

The
GILBANE
REPORTTM

on Open Information & Document Systems

March, 1993

Vol. 1, Number 1

WHAT THE REPORT WILL COVER, AND WHY

Welcome to the first issue of the Gilbane Report, a bimonthly report focused on issues surrounding document and information system technology. The goal of the report is to help IS managers make

well-informed short term business decisions, and to formulate long range strategies that reflect the risks and capabilities of document system technology. Each issue will examine a specific topic in some depth. Occasional case studies will be presented to illustrate particular challenges or benefits associated with implementing document system technology.

This first article sets the stage for future discussions by taking a look at what we mean by 'open information and document systems', and by examining how the notion of 'document' is changing as we take greater advantage of electronic information for strategic business purposes. This will give you a feel for the kinds of trends and the scope of the subject matter the report will track. This issue also examines how to determine what kind of 'document management system' will be the most cost effective solution for your business problem. This is an example of the guidance we will provide for shorter-term decision making.

This is an exciting area with lots of activity (see a partial list of topics to be covered inside the back cover), lots of confusion, and a fair amount of controversy. We look forward to helping you analyze the issues so you can navigate safely and profitably through the document-related technology decisions of the nineties, and we welcome your input.

▲▲▲
Publisher:
Publishing Technology
Management, Inc.
Applelink: PTM
Editor:
Frank Gilbane
fgilbane@world.std.com
(617) 643-8855
Subscriptions:
Carolyn Fine
carolyn@world.std.com
(617) 643-8855
Design & Production:
Catherine Maccora
(617) 241-7816
Associate Editor:
Chip Canty
ccanty@world.std.com
(617) 265-6263
Contributing Editor:
Rebecca Hansen
MCI Mail: 4724078

CONTENTS

What The Report Will Cover And Why	▲ Page 1
Imaging, Document & Information Management Systems— What's The Difference, And How Do You Know What You Need?	▲ Page 9
Calendar of Events	▲ Page 18
Topics To Be Covered In Future Issues	▲ Page 19

THE BUSINESS CHALLENGE

For companies that want to use information technology to automate

business processes and improve competitiveness, the biggest challenge is often riding the technology wave without falling off. If you fall behind the technology wave your competition can lower prices to gain market share or increase margins to heighten interest in their stock, while you drift behind. If you get too far out in front of the curve your business can suffer from 'beta-site-syndrome', also marked by lost efficiencies and unhappy customers.

The trick is be the first to adopt technology when its risks and benefits and cost-effectiveness are first well-understood and real for your business. Today there is no inherent risk investing in a system for product documentation that integrates text and graphics, whereas ten years ago you would have been a pioneer, and five years ago you would have been ahead of the pack. On the other hand, it is still difficult to get a quick return on investment in document systems that integrate audio, video, and high resolution color photographs. (This is not to say that you *can't* get a reasonable return, only that there has not been enough industry experience for you to assume you can without some especially prudent analysis.)

Such multimedia systems, however, will become generally cost-effective in the not-too-distant future. *When* do you invest without making a leap of faith that will keep you up nights? How do you make the right strategic decision today, i.e., one that won't lock you in? Similarly, how do you decide whether you should base a new information system around an object-oriented or relational database, especially if this system is for managing documents or multimedia objects to be used to construct documents.

Another very real decision facing many companies today is whether they should adopt an electronic distribution strategy. What kinds of risks are involved? What are competitors doing? Where does it fit on the technology adoption curve? (See Figure 1.)

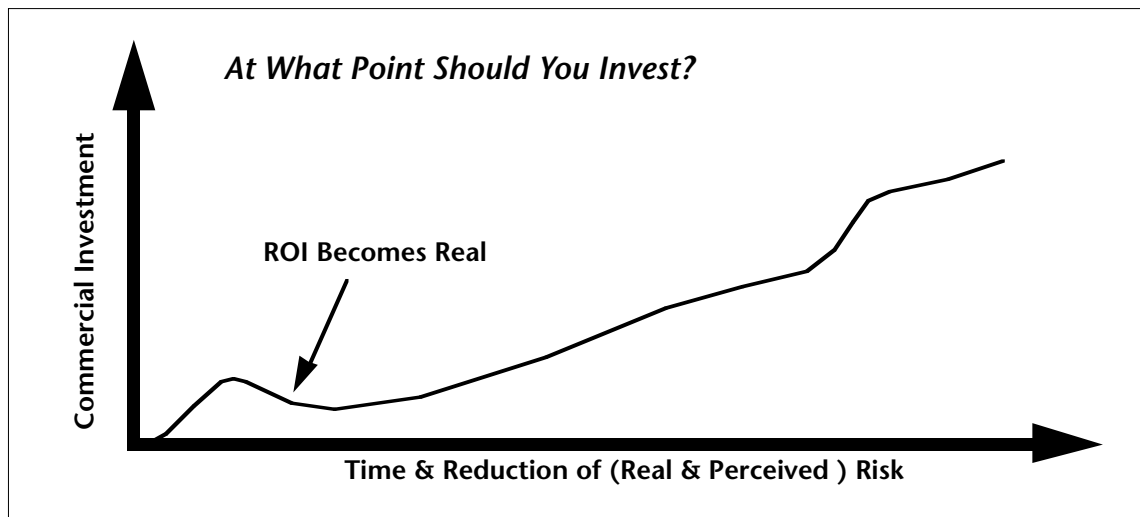


Figure 1
Rate of New
Technology
Adoption

What Is Your Risk Threshold?

There will always be technology leaders in tune with developments and trends who are willing to take risks and reap the rewards as they master new technology. The main risk they take is not usually that the technology won't work *per se*; instead we often find that it does not efficiently solve the business problem at hand. Or the ramp-up required is longer and costs more than projected. Or the technology creates new, unforeseen

business problems. In retrospect it often appears that the company's organization is simply a poor fit with the new technology.

This report is aimed at helping IS managers who want to ride the crest of practical technology to improve their business processes, and make strategic decisions that accurately anticipate the trends in usable technology. This means sometimes advocating new, untried technology when the risks are small and the potential rewards large, as well as discouraging the use of tried and true technology when it will trap you in costly inefficient business practices. Sometimes *not* trying something new is the riskiest course of action. What's important is that you know whether your decision is risky and what the risks are — you can't avoid risk, but if you have the right information you can manage it to your advantage.

THE MEANING OF 'OPEN'

It used to be that such risk lay almost solely with the buyer of systems. Today,

more and more vendors have assumed such risk by advertising their products as 'open systems'. Unfortunately, 'open' is such a popular term these days that it is sometimes criticized as being almost completely meaningless. No major platform or software vendor is going to proudly claim that their product is 'closed' or 'not open'. At best the term will simply be left out of marketing literature. If you ask different types of vendors why their system is open, you receive a variety of answers. To a platform vendor 'open' may mean that a lot of third party applications are available for their platform, or that a version of their operating system code is in the public domain. To an application vendor, 'open' may mean that their software runs on more than one platform. Open may also mean that you (the end user or another vendor) can buy support for an API. Everyone supports at least one, and sometimes many, 'open' standards.

