading international organization focused on helping users understand the challenges associated with managing documents, content, and business processes. Today, AIIM is the leading international authority on Enterprise Content Management (ECM). ECM is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and technologies provide solutions to help users with the four C's of business: CONTINUITY, COLLABORATION, regulatory COMPLIANCE, and reduced COSTS.

AIIM provides:

- **Market Education**—AIIM provides unbiased information through *AIIM E-DOC Magazine* and *mID* (Managing Information and Documents), the leading industry magazines in, respectively, North America and the UK; its 20-city Content Management Solutions Seminar in the U.S. and Canada; the IM Expo event held throughout the UK, and Inf Ireland.
- **Professional Development**—This industry education roadmap provides a variety of opportunities. IM University is a multi-faceted program offered in Europe. The Web-based Fundamentals of ECM Certificate Program familiarizes users with the core concepts and technologies related to ECM. The AIIM Webinars round out user education on key issues.
- **Peer Networking**—Through chapters, networking groups, programs, partnerships, and the Web, AIIM creates opportunities that allow, users, suppliers, consultants, and the channel to engage and connect with one another.
- **Industry Advocacy**—AIIM, as an ANSI (American National Standards Institute)-accredited standards development organization, acts as the voice of the ECM industry in key standards organizations, with the media, and with government decision-makers.



ESG
guida

Authored by Arthur Gingrande, Partner, IMERGE Consulting

Processing Unstructured Documents: Challenges and Solutions

*An AIIM User Guide
By Arthur Gingrande*

*This is one in a series of User Guides from AIIM International.
They are intended to educate and inform readers on a variety of
enterprise content management topics.*

Copyright © 2004 by:
AIIM International
1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
301-587-8202 / 800-477-2446
ISBN 0-89258-395-9

Published by:



www.aiim.org

Processing Unstructured Documents: Challenges and Solutions

The majority of business documents (e.g., invoices, purchase orders, resumes, work orders) are arbitrarily structured and, as so, cannot be processed using conventional “ICR-friendly” document processing approaches. Alternatively, they may be consistently formatted but contain incremental variations from one document to another (e.g., medical claims, IRS forms) that diminish the effectiveness of using customary, ICR-based forms processing applications to the point where they hit unacceptable cost-justification levels. Over the past few years however, advances in automated forms processing technology have dramatically improved recognition accuracy in processing arbitrarily structured business forms. These advances include exponential increases in computing speed and memory, significant improvements in image processing technology, and innovations in neural network algorithms.

Clearly, ICR-based forms processing has come a long way since its introduction to the imaging world in the late eighties. In those early days of automated forms processing, automatically identifying a given form type and then setting it up for ICR/OCR taxed the capabilities of users, integrators, and vendors alike. Typically, the only forms eligible for image-based forms automation were those that had been created by the company responsible for processing the form. In fact, the forms in the early days were designed explicitly to be “ICR-friendly;” that is, they were specifically formatted to fit the needs of intelligent character recognition software, in order to obtain the most accurate text recognition and hand-printed recognition results possible.

This meant that, preferably, the form data were printed in “drop-out” ink—carbonless ink designed to be ignored by an imaging camera—so that only the “active data” filled in by the customer was actually detected by the scanner. Furthermore, since ICR engines encounter extreme difficulty when forced to recognize connected characters, the form data fields had to be framed by graphical objects known as “combs”—strings of

boxes that forced a person to separate hand-printed characters when filling out the form.

The user created a software template to match each form type and define the ICR parameters of each of the fields on the form—check box, hand print, machine print, alpha, numeric, number of characters, and so forth. Due to the limitations of form identification technology, forms were processed homogeneously by the batch, carefully sorted and tightly organized by form type: one batch, one form type, so that the form template precisely fit each and every form in the batch. If the layout of any form in the batch differed incrementally from that of the form definition template, the system rejected the form and sent it to a human for manual data entry.

Over the years, form identification and data location algorithms have improved considerably. Setting up a form type—or a number of form types—for automated recognition has become a relatively simple task for users to accomplish with about any given forms processing software system. An end-user can establish form ID and data field parameters in a few simple steps, using a form setup module that creates a mouse-driven, form definition template with the aid of a user-friendly graphical user interface (GUI). Algorithmic software innovations allow ICR and OCR engines to recognize data elements even when the target data falls outside the zones of the form definition template. Taken collectively, these innovations add up to higher recognition accuracy. For all practical purposes, the problem of processing structured data on forms is solved.

