THE GILBANE REPORT™
on Open Information & Document Systems

Vol. 2, No. 4
September/October 1994

Publisher:
CAPV/Publishing Technology Management, Inc.
ptm@world.std.com
(617) 837-7200

Editorial Director:
Frank Gilbane
fgilbane@tiac.net
(617) 576-5700

Subscriptions:
ptm@world.std.com
(617) 837-7200

Design & Production:
Catherine Maccora
(617) 241-7816

Associate Editor:
Chip Canty
ccanty@world.std.com
(617) 265-6263

Acting Editor:
David Weinberger
self@evident.com
(617) 738-8323

## Corporate Publishing on the Internet — Is It Realistic Yet?

There is no question that the Internet provides an important new channel for organizations to communicate with employees, users, and buyers. But beyond that simple statement, most others end with questions marks. What are the benefits? What are the costs? Which technologies and standards will win? Is HTML a done deal? Will SGML shove HTML aside to assume the position it seems to believe it deserves? Will the Internet become commercialized? How long will it be "free"? What should you be doing now to take advantage of what Internet offers? How can you prepare for what it will offer in the next few years?

In this issue we take a serious look at Internet publishing, asking the one question that conditions all the questions listed above: How realistic is it as a useful and reliable tool for mainstream business use?

## PTM & CAPV Merge

Publishing Technology Management, Inc. (Publisher of this Report) and Charles A. Pesko Ventures completed a merger on December 1. The new company will continue to publish the Gilbane Report, and Frank Gilbane will remain the Editorial Director. The only substantive changes subscribers will see are an expansion of coverage into some areas of the document management market where CAP has extensive and widely recognized expertise (for example, on-demand printing technology).

The combined resources of PTM and CAPV will ensure that both company's existing consulting, conference, and information services keep well ahead of the rapid growth of the document management and computing market and able to provide you with the information you need to make sound strategic decisions regarding document system technology. Welcome to our new partners!

## Contents

# CORPORATE INTERNET PUBLISHING —
## *IS IT REALISTIC YET?*

## EXECUTIVE SUMMARY

### Strategic Overview

- The Internet is now becoming an important channel for corporate communications.
- The growth of Mosaic as the "killer app" that makes the Internet useable means that documents are becoming the primary communications vehicle over the Net.
- Mosaic documents are hyperlinked and active, but are in some ways less sophisticated than "normal" documents. They change the way users interact with information. To create Mosaic documents that are effective requires experience, thought, and skill.
- Now is the time to begin exploring publishing on the Internet, in preparation for its wider acceptance.

### Corporate Publishing & The Internet

- The Internet can be accessed by 15M-30M people today, and is growing exponentially. It can enable you to interact with your markets in new ways.
- It is a free channel, although there are costs associated with building and maintaining the documents and information delivered through that channel.
- Internet publishing is enabled by the growth of the World Wide Web as the infrastructure and Mosaic as the software that makes the infrastructure useable.
- Mosaic has gained rapid acceptance because it is easy to use, powerful, and free.
- Mosaic displays documents that are in HTML format, a simple markup language.
- By clicking on a hyperlink in a Mosaic document, a new document is fetched from somewhere on the planet; users can browse with no knowledge of the network they are browsing.
- Mosaic is being used to publish corporate information internally and externally. (The Internet is also being used by publishing companies as an order-taking and fulfillment channel.)
- To publish on the Internet you can: use an HTML authoring system, write in a specialized HTML authoring system, filter documents into HTML, generate HTML from a database, link to a different electronic viewing tool, or use SGML.
- However you accomplish it, your Mosaic authoring system should be a part of your larger document management system.
- HTML will become a more sophisticated and powerful language for expressing documents.
- Internet will not remain a free channel.

- The Internet will be subsumed by the Information Highway.
- Internet publishing will become a mainstream part of every business as the Internet/Highway becomes accessible from every desktop.

### Risk and Costs

- While the Internet is free and some Mosaic viewers are free, there are still costs to you and users associated with Internet publishing.

### Conclusions & Recommendations

- Begin investigating the benefits and costs.
- Examine what others are doing on the Internet.