"In fact, true open systems are not possible without open information."

The varied use of 'open' is not a conspiracy by vendors to hoodwink customers (although the term is often used a bit too conveniently). Most uses of the term are legitimate when interpreted in the correct context and, I would argue, always relative. Potential purchasers of open systems need to be knowledgeable and specific about the kind and degree of openness they want to buy. Users have to be responsible for buying intelligently, just as vendors should be responsible for accurate marketing.

In the context of information and document systems, 'open' can be used appropriately in many ways. A crucial aspect of open systems is that open systems are meaningless if the information they are supposed to share is not open. In fact, true open systems are *not possible* without open information.

This fact is largely ignored in the computer press. What good does it do you if your sixteen applications run on all four of your installed platforms, but the data they process can't be easily shared? Is this really what you mean when you say you want 'interoperability'? Have you solved your business problem? You *have* solved a big piece of the problem, but you haven't solved it all.

What is 'open information'? Ideally, open information is information that can be shared across platforms and applications without loss, corruption or compromise. The richer the information, the more difficult this becomes. It is easy to share ASCII textual information, but ASCII text is not very interesting without additional information about how it should look (what typeface, how much white space), its organization or structure (which series of characters make up the chapter title, how titles relate to paragraphs), and sometimes other *meta-information* about it (what version it is, whether it is copyrighted).

Furthermore, documents no longer contain just text, they contain different types of graphics and will soon contain voice, video, and animation. One measure of information openness is the level of complexity and the cost of conversion. How realistic is open information? How much does it cost? We'll talk more about this in future issues.

STANDARDS

Interoperability is rapidly becoming the most competitive feature a product can

have. This is creating a different view of standards — a new playing field. Standards that support interoperability between information and multiple applications are becoming as important, if not more so, than standards that support interoperability between applications and platforms.

The Gilbane Report will focus on standards that contribute to a greater level of information sharing between document and information system applications. We will analyze standards according to how well they contribute to interoperability and how well they fit with emerging technologies. Examples of the kinds of standards to be covered include standards for *encoding* information such as SGML and TIFF, standards for *accessing* information such as SQL, and standards for *presenting* information such as PostScript™. We will also consider what standards cost, what they buy, what they don't do, and whether they should even be considered 'open standards'.

EMERGING DOCUMENT SYSTEM TECHNOLOGIES

Standards that promote interoperability are important for all computer applications, but especially so for a new class of document systems that act as an

intermediary between a variety of computer programs and the computer user. Businesses are under intense competitive pressure to reduce costs and to bring new products to market faster. A number of document system technologies are emerging and converging that promise to do just that. While many of these technologies are ripe for adoption, keeping up with the rapid development and determining when they meet your business needs is a strenuous challenge, especially if this is only one of your responsibilities.

Organizations that take advantage of new document-oriented technology in an intelligent way will remain balanced on the curve. For example, the costs of documentation have become so high that some companies have actually given away CD-ROM drives to customers because it is so much cheaper to provide them with information on CDs than on paper. The costs and logistics of creating, storing, shipping, and updating large paper documents have become unwieldy. This cost is reflected in either higher product prices, or passed on to customers who want paper manuals instead of electronic documents. *If* 10% of the cost of a product is the cost of delivering product information, and *if* a particular electronic delivery strategy allows you to cut that in half, then you can either increase your margins or reduce your prices 5%. This is independent of increases in quality, customer satisfaction, added value and increased market share.

Such a strategy sounds easy, but the challenge is to ensure that your assumptions about how to get there are correct. These are the kinds of assumptions we will help you take a critical look at.

Product documentation is only one of the types of strategic information that can be managed more effectively with document system technology. Design, maintenance, administrative, experimental, and customer support data are other examples of areas where effective management of document information has become costly. According to

“Interoperability is rapidly becoming the most competitive feature a product can have.”

the Gartner Group 90-95% of all corporate data is in document (i.e., paper) form. This makes it imperative for corporations to focus on the problem of information management and distribution.

“A critically important characteristic of these emerging technologies is the trend towards their convergence in products.”

Electronic distribution is only one member of a family of emerging document-oriented technologies that are poised to become mainstream technologies in the next few years. Electronic publishing systems are moving rapidly from their bias towards paper output, and many now support both paper and electronic output equally well. Other interrelated technologies coming into their own in the next year or two include imaging systems, document management systems, full-text and structured data retrieval systems, object-oriented databases (and relational database support for objects), structured editors, and multimedia both on the desktop and ‘on the road’. Many of these technologies employ graphical user interfaces based on document metaphors, and all revolve around documents in one way or another.

A critically important characteristic of these emerging technologies is the trend towards their convergence in products. For example, it is becoming less clear where a publishing system and a database system diverge. Database products have always presented information electronically, and are now generalizing and adding publishing-like features to that capability. Publishing system vendors now offer database querying capability and provide electronic output. Both kinds of systems will increasingly add hypermedia and multimedia functionality, as well as full text and image retrieval.

Think of how this complicates interoperability. Instead of having to integrate a publishing system and a database system where there is most likely a single (though messy) point of interaction and mapping, you may have to integrate one hybrid publishing/database system with another hybrid database/publishing system. Some firms will benefit because one publishing/database product will meet all their needs. But others will have to integrate systems with those of other departments, divisions, or business partners. As for avoiding this issue by legislating the purchase of specific publishing/database systems across an enterprise, forget it. It did not work with computer platforms or software applications and there is little reason to expect it to work here.

WHAT IS A DOCUMENT?

We all have a pretty clear notion of what paper documents are. Most of the time at

least we can point to something and agree that it is or is not a document. (But not always! e.g., is a milk carton a document? An album cover? A shipping label?) For the moment at least, we can probably agree to define a paper document as a collection of information — text and graphics — recorded onto one or more pages (however formatted or organized).

This general agreement does not carry over to electronic documents. Our view of what a document is is rapidly changing as electronic documents come of age. We need a more general notion of what a document is that incorporates the additional characteristics associated with, or even required by, electronic documents. There are a number of characteristics that apply to both types of documents. For example, documents can be static or dynamic containers of information (paper documents can have multiple editions or change page additions); a document is a mechanism for applying scope to information for human consumption; a document is the interface to an information system (body of knowledge or database).