Nowadays, however, it is unstructured data that is getting all of the attention. *Unstructured data accounts for nearly 80% of all of the corporate data on record.* This percentage applies across the board to all data types, including email and voice messages, slide presentations, videos, attachments, paper documents, and paper forms. The overwhelming volume of the information involved makes processing unstructured data for storage and consumption a major priority for most corporations. This preoccupation with unstructured documents prevails currently in the world of paper forms, where the hottest document management application is the processing of unstructured forms such as purchase orders, invoices, medical claims, and explanation of benefits (EOB) forms.

Chapter One:

Defining the “Unstructured” Document

The term “unstructured” document, of course, is a misnomer when applied to forms. After all, the essence of a form is structure. One could go so far as to say that, in reality, a form is nothing but structure devoid of content—in effect, a form is an empty vessel waiting to be filled with data. Moreover, there is no question that, purchase orders and invoices, for example, are consistently structured from one instance to another—at least from the standpoint of the issuing company. The term “unstructured” becomes meaningful if and only if it is interpreted from the perspective of the company on the receiving end that must process the myriad of purchase orders, invoices, shipping documents, medical claims, or explanation of benefits (EOB) forms that arrive daily at its headquarters in a variety of layouts that differ arbitrarily from one sender to another.

These random differences over a series of functionally similar forms mean that the forms cannot be processed using the traditional template-based approach, in which one software template matches each and every data field on all the forms in a presorted batch. Instead, the data on these forms must be entered manually by human operators, and at considerable cost.

If the operation is automated, then it must be capable of processing unstructured forms “on the fly,” using sophisticated algorithms that locate the data on a diverse array of form layouts within a given form type. Forms that originate from a continually changing number of sources, and that arrive in arbitrary sizes and formats, require a system that employs relatively little explicit, programmatic knowledge about the formats of incoming documents. In other words, accurately automating the processing of unstructured documents is successful under conditions where the forms automation software is indifferent to the targeted form data. Jerry Fisher, Vice President of Symbus Technology at the time, once remarked that, “forms processing will penetrate the mainstream market only when the software that drives it is as indifferent to the form data as the form itself is.”

Accordingly, the following definition represents the consensus of the major forms processing vendors, based upon interviews with them conducted by IMERGE Consulting. It will govern the use of the term “unstructured” document in this paper. An unstructured document is a class of forms in which:

- The individual forms are not designed by the processing agent;

- The forms are received in a variety of layouts that can differ from sender to sender; and
- The forms cannot be processed using the traditional form template that consistently matches up with each and every data field on all the form documents in a presorted batch.

In other words, *the term “unstructured” document refers to a group of functionally alike forms with dissimilar layouts, which are received by the processing agent in collectively high volume and that would normally require manual key entry in order to capture their data.* In reality, a better term, slowly coming into use, would be “semi-structured” forms.

Major applications

Unstructured document processing technology is just starting to vie for mainstream user adoption. In a survey conducted by IMERGE Consulting, vendors reported that the most popular applications include the following, listed in the order of consumer demand:

- **Invoices and purchase orders**—Every major company processes mountains of invoices every day, not to mention their sister form, the purchase order. Their arbitrary format makes it necessary to enter the data contained manually. Nowadays, every major forms processing vendor offers an invoice solution that can also do purchase orders.
- **Medical claims**—Data on an HCFA-1500 medical claim form is contained within 31 numbered regions, which in turn contain over 100 inconsistently placed, densely packed data fields within them, which can include machine print, mark sense, hand print, bar code, columnar, signature, and, in some instances, bar code symbols. HCFAs, UB-82s and UB-92s, and dental claim forms are composed data fields. Each sports its own peculiar array of lines, boxes, and fine print instructions—printed in red, green, or black—and each color presents its own set of imaging problems. Add in the OCR complications of degraded font recognition created by claims filled out by faded ribbons from old dot matrix impact printers chronically used by doctors, and processing medical claims becomes the most challenging task in forms automation.
- **Explanation of Benefits (EOB) forms**—Derived from a medical claim, an EOB statement is no ordinary form. Essentially, it is one big table with multiple columns of data that frequently spans many pages. The number of data items within each column is indefinite, and vertical positions on a page

are unpredictable. There may also be intrusions of summary data items sandwiched between consecutive patient accounts. Moreover, within a single column, several different data items may or may not be staggered vertically. The overabundance of variable-width columns and tightly-packed data fields, often spread over 20 pages, makes EOBs particularly difficult to recognize by conventional recognition systems.