What makes electronic documents so different? For one thing, electronic documents can contain things that you can’t reasonably represent on paper, e.g., video. Electronic

//
...many
products
already
exist...that
have and use
very different
concepts of
'document'.

documents can change in nanoseconds, can interact with live data and can have built-in navigational and hyperlinking tools. The list goes on. Of course, you can also have electronic documents that don't have any of these characteristics and behave almost like paper pages — images of pages that you can view in sequence. When someone tells you they are going to send you a paper document you know pretty much what to expect; when someone sends you an electronic document, all bets are off unless they provide you with a lot of additional information, and maybe some software.

We don't propose to settle the issue of what an electronic document is here and now, rather this will be a recurring topic as we analyze what different applications think a document is. Why is this important? Because many products already exist (especially 'document management products') that have and use very different concepts of 'document'. What happens when you try integrating an imaging system, a database system, and a distribution system that all make different assumptions about the definition and role of a document? You need to think about it because the document system you buy must deal with the same kind of documents that you have to manage, and it is unlikely that marketing literature will include definitions of 'document' that are exclusionary. (We cover this question in more depth in the following article.)

You'll be hearing a lot more about documents in the next year, and not just from this report. In the January 1993 issue of BYTE, editor in chief Dennis Allen, calls 1993 'the year of the document'. This was in the context of the rapidly growing number of electronic viewing technologies, and the availability of CD-ROM recorders at about the same price as a desktop laser printer. In the December 1992 issue of BYTE, Cary Lu opened a series of articles on objects (as in object-oriented information and application objects) by talking about 'document-oriented interfaces' as the up-and-coming way for humans to interface with objects.

Documents As Interfaces?

One advantage of a document-oriented interface is that it can protect users from having to deal with several different concurrently running applications. For example, instead of using three different applications to edit text, graphic, and spreadsheet data, your interface is a single document view that transparently provides text, graphic and spreadsheet editing. Three or more applications may be operating but you don't have to worry about finding them, or how they interact, the document interface manages all this for you. You can concentrate more on getting your job done and less on fiddling with software tools. (This is similar to much of the motivation behind pen computing interfaces.)

It is extremely difficult for us to think of a collection of information independent of the notion of a document, so why fight it? Even with database information which is something of an exception to those familiar with data structures and records and fields, we still feel more comfortable with such information if it is presented in the context of a document.

Is 'Document' Politically Correct?

Documents have not always been regarded as interesting in the information technology world. They have been associated with publishing, which has been considered a dirty back room process involving chemicals, ink, and sweat, or at least irrelevant to the problems of data processing. IS managers dealt with records and fields, but rarely *pages*.

Now document imaging and electronic distribution systems are finally bringing documents into favor in mainstream IS organizations. The concept of 'pages', on the other hand, is

likely to be controversial for some time. Pages have some baggage that gets in the way of a system supporting equal opportunity among the variety of output media and presentation devices. And, of course, pages have little meaning to most database applications. How do you map pages to objects? Will objects be the pages of the future?

THE RELATIONSHIP BETWEEN DOCUMENTS AND INFORMATION SYSTEMS

Even today, document and information systems are still largely viewed as entirely different kinds of animals, each belonging in its own domain, and almost entirely inappropriate to the other. This is

changing. Information systems of the future *will be* document systems. Documents as well as the pieces of information that make up documents, in addition to traditional numeric data, are what will be managed. Document servers will function much as database servers do today. A client will send a request or query to the server (using an advanced version of SQL or a next-generation query language) to retrieve a particular set of document elements (records or objects), and assemble and present the result within a document (form). Databases will still perform data-crunching functions that feed into yet more data-manipulation functions. But for human consumption, the vast majority of information will be in document form.

This makes sense. A document is a vastly richer, more flexible, and friendlier metaphor for interfacing with information than, for example, a relational database. Even when you want to see a table of values that illustrates relational information, it is easier to understand in the context of a document, surrounded by explanatory text and perhaps graphic information in chart or animated form. Ideally, we want to manage strategic document data with the same power and flexibility with which we manage database information. Actually, we want *more* power and flexibility since databases do not yet deal very well with some of our requirements for managing multimedia objects.

The distinction between document elements, and database records and objects will become even less clear in the next couple of years, as platform vendors start supporting compound document elements at an operating system level. This will force a profound change in the way applications work together. Until we have that level of support, this convergence will be implemented through various kinds of application languages and 'middleware'. ('Middleware' is commonly used to describe any software in between an operating system and an application.)

WHAT ALL THIS MEANS TO VENDORS (AND WHY WE CARE)

Our primary focus so far has been on how such changes will affect the purchasers of open document and information system technology. It is also important to consider how this evolution affects the vendor

community. The smart buyer not only keeps in mind the issues the vendors face, but also maintains a productive dialog and business partnership with the suppliers of the technology that can solve business problems.

Vendors of traditional publishing, imaging, document management, database, and text retrieval applications are confronted with two fundamental challenges as their technologies and products become more integrated: product differentiation, and interoperability.

Differentiation in the context of the merging technologies we are talking about means more than just identifying a feature or six that the competition does not have. It also

means deciding and communicating what your business is, which in many cases means reinventing your identity. The categories that traditionally have differentiated vendors no longer do so. Database, imaging, document management and electronic publishing vendors are finding themselves with products that increasingly overlap and compete. This means re-evaluating not only product plans, but concomitant business decisions and alliances as well.

At the same time, vendors are under severe pressure to make sure their products can 'interoperate' with others, whether partners or competitors. Many suppliers now realize they need to be in the interoperability business. (This is why platform, application vendors and system integrators are building their own 'middleware'. They, and you, need to be careful that this 'middleware' supports, rather than hinders, interoperability.)

Differentiation and interoperability appear at first glance to be incompatible goals. But they're not necessarily. There is a difference between the features and functionality of information processing, and the way it is shared with another application. Applications can share *what* they do without sharing *how* (well) they do it. In any case, reconciling differentiation and interoperability demands is as much of a challenge for vendors as it is for you.

WHAT'S COMING

These are the kinds of issues that will be covered in The Gilbane Report. In future issues, we will return to many of the areas we touched on here, analyze them in greater depth, and recommend courses of action. The inside of the back cover of this issue lists some of the specific topics to be covered.

IMAGING, DOCUMENT & INFORMATION MANAGEMENT SYSTEMS

*WHAT'S THE DIFFERENCE,
AND HOW DO YOU
KNOW WHAT YOU NEED?*

EXECUTIVE SUMMARY

Strategic Overview

- New document management systems can solve business problems involving the management of a variety of types of strategic corporate data. Despite growing interest in them, there is also much confusion about what document systems are and how to differentiate among the wide assortment of products available.
- Managers need a top-level methodology for determining how to investigate and evaluate potential document management solutions.
- A three-pronged approach is recommended for: (1) analyzing existing business processes and determining how solutions based on different technologies can change them, (2) examining the characteristics of the information you need to manage, and (3) mapping these to appropriate terminology and product categories.