- **Transportation documents**—These include shipping documents such as bills of lading and customs declarations. Like invoices, these forms are processed by parties that did not create them, so their layouts appear arbitrary to the processor. Moreover, the use of carbon copies and the beating that these documents take en route to their destinations compound the recognition problems involved.
- **Tax forms**—The tax form is one of the most difficult forms to recognize by machine, for many reasons, including the complexity and density of the form, its myriad parts, its multiple-page nature, and the variance in possible layouts, especially since many of the state IRS departments accept forms from accounting firms and tax preparation software packages that do not match the government-issued form. The sizes of the data fields on an ordinary tax form do not encourage the taxpayer to separate characters legibly from one another. Moreover, since the data fields can be filled out in either machine-print or hand-print, the ICR engine must be able to switch from one recognition mode to another ad hoc. Statistics are unavailable on how many invoices are processed by American businesses each year. Since invoices are the documents that demand payment—and hence are the form-based lifeblood of America’s capitalistic economy—it is intuitively obvious that they are the most prevalent and mission-critical of all unstructured form types. At last count, there were over 75 million pages of medical claim forms processed daily, at a cost that exceeds \$125 million per day—and there is at least one EOB form for every medical claim. Transportation documents proliferate in obvious abundance, but the degraded conditions of many of them, because they are attached to the outside of packages and are often carbon copies of originals, create additional problems that make them the province of specialist integrators.

Tax forms processing projects, because they involve state and federal government documents, are usually done on a special bid basis, and require massive amounts of tailored validation routines and code-writing by developers and systems inte-

grators. Over 30 states already have systems in place, and the federal government has a number of programs dedicated to automating tax processing and auditing procedures. Consequently, like transportation document processing, tax forms processing is an application for which a software solution is not in very high demand by corporate America—especially when compared to invoice processing, claims processing, and EOB processing.

Chapter Two: Properties and Procedures

It would be incorrect to treat all unstructured forms as one document class with a set of variable but definable characteristics. Clearly, an EOB differs markedly in appearance from an invoice or purchase order. Each form type must be processed according to its own set of business rules. Each type of unstructured form has idiosyncratic characteristics that present a unique set of problems to the I.T. professional, which, taken together, require a customized approach to process them successfully. In fact, each application has become so customized that the current practice among form processing vendors is to produce a separate software product for each major unstructured forms processing application.

Hence, in today's marketplace, the typical publisher of forms processing software makes available to its customers one software product for invoice processing, another product for medical claims, and still another for EOBs. Alternatively, a vendor can offer a generic "any form" processing engine that accepts plug-in modules that are tailored on an application-specific basis to process invoices, claims, or EOBs, accordingly.

Because these are the three highest-volume applications, each one lays claim to its own product. As time goes by, no doubt, vendors will publish a software package that features a transportation forms processing wizard, but currently there is none.

Factors to consider when processing unstructured documents include data location variability, data field ambiguity, data field density, form removal, multiple pages, knowledge of the business process, and foreknowledge about forms.

• **Data field location variability**

As previously discussed, due to the variety of graphical layouts and data formats, the major difficulty in processing unstructured documents involves simply locating the array of suspect fields that must be captured in order to yield the specific data required to process a given form type. There are a variety of ways to locate suspect data fields on forms processed "on the fly," which will be described later in this paper. For example, in order to locate a sum at the end of a column, brute force OCR can be used to recognize every field in a certain region of a form, and then the total can be discovered by looking for a "total" or "sum" field. Alternatively, the morphology of the form itself can be examined in order to detect the last entry in a columnar "blob" that is then passed on to an OCR engine to be recognized.

• **Data field ambiguity**

Due to the imperfections of the real world, objects on forms, such as invoices, can appear as many things, open to a variety of differing interpretations. For example, not every amount on an invoice form is the amount that the customer must pay. There may be the original billing amount, the net billing amount "if paid within thirty days," an amount that is a subtotal of the final total, and the amount before sales taxes, just to name a few. The same goes for dates. There is the "pay by" date, the "date sent," and the "date received" on the invoice. The latter field could be stamped on the invoice in ink and, if so, might have to be filled in by hand. Consequently, the system must employ intelligent algorithms to figure out automatically what each individual amount and each date means. The same is true for other unstructured documents, such as purchase orders and medical claims.

• **Data field density**

Data field density can be a problem, especially if the form is cluttered with a lot of "noise" that partially fills in white spaces between the fields and makes them difficult to extract by OCR engines. In particular, healthcare forms, such as HCFA, UBs, EOBs, and dental claims, are extremely packed with data fields of all types: text, hand-print, check boxes, bar codes, and signature fields. On an HCFA form, there can be over 100 fields on a page, squeezed into 31 numbered regions. State-of-the-art form removal, special data parsing rules and scripts, as well as accurate OCR and ICR, are essential for extracting information from tightly packed data fields.

• **Form removal**

Invoices and purchase orders typically contain fewer graphical objects, lines, or boxes on them than do medical claim forms. Separating the "passive" healthcare form itself from the "active" data supplied by the medical personnel filling out the form is an extremely complex task, requiring sophisticated forms removal technology. This is primarily because the typical healthcare form is composed of densely-packed data fields, each with its own peculiar array of lines, boxes, and fine print that instructs users on what information to enter in those data fields. White space on a healthcare form such as an HCFA is a scarce commodity, which makes the process of extracting data fields from their prison of passive form elements a far more exacting task than with other types of forms, such as invoices and transportation documents. In

order to facilitate the degree of precision that is required to separate an HCFA form from its medical data elements, the form removal process must be tightly integrated with the form ID and registration process.

- **Multiple pages**

Extracting data from multiple page documents pose significant location and classification problems, due to the variable number of item lines that can be contained on a given page. One invoice, for instance, might be a single-page document, while another one could span many pages. Simply finding the grand total on an invoice or claim form can be a challenge that requires complex and sophisticated algorithms that formulate hypotheses about data content, iteratively test and modify them based upon results, then rerun the new hypotheses through the entire procedure in a continuous feedback loop. Because of the parallel processing capabilities and enormous power available from contemporary computing technology, these routines can hone and refine recognition accuracy tremendously.

- **Knowledge of the business process**

To ensure accuracy, a thorough knowledge of the business process of the user, as well as the invoice forms themselves, is essential. For example, with invoices, the vendor number typically is not on the invoice itself, but is assigned internally and kept in a separate database. After OCR, the vendor is identified automatically by matching the vendor name to a number in its own database. Within the business process, security considerations also enter the picture: business rules could require that high-dollar invoices be flagged and routed for manual review before they are approved for payment. In the case of medical claims, they must go through an adjudication process and be checked for fraud and illegal “bundling” procedures before reimbursement can be authorized and paid.

- **Foreknowledge about forms**

Because a static template cannot be used to recognize a functionally defined class of forms on the fly, prior knowledge about forms takes on extreme importance. In fact, prior

user knowledge about forms is essential to ensuring document classification and character recognition accuracy. Such knowledge could include, for example, the relationship between data and keywords, tabular structures and isolated fields, and information about data formats.

The more advanced knowledge about the forms that can be employed, the better. An invoice, for example, contains an invoice number and purchase order number, “bill to” and “ship to” address information, and a total, as well as a variable number of item data lines with product codes and/or descriptions, quantity, unit cost, etc. “Stuffing” the forms processing engine with this prior knowledge by integrating it with an existing purchase order database enables the engine to verify vendor information and compare purchase order amounts to invoiced amounts. This process narrows the range of the search and accelerates validation of recognized invoice data, which speeds up the data capture process significantly.

Once trained, an invoice processing system can be set up to learn about user documents as it goes along. In fact, some systems can be systematically retrained to take into account the elimination of old vendor’s invoices as well as the addition of new invoices from new vendors as these changes modify the processing set. Each system learns more each time it processes an invoice. It recognizes invoices from the same vendor and remembers the location of the required data, which improves accuracy and speeds up processing time with each successive invoice. Other software systems use neural network technologies to automatically learn invoices on the fly and store the results as a series of quasi-dynamic templates.

Processing unstructured forms involves taking multiple approaches that create alternatives that are compliant with classifying consistently similar data elements. It takes a powerful and sophisticated combination of document classification, image parsing, and character recognition algorithms to sort out the results correctly.

Chapter Three: **Enabling Technologies**

Processing unstructured documents is nothing new to the field of forms automation. Techniques for processing “forms on the fly” were pioneered in mid-1990 by Nestor, Symbus (renamed Captiva), Mitek, Daimler-Benz (changed to OCE), MTI, and the now-departed TeraForm. Most of these companies typically used neural networks to analyze and generalize the image morphology of a particular class of unstructured forms such as faxes, invoices or medical claims. Advanced feature extraction techniques helped to facilitate page decomposition, and complex search routines located specific data fields on variable and problematic forms. The software could even find form data fields even if they were discovered separated from their expected locations by as much as a half-inch. However, due to the complexity of the algorithms and the expensive, pre-Pentium hardware requirements, attaining acceptable processing speeds at reasonable accuracy levels was costly; hence, during the 1990s, the software failed to achieve mainstream user adoption.