Imaging, Document, and Information Management Systems

- Modern imaging systems, document management systems, publishing systems and database systems overlap greatly in the functions they perform, but each has its own areas of strength, and each is appropriate for different applications. Understanding how they differ and how each is evolving is essential in selecting among them.
- Most important is understanding exactly what kinds of information you need to manage. In choosing one type of solution over another you should also examine how complex that information is, what kinds of processing you need to perform on it, and how often such processing must take place.

Risks & Costs

- There are many hidden costs and risks to consider when buying systems. Besides on-going operational costs, these include risks deriving from continuing technological change and the cost of migrating data from today's technology to tomorrow's.
- Certain types of standards can reduce long term risk by preserving the integrity of strategic information.
- Spreading implementation costs over time often improve return on investment and cost/benefit calculations, over both the short and long term.

Conclusions & Recommendations

- Today's document management technologies are mature enough to offer real benefits with relatively low risk. Many products can help companies reduce the cost of managing documents and/or the information contained within them.
- A balanced analysis of your business processes, the available technology and trends, and the terminology used by different suppliers helps to narrow the scope and cost of evaluating such products.
- Plan now for sharing document data with other applications in the future. With today's

trend toward standards, systems that fail to provide for *interoperability* of data may become obsolete before yielding any return on investment. Document management problems should *never* be solved in isolation.

STRATEGIC OVERVIEW

It is intriguing these days to ask IS professionals and vendors to define document management. What classifies as a document management product? Is document management something you accomplish with imaging technology, publishing technology, database technology, or all three? The diversity of products available has resulted in a wide variety of views.

If you think document management is something you need, how do you go about determining which technologies and products to look at? How do you narrow the scope of your search?

This article clarifies issues facing managers who are thinking of investing in a document management system. It also suggests how to begin an analysis that maps your information management requirements to the most appropriate technology. Its objective is to provide a framework for evaluating your own information management needs against the many overlapping technologies and products available to meet them.

In particular, it addresses the respective roles of imaging systems, electronic publishing systems, and document management systems, and it discusses how all three can be combined with database technology to manage images, documents or other elements of information.

It does not pretend to cover everything you need to know or do in choosing such a system; instead it describes an overall approach proven in large information management strategies now being deployed in industries such as defense, aerospace, automotive, telecommunications, pharmaceutical, financial and insurance. Future issues of this report will analyze specific applications of this approach, as well as specific technical and business issues encountered in these industries.

The following sections cover: (1) how to analyze the role of imaging and electronic publishing systems in managing information, (2) how to begin the process of analyzing your own requirements, (3) how to protect your investment in information, and (4) how to balance start-up costs against the ongoing expense of managing documents and information.

First Things First

The first step in determining what sort of information management system is most appropriate to your organizations needs is to identify exactly what those needs are. While this may seem obvious, it is tempting to look first at the technology and products available. Too often this results in a forced match between your needs and a particular system that may not produce significant cost savings or other benefits.

On the other hand you also should not set business processes in stone without evaluating all available technology. In many cases, new technology will encourage if not require fundamental changes in how you run your business. You cannot know what these changes will be without thoroughly analyzing what the technology can do.

Suppose for example that you install an information distribution system that allows customers to access product information over a network or by phone. Doing so will almost certainly force you to re-evaluate other business processes. The shipping department, the corporate publishing department and the engineering organization will

all be affected, sometimes in unexpected ways. Fundamentally new processes will be needed for moving information in and out of this information hub, which in turn will require changes in every department contributing or accessing this information. New possibilities for inter-departmental communication will emerge. You will not save anything just by dropping in a new product information system. Both the technology and the business processes must be integrated. In other words, your business processes must be re-engineered.

Many types of analyses are useful in reviewing existing business processes and organization. The most relevant questions are:

1. Where does information come from?
2. Where does it need to go?
3. What needs to be done with the information?
4. Which kinds of pieces of information do you need to isolate, manipulate and manage?

The answers to these questions will vary depending on your business and your function within it. Those who manage personnel records will give very different answers than those who manage engineering processes. The *complexity* of the information varies widely; so does the level of *dynamic access and interaction* required to process it.

How you answer question 4. is *critical* to a successful decision about the kinds of information management system you should consider. Unfortunately, this question too often is not even asked.

IMAGING, DOCUMENT, AND INFORMATION MANAGEMENT SYSTEMS

The terms imaging, electronic publishing, document management, and information management are used in many confusing and conflicting ways. While it may seem clear-cut at first glance which kind of

system you need, there is in general no way to find which solution is best for you without examining the capabilities of more than one of these systems. For example, a document management or electronic publishing system might provide imaging capabilities more appropriate to your requirements than an 'imaging system'. To resolve this confusion, you first need to determine what you mean by these terms; only then can you communicate effectively with suppliers, and make them address your particular requirements.

Imaging Systems

Imaging can be defined as the capture or conversion of information from hard copy media to digital raster files. The hard-copy information could include text, graphics, photographs and forms. Creating electronic images of paper documents provides little benefit unless those images can be easily accessed; for this reason, suppliers of imaging systems typically provide some way to identify, index and retrieve image files.

Imaging systems have been developed for many different applications, but most fit into three categories. Some are designed to manage engineering data (e.g., CAD drawings), others to handle business data (invoices, personnel records), and still others to manage graphic arts data (color photographs). How these images are stored, retrieved and otherwise managed is what distinguishes one vendor's products from another's. Because these solutions are so different, it is especially important that buyers study these products carefully and that they understand this distinction.

Some imaging systems include, or are supplemented by, some form of optical character recognition (OCR) or raster-to-vector conversion capability. Both transform raster images

into other more useful forms of processable data, thus enabling imaging systems to behave more like database or publishing systems, and dramatically increasing the number of business applications that such systems can handle.

Publishing Systems

Publishing — which we define broadly as the distribution and presentation of any human-readable information on any medium — includes the preparation of information both for distribution on CD-ROM or paper and for presentation on paper or electronic displays. Given this definition, most (if not all) imaging and database systems also publish information, although usually with fewer options for formatting it. On the other hand, mainstream publishing systems today do much more than simply format information; many provide ways to create and manage it prior to publication. This information may include images as well as other forms of electronic text and graphics.

Imaging and publishing systems both provide some level of information management and often overlap in functionality. Imaging systems typically manage images and documents, while publishing systems typically manage documents and combinations of document elements; chapters, pages, paragraphs, images, and such other data elements as numerical data extracted directly from a spreadsheet. Figure 1 illustrates this overlap. (If you're familiar with database management you will recognize that the process pictured is analogous to data capture, organization and manipulation in databases, and presentation of queried information.)