Today, however, given the incredible computing horsepower and memory that resides in the average desktop PC, a variety of techniques, including ICR brute force, can be run in parallel to achieve remarkable results. Blob analysis, edge detection, multi-line character segmentation, and long-line detection can be used to find form objects, columns, and data fields based solely on their topology. Sometimes the geometrical and spatial relationships between the text data elements, such as rows or subheadings (rather than graphical objects), are used to locate the places where data most likely will be found. In fact, the process need not involve character recognition at all; the text can be treated as a pattern of blobs.

Conceptually, there are two basic approaches to processing unstructured forms, located at opposite ends of the methodological spectrum. The document image understanding (DIU) approach treats the form primarily as a pattern of specific images, whereas the character recognition-driven approach treats the form primarily as a concatenation of text data elements. Neither technique is used solely by itself; contemporary vendors each use a hybrid version that uses techniques from both schools of thought.

The document image understanding (DIU) approach

In the neural network-driven approach, the software engine analyzes, learns, and then generalizes the image morphology of

a particular class of unstructured forms, such as invoices. Neural network technology is used to train on a large training set of images consisting of hundreds, even thousands, of different samples of the same form type, in conjunction with feature extraction technology that distills and extracts a wide variety of topological elements—text lines, columns, graphical objects, and pictures—from the training samples. When linked to a conventional forms processing system, the DIU software engine facilitates page decomposition, which dissects and then automatically locates specific data fields on variable and problematic forms. The DIU engine also locates variable numbers of fields across variable numbers of columns, rows, or subheadings.

In the dynamic template-driven approach, the user defines a dynamic template or a map that describes the data zones contained on the form type in question. The DIU engine uses techniques such as blob analysis, edge detection, multi-line character segmentation and long-line detection to locate form objects and data fields. Defined characteristics could include the width of one data field relative to the width of another, the number of characters in a field, the size of an expected logo relative to the data columns on the page, and other factors. Fields can be detected at the pixel level with extreme precision, which allows for extremely fine separation of wanted and unwanted image elements, especially when locating tightly-packed data fields, such as EOBs. Once the data fields are located, they can then be sent to an ICR/OCR engine to be recognized and then validated.

Utilizing an intelligent scripting language approach, the end-user can employ a DIU engine to write a set of image parsing rules in a scripting language that guides the DIU engine in its search for data fields. The script is essentially run as a Visual Basic program that instructs the recognition engine on how to locate the required data. The script actually reads the image, using predefined information about intrinsic relationships—for example, the location of a “total” field relative to columnar lists of item values, geometrical location of data elements, such as corporate logos and address blocks, augmented by text prompting.

Another form of DIU involves focusing more on the spatial relationships between the text data elements rather than graphics in order to locate the places where invoice data is most likely to be found. The process of locating clues based upon text need not involve character recognition: the text can be perceived simply as a pattern of blobs. However, once the text is found, an ICR or OCR engine is used to recognize and validate the found data.

Processing does not stop once the data has been accurately recognized. Complex algorithms are used to sort out the recognized data and make sure that it fits the criteria for a given form type. Totals must be balanced, dates must be checked, and application-specific business rules must be followed in order to determine that the right sums and payment dates are reported.

The character recognition-driven approach

The character recognition-driven approach relies upon a powerful, OCR-based, literal search technique to locate and extract data from pre-designated regions on a form, and then processes that data as part of a user-defined routine. Instead of a special scripting language, users can set up routines using a special dialog box to edit, graphically map, and then view the procedures. Searches can be set up according to predefined data definition parameters, often exhibited as a flowchart, then the compiled results can be displayed in a number of graphic formats. Rules for accepting the data as valid can be established contingent on the confidence values of the OCR results.

The user can replace a defined character with another character, remove questionable flags, hide rows, or flag rows based on whether or not a character is found or not found. The user can also hide or flag rows, remove tables, and/or execute another procedure contingent on the results of the first procedure.

Conditions for launching additional procedures are based upon whether or not the criteria set for a primary or secondary search are met. This can involve applying specific business rules to the information found on invoices, (e.g., the length of a purchase order number must be greater than nine numeric characters, must begin with the characters "P1," and that payment terms must be net 30 days, etc.).

Almost all vendors support their form processing engines by

using a dynamic template library to speed up form identification and data element location. Using this method, the user designs a series of generalized form templates to help recognize the invoices received by classes of customers, grouped both by company and invoice format. These differ from the conventional, static, form definition templates in that they do not define data element regions by exact pixel location; rather, data location can be defined by general regions of the form, and by rules to follow once certain graphical structures or text elements are discovered at those designated regions of the form.

When a document is scanned, it is first checked to see if it matches a similar form type in the predefined dynamic template library. If it does, then the template can be applied immediately to find the data without a further search. If it does not, then the regular invoice analysis process is launched; but once the invoice is processed, the variations between that invoice and the nearest-matching template are noted and stored either as a new or variant template by the system for later use.

As mentioned earlier, none of these forms processing techniques is radically new. Vendors have been using them since the early nineties. Every vendor uses its own blend of techniques to optimize document processing accuracy. Many employ a multi-stage approach: they use one technique to recognize the form, and then another to locate the fields and their values before passing them to an ICR engine to recognize. Some vendors employ a neural network-based, DIU approach for doing form identification, and then apply a powerful, OCR-driven, literal search technique to find data elements where their neural network method leaves off. Others utilize still a third technique to validate found data values. There are also vendors that employ different methods simultaneously throughout the various data capture phases, in a voting scheme.

Chapter Four:

Completing the Processing Cycle

• Adding workflow to the process

Every unstructured forms processing application has a workflow component, necessary to successfully complete that application. For example, before payment for an item can be authorized in an invoice processing operation, it must be verified against a purchase order and/or other documents that describe the item, its price, and that the item has been received and accepted by the company that placed the order. As previously referred to, with respect to medical claims and EOBs, those forms must go through an adjudication process and be checked for fraud and illegal “bundling” procedures before payments can be authorized. Moreover, in the case of processing Medicare and Medicaid reimbursement forms, they must meet certain deadlines and turnaround standards, and might additionally be required to interface with a state EDI network to receive speedy reimbursement.

Setting up the correct workflow for an application demands a thorough knowledge of its underlying business processes, an understanding that is deep enough to enable the user to define explicitly all steps of the workflow and the exception procedures. Along these lines, there are two approaches for adding workflow to document processing procedures: (1) the vendor can supply its own workflow engine fully integrated with the forms processing software; or (2) the vendor can supply an application programming interface (API) for integrating its software with a number of existing workflow engines. Each workflow and business process for each unstructured forms processing application requires its own work process analysis.

In this regard, we raise a word of caution. Whether it involves training a neural network, setting up a series of dynamic form templates, predefining ICR thresholds, or integrating a forms processing system with a workflow engine to create a “total solution,” none of these tasks should be taken lightly. Despite the ease of a given user interface, no unstructured forms processing or workflow solution works smoothly “out of the box.” Setting up an efficiently running system for processing unstructured forms is at best a complicated and rather trying procedure that requires the assistance of a trained professional—either a consultant in process analysis with an ICR background, or a systems integrator that knows forms processing.

• Measuring accuracy

Interviews with industry vendors report that, when it comes to automatically processing invoices and purchase orders, per-document recognition accuracy can be achieved reliably and consistently at rates greater than 60%. This means that, out of every 1,000 invoices processed, today’s automated invoice processing software can be counted upon to classify, validate, and insert at least 600 of the documents into a company’s business process at a zero error rate with no human intervention required. The remaining invoices are rejected by the processing engine and passed to human operators to review and correct, because recognition results failed to meet user-defined parameters based upon preset ICR confidence levels.

Vendors report similar results with EOB forms and medical claims. The result is labor savings that can translate into at least a 60% reduction in overall data entry costs. For now, however, these numbers are purely anecdotal, for no independent, third-party, industry benchmark tests have been conducted to establish industry-wide performance standards.

The fact of the matter is that the specific mix of documents that a company receives in its incoming mailbox creates a document management profile that is a product of the type of business in which a company is involved and the companies with which it does business. Moreover, since America is in the early stages of user adoption regarding automated forms processing software, it is too early to know what generalities will hold true across the board with respect to estimating the kind of success that a given company can expect from automating, say, its invoice payments operation.

This brings up an important point: virtually every forms processing installation requires some customization, regardless of whether the forms are structured or unstructured. As previously mentioned, there is no such thing as a strictly “out of the box” solution—that is, if maximum accuracy is desired. Each forms processing installation has a unique profile—based upon variations in skew, print shop preferences, color of ink, paper reflectivity, and font degradation and other physical and design idiosyncrasies—that requires analysis and experimentation in order to optimize recognition accuracy. In other words, the complexity and variety of these factors will determine the extent to which the application must be tailored. Therefore, the best practice for a business user to follow when evaluating the potential effectiveness of an invoice

processing, claims processing, or EOB processing software solution is to request that a vendor run a pilot project using the client's own business forms. That way, the translation of laboratory theory into practical reality can be empirically tested and observed at low risk to the potential customer. It is important for the user to ensure that the pilot is, in reality, upwardly scaleable. The pilot must involve a random assortment of several thousand forms in order to enable the user to tune the forms processing system to a document profile that is upwardly scalable and truly representative of the installation. A pilot that does not accurately portray the variety of documents that a user processes daily is useless.