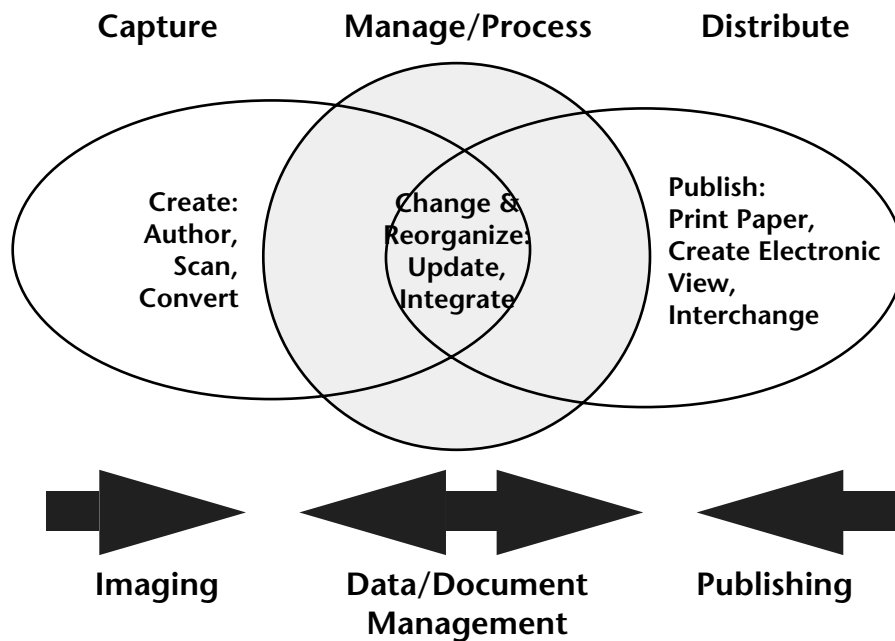


FIGURE 1
The Overlap
Between
Imaging,
Publishing,
Document
Management,
And Database
Suppliers.

There is a symmetry here: Both imaging and publishing systems began with a focus on processing static chunks of information (raster images and paper pages, respectively), and both are evolving towards each other as they increasingly concentrate on managing the *information content* on these chunks of information dynamically (hence the inwardly pointing arrows in Figure 1). An equally important trend is that of database vendors responding to market pressure to incorporate images and other types of document elements in their products (thus the expanding arrows in Figure 1). Understanding these trends will help you put the spectrum of products in perspective.

“Document management systems, like other information management systems, manage pieces, or objects, of information.”

What Does The Product Landscape Look Like?

A large number of products target image and document management applications. Even if you exclude related products focused on authoring, electronic publishing, workflow management, data conversion, and text and data retrieval (not to mention products developed by system integrators), you still face a bewildering array of choices.

Products that fit into the bubble on the left in Figure 1 include imaging systems from companies like FileNet, IBM, Wang, Plexus, Mainstay, & PRC. In the bubble on the right are systems from electronic publishing suppliers such as Interleaf, Datalogics, Xyvision, XSoft, and Agfa. Companies concentrating on document management include Documentum, Micro Dynamics, Boss Logic, and Workgroup Solutions. (These of course are partial lists.)

Which category, or combination of, products should you look at to solve your business problem? One way to narrow down the field is to evaluate carefully what kinds of objects you need to manage. This will tend to point you towards one category or another.

Document Management & Information Objects

Document management systems, like other information management systems, manage pieces, or objects, of information. These information objects can be either documents, images, database records, or other discrete data elements such as a part numbers or mean-time-between-failure values.

Defined this broadly, therefore, the term document management system in itself carries little meaning and is not very useful in sorting out the differences between, say, publishing and imaging systems. *The first thing to find out about either so-called document management or information management system is what kind of information objects it is designed to manage.* Are these the same objects you need to manage? Which level of object management will provide you with the most cost/benefit?

Figure 2 shows how different the information objects in need of management can be; it also demonstrates, however, that one can start with the relatively simple task of managing whole documents or pages and evolve into the more complicated job of managing pieces of information contained within a page. (Note, too, that the cost of implementing and maintaining an information management system increases in the direction of the arrow.)

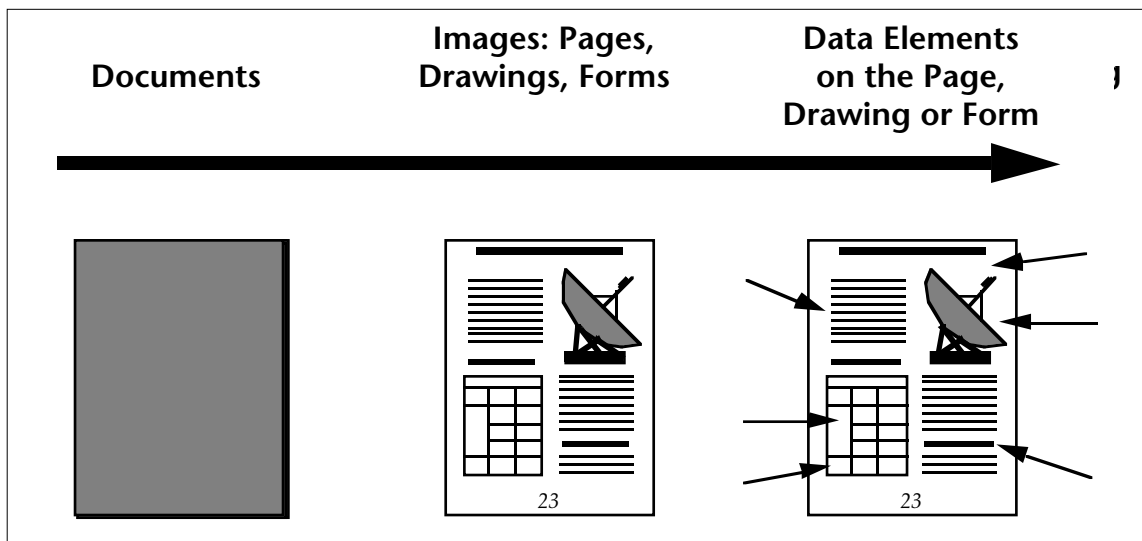


FIGURE 2
At What Level
Do You Need
To Manage
Information?

If analysis determines that the most effective way to run your operation is simply to manage only entire documents, your implementation costs and technology risks will be minimal. You may be able to obtain a significant return on your investment simply by storing documents electronically and reclaiming real estate now used to store paper documents. If, however, your business requires that you retain and retrieve individual forms, drawings or pages of documents, then you must invest instead in a system that manages these objects. Such a system requires more sophisticated technology and costs more, but if that's what you need, then that is what you must plan for. Here the return on investment comes from faster data retrieval — finding paper documents can be so slow — and from faster customer service. Finally, if you need to manage discrete information *within* such forms, drawings or pages, a much more significant up front investment is required. In return, however, you get more flexible access to and control over your data, such as the ability to automatically verify or update all part numbers in your documents.