A vendor that wants the business will comply with a request for a pilot because they know that this is the price that they must pay in order to ensure successful early adoption.

However, a potential buyer should be prepared to define for the vendor the terms of success that, if achieved, will trigger their decision to purchase.

- **Performance Improvement Drivers**

Since 1996, the forms processing industry has been governed or driven as much by Moore's Law and advances in computer

technology as it has been by innovations in ICR/OCR technology or in new and sophisticated form classification algorithms. Because under Moore's Law, computing power exponentially grows by doubling every eighteen months, the enormous power available in today's PCs makes it possible for brute force to be introduced as effectively as elegance into the equation. Vastly increased computing power also enables forms processing software engines to run several different or orthogonal algorithms in parallel with each other, and even to feed the results into voting algorithms. As computing power increases, unstructured document processing can only continue to improve.

In the past, applications rooted in ICR-based, forms processing and image data capture technologies failed to meet great expectations of user adoption—largely because of the applications' narrow scope of use and perceived accuracy limitations. Now, armed with market-broadening capabilities of processing unstructured documents, mainstream application adoption is a far more likely possibility. Perhaps, at long last, a preponderance of end users will finally realize the promise of these remarkable technologies.

Major Forms Processing Applications

<i>Market Segment</i>	<i>Form</i>	<i>Structured*</i>	<i>Unstructured*</i>
Healthcare	Medical Claims	X	X
	Explanation of Benefits (EOB)		X
	UBs (Series 80 & 90)	X	X
	Pharmaceutical Tests	X	
Banking & Financial Services	Mortgage Applications	X	
	Loan Applications	X	
	Mutual Fund Shareholder Proxies	X	
	Checks	X	X
Transportation & Utilities	Bills of Lading		X
	Custom Declarations		X
	Utility Bills (Remittance)	X	
Retail	Order Entry	X	
	Warranty Cards	X	
	Invoices		X
Insurance	Policy Applications	X	
	Policy Claims	X	X
Surveys and Questionnaires	Sweepstakes	X	
	Market surveys	X	
	Educational Tests	X	
Government	Internal Revenue	X	X
	Expense Requests	X	
	Court Documents		X
	Licenses and Permits	X	
	Violation Notices	X	

An "X" in either of these columns indicates if that document application is predominantly structured or predominantly unstructured. In a few instances, there are significant proportions of each.

Arthur Gingrande, Partner, IMERGE Consulting

Arthur Gingrande is a nationally acclaimed expert and pioneer in image-based intelligent character recognition (ICR), electronic forms, and forms automation. In 1988, he founded a neural-network development firm called Symbus Technology, now known as Captiva (San Diego), which has grown to become America's largest ICR-based, forms automation software firm. He is also the former director of marketing and business development for Nestor (Providence, RI), one of America's most prominent neural network software development companies.

Since 1991, over 200 of his articles have been published in various trade periodicals such as *KM World*, *AIIM E-DOC Magazine*, *Business Solutions*, *Integrated Solutions*, *Imaging Business*, *inform*, *Imaging and Document Solutions*, *VAR Business*, and *Imaging World*. The topics of these articles have included workflow, document management, electronic imaging, COLD/ERM, ICR/OCR, e-forms, e-business, CRM, knowledge management, wireless communications, processing unstructured form data, and content management. He has also written numerous white papers on those topics, including research reports for Dataquest and BIS CAP, now known as GIGA Information Group. Recent white paper clients include IBM, Tower Technology, SER Macrosoft, AIIM International, TAWPI, CharacTell, and Ceresoft. He has also written four patents in the areas of intelligent check recognition, ATM-based direct payment systems, and smart cards.

Mr. Gingrande is a founding publisher and former director of document and image management at ISIT.com, a website dedicated to integrated solutions in information technology. He is also editor and publisher of *Contemplor*, a newsletter

dedicated to document management and forms automation technology. He has participated in scores of industry trade shows and business conferences as a coordinator, guest speaker, panelist, and industry commentator.

Mr. Gingrande is the author of *Forms Automation—from ICR to Electronic Forms to the Internet*, published by AIIM International, a book about the role of forms automation in document management and electronic commerce. He wrote *Cost Justifying an ICR Solution*, published by The Association for Work Process Improvement (TAWPI). He also wrote *Technology Convergence, Document Management, and E-Commerce*, published by AIIM International, which shows the impact that the convergence of digital technologies—including the Internet, Web TV, wireless communication devices, EDI, and smart cards—will have on the application environment of the future. He also authored the AIIM white paper entitled *Enterprise Application Integration: Connecting the New Application Frontier with the Old*. Other publications include two white papers written for IBM on the topics of customer relationship management (CRM) and Web catalog management to introduce IBM's new WebSphere products in those areas.