RISKS & COSTS

Protecting Your Investment in Information

Perhaps the most important decision to make, especially for information that is strategic or that must last for a long time, is how to ensure that the information is accessible and accurate for as long as you need it.

If your information must last more than a couple of years, you need to be concerned now about converting the information later to new computing platforms or software applications. *This conversion cost can easily be greater than the cost of the new hardware and software.* In addition, the integrity of the information is almost certain to be compromised. There is no such thing as 100% automatic conversion. To protect your investment, you need a way to ensure that your information will be available to you when you need it, and that it is in a form that is still useful to you.

Open Information, Standards & Conversion

The value and quantity of information organizations have to manage will continue to grow rapidly. Some of the largest information handlers in the world have embraced similar strategies for protecting their investment in strategic information. Inspired in part by the U.S. Department of Defense's CALS¹ initiative, industry consortiums and associations in the defense, commercial airline, aerospace, automotive, and telecommunications markets have adopted internationally accepted information encoding standards as the foundation of systems for information interchange and delivery. Three standards these efforts have in common are SGML for text, CGM for graphic illustrations and CCITT Group IV for raster images.² The chief benefit of these encoding standards is that they are completely independent of hardware and proprietary software, and thus provide important data with the requisite *staying power*.

The term 'standard' is so overused these days that it has become almost meaningless. Dozens of would-be standards now apply to information management — far too many to be either useful or cost-effective. To meet the criteria of providing vendor independent staying power, information encoding standards today must meet three requirements:

1. They must be internationally accepted, either by a neutral standards-setting body such as the International Organization for Standardization (ISO), or by an industry consortium with support from all major players in the market.

¹ CALS—Computer-aided Acquisition and Logistics Support.

² SGML—Standard Generalized Markup Language, CGM—Computer Graphics Metafile, and the CCITT International Telephone and Telegraph Consultative Committee recommendation for encoding of raster files.

2. They must be supported by products and tools in the marketplace, including a variety of off-the-shelf products.
3. They must perform the required functionality, by which we mean they must be capable of expressing the richness of the information.

SGML, CGM and the CCITT raster standard all meet these requirements more than others. Explosive growth in the demand for SGML is due in large part to its richness — its ability to describe complex multimedia databases as well as simple text — making it particularly suitable for use in systems that manage different types of information objects. Though not the only important standards, these are clearly among the safest to adopt today, given the large segments of government and industry subscribing to them and the resulting market demand that forces vendors to support them.

“It is a fact of life, however, that there will always be a need for data conversion.”

Standards Are Essential, But They’re Not A Panacea

In general, conversion of information from one encoding scheme to another, whether a standard or not, is something to avoid for two reasons. One is that it is hard to preserve data integrity in conversion; another is that it costs so much. It is a fact of life, however, that there will always be a need for data conversion. Standards cannot eliminate this need — indeed, different applications of the same standard often create a need for conversion — but standards make this problem significantly more tractable and less costly, and help us get on with our jobs.

There are alternative ways to deal with this problem. Under one scenario you would convert all your information to conform to the new encoding standards, then migrate to information capture and creation tools based around the same standards. This approach works best when the conversion process is finite — that is, when you control the creation of information or can dictate how others deliver it to you.

A quite different scenario calls for you to adopt an ongoing conversion process as part of the day-to-day information flow. In this case you do not replace your data capture and creation tools, you simply add data conversion tools. If conversion technology were

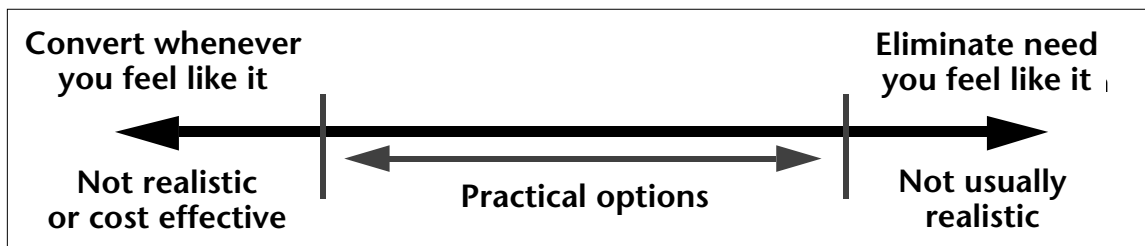


FIGURE 3
The Spectrum of Conversion Options

perfect and inexpensive, it would not matter (much) that you were forever converting data. But it isn't, and few companies enjoy total control over data creation. As a result, for most firms the most practical alternative is a strategy that falls somewhere between these two extremes. (See Figure 3.)

Approaching a Cost/Benefit Analysis

Typical cost/benefit studies or return-on-investment analyses for computer-based systems tend to be biased towards initial equipment costs and ongoing equipment maintenance. Often other direct costs and indirect cost savings associated with information life-cycle maintenance are neglected. Factoring in these other costs and benefits can significantly enhance the appeal of an investment in new technology, even in the face of pressure to keep equipment costs low.

Indirect costs savings include those associated with getting the information portion of a

product to market on time, sharing information with business partners, serving customers better due to faster information access, and using information more efficiently. Such savings can be substantial and can represent value that cannot be obtained through other means.

Spreading Out the Cost

Firms should also consider spreading out the costs of implementing a document management system by gradually migrating in the direction of the arrow in Figure 2. No matter how appealing a sophisticated information management system might be, many firms simply do not have sufficient resources to implement one right away.

In such cases it sometimes makes sense to migrate slowly toward the ideal system by first scanning all documents into an imaging system, then extract more meaningful data from the images only as needed over time. Since conversion represents a substantial part of the overall cost of implementing a new system, spreading it out in this manner might make such a system more affordable.

Another way to save money is to focus on the information that is most frequently used. You may even find that converting as little as 15-20% of your information yields 80% of the benefits that you could enjoy by converting all your data.

Incorporating conversion into the on-going process of updating documents also significantly reduces costs. For example, if an author must re-write a personnel procedure, it may take only 5% more effort to convert the data to the new format at the same time. If when writing the new procedure the author uses tools that directly support the target encoding standard, then the conversion may involve no extra effort at all.

Initial & Life-cycle Costs

The initial costs of implementing an information management system can vary widely, depending to a large degree on what kinds of information you need to manage. Even very high initial costs can appear insignificant in the context of life-cycle maintenance costs, especially for information that must last a long time.