As a partner of IMERGE Consulting in Arlington, MA, Mr. Gingrande has written the marketing or business plans for six of the leading software development firms in image capture and automated forms processing, which collectively make up 80% of the market. For IMERGE, he also consults to end-users in the areas of needs analysis and implementation oversight of automated document processing systems. Readers may contact him at 781-258-8181 or by email at arthur@imergeconsult.com.

AIIM International

1100 Wayne Avenue, Suite 1100
Silver Spring, MD 20910 USA
800-477-2446 / 301-587-8202
aiim@aiim.org
www.aiim.org



For over 60 years, AIIM has been the leading international organization focused on helping users understand the challenges associated with managing documents, content, and business processes. Today, AIIM is the leading international authority on Enterprise Content Management (ECM). ECM is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and technologies provide solutions to help users with the four C's of business: CONTINUITY, COLLABORATION, regulatory COMPLIANCE, and reduced COSTS.

AIIM provides:

- **Market Education**—AIIM provides unbiased information through *AIIM E-DOC Magazine* and *mID* (Managing Information and Documents), the leading industry magazines in, respectively, North America and the UK; its 20-city Content Management Solutions Seminar in the U.S. and Canada; the IM Expo event held throughout the UK, and InfoIreland.
- **Professional Development**—This industry education roadmap provides a variety of opportunities. IM University is a multi-faceted program offered in Europe. The Web-based Fundamentals of ECM Certificate Program familiarizes users with the core concepts and technologies related to ECM. The AIIM Webinars round out user education on key issues.
- **Peer Networking**—Through chapters, networking groups, programs, partnerships, and the Web, AIIM creates opportunities that allow, users, suppliers, consultants, and the channel to engage and connect with one another.
- **Industry Advocacy**—AIIM, as an ANSI (American National Standards Institute)-accredited standards development organization, acts as the voice of the ECM industry in key standards organizations, with the media, and with government decision-makers.

About AIIM

The Enterprise Content Management Association

AIIM, the international authority on Enterprise Content Management (ECM), is leading the way to the understanding, adoption, and use of the technologies, tools, and methods associated with managing documents, content, and business processes. ECM tools and technologies provide solutions to help users with four key business drivers: continuity, collaboration, regulatory compliance, and reduced costs.

As a non-profit association for more than 60 years, AIIM provides industry news and information, educational events and career development, industry standards and advocacy, and professional networking.

Our year-round membership programs include:

Market Education - AIIM provides events and information services that help users specify, select, and deploy ECM solutions to solve organizational problems.

AIIM E-DOC Magazine - The industry's leading magazine in North America.

M-iD Magazine (Managing Information and Documents) - Published in partnership with Infoconomy, the leading publication for the industry in the United Kingdom.

Content Management Solution Seminars - General and vertically focused educational seminars held in 20 cities throughout the U.S. and Canada.

IM Expo - The major educational event for the ECM industry, held in five or six locations each year across the United Kingdom.

Info Ireland - A two-day educational event held annually in Ireland.

Professional Development - AIIM provides an educational roadmap for the industry.

Fundamentals of ECM Certificate Program - A Web-based professional certificate program designed to help users become familiar with the core concepts and technologies related to Information Management Compliance.

AIIM Webinars - Provide education on the key issues and trends affecting the industry.

IM University - A multifaceted education program (Web-based education, single-day events, and multi-day residential programs) offered in Europe.

Peer Networking - AIIM creates opportunities that allow, users, suppliers, consultants, and the channel to engage and connect with one another - through chapters, networking groups, programs, partnerships, and the Web.

AIIM Expo and Conference - Produced by Advanstar Communications in cooperation with AIIM, the premier networking and education event for the industry.

AIIM Chapters - A network of 39 chapters in North America providing educational and networking opportunities at the local level for AIIM members.

AIIM Partners - A global network of organizations similar to AIIM committed to helping grow the industry.

AIIM ChannelConnection - A special series of programs and services specifically designed for document and content management VARs, system integrators, and service companies.

Industry Advocacy - AIIM acts as the voice of the ECM industry in key standards organizations, with the media, and with government decision-makers.

AIIM Standards - AIIM is an ANSI (American National Standards Institute) accredited standards development organization. AIIM also holds the Secretariat for the ISO (International Organization for Standardization) committee focused on Information Management Compliance issues, TC171.

Industry Watch - AIIM user-focused industry research and analysis.