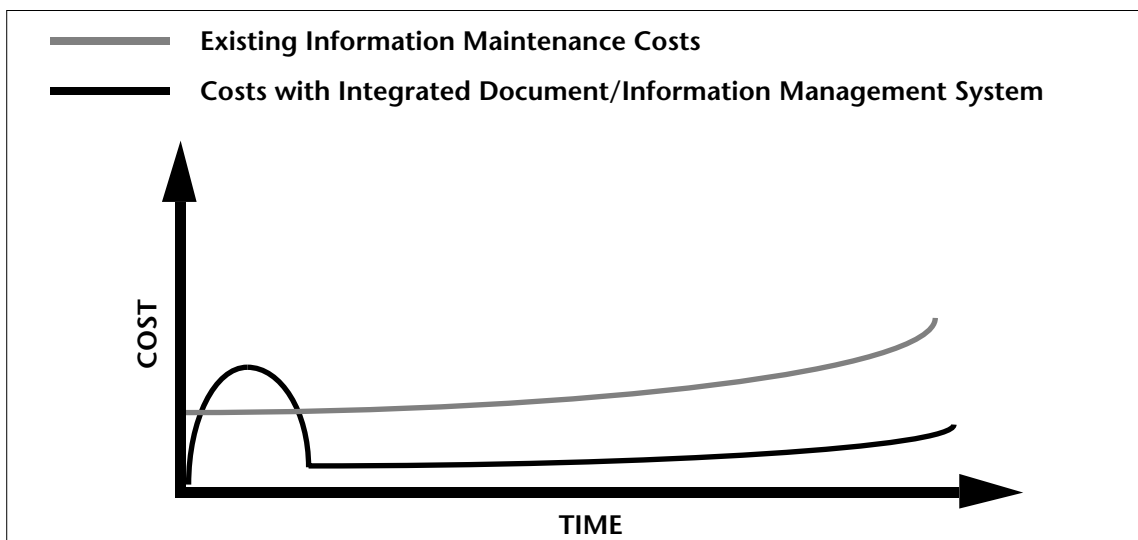


FIGURE 4
*Relative Initial
and Life-cycle
Costs*

Figure 4 illustrates the relation between initial and life-cycle costs. The grey line shows gradually increasing costs of maintaining information with existing technology and processes. The black line represents the increase in short-term costs in relation to longer-term savings. Note that the difference between the lines is large enough to show a return on investment at time T1. Drastic savings are realized, however, as they accumulate up to time T2 and beyond.

Whether the time in Figure 4 is measured in weeks, months, or years will depend on your own business, and the differential will vary. Nonetheless, if you carefully analyze your business processes, your information flow and the kinds of information objects you need to manage, and if you map these to appropriate standards and technology, a version of this illustration based on your own numbers should exhibit a similar relationship. Expressed this way, such savings can be impressive enough to convince even the most conservative financial managers.

CONCLUSIONS & RECOMMENDATIONS

Analyzing, designing and implementing a document/information management system of any kind is a complex and often challenging task. This article addresses

some of the top level management issues that arise in planning such a transition, and it provides a framework to help start this process.

Today's imaging and document management products can provide significant benefits to many different kinds of firms. The technology for managing text elements, images and other still graphics is widely available and maturing rapidly. As a result, the technology risks are relatively low. The remaining challenges are (a) in configuring and integrating all such products so they work together effectively, (b) developing a strategy that ensures that the implementation maps appropriately to the company's business processes, and (c) anticipating and pro-actively addressing future interoperability requirements. It helps to view information technology adoption as a process of evolution, and to treat implementation strategies as migration strategies.

In summary, before selecting and implementing any document management solution, you should:

1. Analyze existing business processes and practices and determine how they might become more efficient.
2. Analyze the information flow and the characteristics of information objects that you need to manage. Here you can start thinking about the relative merits of imaging, document or other object management technology. Determine whether your use of document terminology differs that of your vendors.
3. Decide which encoding standards make the most sense for the information objects you need to manage and the level of interoperability you need today and in the future. Where appropriate standards do not exist or are still emerging, factor additional risk into your analysis.
4. Perform a cost/benefit analysis that includes both short-term and life-cycle costs, as well as strategies for spreading out costs over time. Validate your plan by doing cost benefit analyses as well for other kinds of systems and for the same system under other assumptions about your future needs.
5. Choose an imaging, publishing, or document/information management system that satisfies the requirements of 1, 2, and 3.
6. Implement a manageable, well-defined solution where the major risks are understood, and quantifiable. (The risks may be high or low depending on your risk tolerance and market pressures.) This solution should be designed to test your assumptions about the technology and about the effect of the new technology on your business process.
7. Scale-up with modifications and fine-tuning based on your real-life experience.

“Today’s imaging and document management products can provide significant benefits to many different kinds of firms.”

CALENDAR OF EVENTS

Below is a selection of key events covering open information and document system issues. There are many other conferences

and shows covering related topics. We will attempt to keep this list to those events that focus on areas most directly related to the areas covered in our report.

On-line Publishing. March 22-24, 1993, Pittsburgh, PA. GCA conference for electronic information publishers. Call (703) 519-8160, Fax (703) 548-2867.

Intermedia '93. March 30-31, 1993, San Jose, CA. Multimedia and CD-ROM conference. Call (203) 964-0000, Fax (203) 964-9209.

AIIM Show. April 5-8, 1993, Chicago, IL. The Association for Information & Image Management annual conference and exposition on imaging/document management. Call (301) 587-8202, Fax (301) 587-2711.

Seybold Boston '93. April 13-16, 1993, Boston, MA. The annual gathering of the computer publishing elite. Conference and exposition. Call (310) 457-8500, Fax (310) 457-8510.

Marathon Week. April 18-23, 1993, Boston, MA. Patty Seybold's new combined show, including the Distributed Object Technology, Open Systems, and Executive Forum. Call (617) 742-5200, Fax (617) 742-1028.

SGML Europe '93. May 9-12, 1993, Rotterdam, The Netherlands. This new GCA conference replaces 'International Markup'. Call (703) 519-8160 or +44 793 512 515, Fax (703) 548-2867 or +44 793 512 516.

Documentation Europe. May 13-14, 1993, Rotterdam, The Netherlands. GCA's combined conference and exhibition. Call (703) 519-8160 or +44 793 512 515, Fax (703) 548-2867 or +44 793 512 516.

Seybold Paris. June 1-3, 1993, Paris, France. A new Seybold conference and exhibition on electronic publishing. Call +44 32 341 0561, Fax +44 32 341 0279.

TechDoc '93. August 22-27, 1993, Anaheim, CA. GCA tutorials and seminar. Call (703) 519-8160, Fax (703) 548-2867.

CALS Europe '93. September 1993, Berlin, Germany. Conference and exhibition on CALS-related activity in Europe. Call (202) 775-9556, Fax (202) 775-8122.

CALS Pacific '93. October 1993, Melbourne, Australia. Conference and exhibition for CALS activities in the Pacific Rim. Call (202) 775-9556, Fax (202) 775-8122.

Seybold San Francisco. October 20-23, 1993. San Francisco, CA. The enormous computer publishing exhibition and conference. Call (310) 457-8500, Fax (310) 457-8510.

CD-ROM Expo. October 27-29, 1993, Boston MA. Conference, tutorials, and exhibition. Call (617) 361-8000, Fax (617) 361-3389.

CALS Expo '93. November 1-4, 1993, Atlanta, GA. The annual conference and exhibition. Call (202) 775-1440, Fax (202) 775-1309.

TOPICS TO BE COVERED IN FUTURE ISSUES

The subjects listed below are some of the areas we will be covering, in no particular order. If you have an opinion about which topics you would like to see added or covered sooner rather than later, let us know.

- Electronic Distribution** — Where Are The Cost Savings? What Real Life Experiences Have Companies Had?
- Electronic Distribution** — Does One Size Fit All? Who Are The Players? What Are The Options? Are Pages Important?
- Documents As Interfaces** — Is This An Option For Today? What Will The Future Bring?
- Query Languages** — What Makes Sense For Structured Text? What Makes Sense For Document Systems?
- SGML & Presentation Interchange** — What Standards Are Available Or Appropriate? (DSSSL, OS/FOSI, HyTime, ODA, etc.)
- Data Conversion & Capture** — What Level Of Data Integrity Can You Expect?
- Cost Analysis And Return On Investment Strategies For Document System Technology** — What Methodologies (And Magic) Are Available?
- Document Management & Databases** — How Do Current Document Management Products Integrate Database Technology? How Will Future Products Do It?
- From Bits To ASCII To Information Objects** — How Does This Evolution Affect Information And Document Systems?
- Authoring Systems** — Do You Need Different Kinds For Different Media?
- “Middleware”** — What Is This Layer Of Software In Between Operating Systems And Applications? Is It The New Proprietary Trap? What Does It Mean To Your Decisions About Document Systems?
- ISO 9000** — What Kind Of Document Management System Do You Need To Meet This Quality System Standard?
- Open Systems & Client Servers** — What Are They? How Do They Relate To Document System Technology?
- Document Elements & Distributed Objects** — How Do They Relate To Each Other?
- CALS & IETMs**— What Are They? How Do They Influence Open System Technology?
- Imaging Technology** — How Is It Evolving?
- The Airframe And Airline Industry’s Strategy For Sharing Product Information** — What Can You Learn From It?
- The Telecommunication Industry’s Strategy For Sharing Product Information** — What Are The Expected Benefits? Who Is Signed Up?
- The Automotive Industry’s Strategy For Distributing Emission Control Service Information** — Will or Should This Strategy Be Applied To Other Information?
- New Drug Applications** — What Document System Strategies Make Sense For The Pharmaceutical Industry?

Order Form

- Please start my subscription to: The Gilbane Report on Open Information & Document Systems (6 issues). Back issues available for \$45.

Special Introductory Offer!

BEFORE April 1, 1993
AFTER April 1, 1993

U.S.A.	Canada	Foreign
\$195	\$202	\$212
\$225	\$232	\$242

Please send me additional information on:

- Consulting Services Special Reports
 On-site CALS Strategic Planning Seminar
 On-site Open Document System Strategic Planning Seminars
- My check for \$ _____ is enclosed Please bill me
- Please charge my credit card MasterCard Visa American Express

Name as it appears on card _____ Number _____

Signature _____ Expiration date _____

Checks from Canada and elsewhere outside the U.S. should be made payable in U.S. dollars. Funds may be transferred directly to our bank: Baybank Boston NA, 175 Federal Street, Boston MA 02110, S.W. code BAYBUS33, into the account of Publishing Technology Management, Inc., number 1444-89-63. Please be sure to identify the name of the subscriber and the nature of the order if funds are transferred bank-to-bank.

Name _____ Title _____

Company _____ Department _____

Address _____

City _____ State _____ Zip _____ Country _____

Telephone _____ Fax _____ Email _____

Mail or fax this form to: Publishing Technology Management, Inc., 46 Lewis Avenue, Arlington MA 02174-3206
Fax: (617) 648-0678 • To order by phone call: (617) 643-8855

How To Find Out More About Companies Mentioned In This Issue

Agfa
200 Ballardvale Street
Wilmington, MA 01887
(508) 658-0200

Boss Logic
1901 Landins Drive
Mountain View, CA 94043
(415) 903-7000

Datalogics, Inc.
441 West Huron St.
Chicago, IL 60610
(312) 266-3125

Documentum
5724 West Las Positas Blvd.
Pleasanton, CA 94588
(510) 460-4115

FileNet
3565 Harbor Blvd.
Costa Mesa, CA 92626
(714) 966-3400

IBM Corporation
Old Orchard Road
Armonk, N.Y. 10504
(914) 765-1900

Interleaf
Prospect Place, 9 Hillside Ave.
Waltham, MA 02154
(617) 290-0710

JTS
4211 Young Street, Suite 620
Willowdale, M28 2A9
CANADA
(416) 512-8910

Mainstay
5311-B Derry Ave
Agoura Hills, CA 91301
(818) 991-6540,
Fax (818) 991-4587

Micro Dynamics, Ltd.
8555 Sixteenth Street
Suite 701
Silver Spring, MD 20910
(301) 589-6300

Plexus
1310 Chesapeake Terrace
Sunnyvale, CA 94089
(408) 747-1210

PRC, Inc.
1500 Planning Research Drive
McLean, VA 22102
(703) 556-1000


Wang Laboratories, Inc.
One Industrial Avenue
Lowell, MA 01851
(508) 459-5000

Workgroup Solutions
76 Blanchard Rd.
Burlington, MA 01803
(617) 229-9000

Xerox
XSoft Division
10200 Willow Creek Rd.
San Diego, CA
(619) 695-7700

Xyvision, Inc.
101 Edgewater Dr.
Wakefield, MA 01880
(617) 254-4100

© 1993 Publishing Technology Management, Inc. All rights reserved. No material in this publication may be reproduced without written permission. To request reprints or permission to distribute call 617-643-8855.

The Gilbane Report and  PTM are registered trademarks of Publishing Technology Management, Inc. Product, technology and service names are trademarks or service marks of their respective owners.

The Gilbane Report on Open Information & Document Systems is published 6 times a year.

The Gilbane Report is an independent publication offering objective analysis of technology and business issues. The report does not provide advertising, product reviews, testing or vendor recommendations. We do discuss particular pieces of product technology that are appropriate to the topic under analysis, and welcome product information and input from vendors.

Letters to the editor are encouraged and will be answered. Mail to Editor, The Gilbane Report, Publishing Technology Management, Inc., 46 Lewis Avenue, Arlington, MA 02174-3206, or fgilbane@world.std.com or APPLELINK:PTM

ISSN 1067-8719