## Strategic Overview

Although the Internet has been around for 25 years, it is only now becoming recognized as an important tool for businesses and other organizations. In a strong sense, this fact is evidence of the power of documents.

The "killer app" for the Internet is, as everyone has probably heard by now, Mosaic — it is the application that gives people a reason to use the Internet. Mosaic has achieved prominence as the *de facto* application standard because it brings the power of documents to the Internet, and brings the power of the Internet to documents. It provides a window to information that looks like a traditional document, but animates that document with hyperlink buttons that can load another document anywhere around the world. The linking has been important to Mosaic's acceptance, of course, but the fact that this occurs within a document — the most user-friendly graphical user interface to information ever invented — is just as important.

Now organizations are faced with an "insurmountable opportunity." They have a channel, virtually free, for delivering information to perhaps 30 million people around the world. What to do with it?

In one sense, the Internet is just another output for a publishing system (and, don't forget, that means it needs to be built into your document management system). But, in another sense, it is a medium with its own strong message. Mosaic documents can do things normal documents cannot. They also can't do some things publishers take for granted — like centering text. Perhaps more important, the nature of the user's interaction with Mosaic changes the publishing dynamic. Mosaic documents are meant to be browsed, not read. They serve as jumping off points. The nature of the Mosaic technology combined with the Internet's counter-cultural culture require an organization to present itself appropriately when making its presence felt on the Net.

Now is the time to begin. The Internet publishing web is only in its infancy, but it will rapidly achieve critical mass; this is what will draw people onto the Web, beyond today's mainly technical audience.

And it is the right preparation for that of which the Internet is only a precursor — *the Information Highway, The Infobahn, Rue de Al Gore*, call it what you will. Those who succeed at Internet publishing now not only receive the benefits of the Internet channel but will be well-positioned to move into the next new world of publishing even more rapidly.

# Internet Publishing — Is It Realistic Yet?

A conventional wisdom has emerged. Internet is the new way to communicate. Mosaic is the killer app. Mosaic makes documents hyperlinked, multimedia interactive, and possibly hyperactive. The Big Name corporations are jumping on board.

From this conventional wisdom comes a conventional anxiety. Is it time to get your feet wet, to dive in, or merely to watch from the shoreline?

In this article, we'll provide both background and perspective. Yes, the Internet is an important new way to publish. No, it is not cost free, and no, it is not going to do much for your top or bottom lines — at least not yet. And, yes, it is time to get your feet wet and, very likely, to dive in.

## The Potential

What's so great (potentially) about the prospect of publishing on this super-network of interconnected computer networks?

### *Huge market*

First (and second and third), it's the sheer size of the Internet. Estimates of the potential Internet audience range from 15 to 30 million users worldwide today. Most observers agree that the Internet has been growing exponentially in recent times. This is the driving force behind all of the hype now surrounding the subject.

One underlying reason for the Internet's growth is that the medium grows richer in content with each added subscriber, because – unlike on the commercial services – every person who signs on can be a producer as well as a consumer of information. More content, more value, more users ... more users, more content, more value.

For the majority of Internet users worldwide, access to the net is provided through their place of business or their school. But access from the home is growing rapidly, especially in America. Some 30% of US households have home computers; of those homes with children, the figure is closer to 50%. Most computers sold in the last 12 months are multimedia-capable. Many came equipped with modems. Suppliers of Internet access are rising like mosquitoes from the grass to meet this exploding market need.

### *Growth of the Internet*

The number of people who have access to the Internet is a matter of some debate. The commonly used figure is 20 million, but some researchers claim that 8–12 million is more accurate. The argument turns on the assumed number of users per Internet address. Some, like a user's home machine, host a single user; some others, like aol.com, host a million users.

Of those, the number who actually use the Internet is smaller. (Example: AOL customers.)

By a number of measures the Internet is experiencing exponential growth.

The number of providers of "space," or presence, on the Internet grew at a rate of 8% per week over a recent 3-week period (5400% compounded per year).

The umber of commercial entities with Internet presence has been growing at a rate of 7% per week (3300% compounded per year).

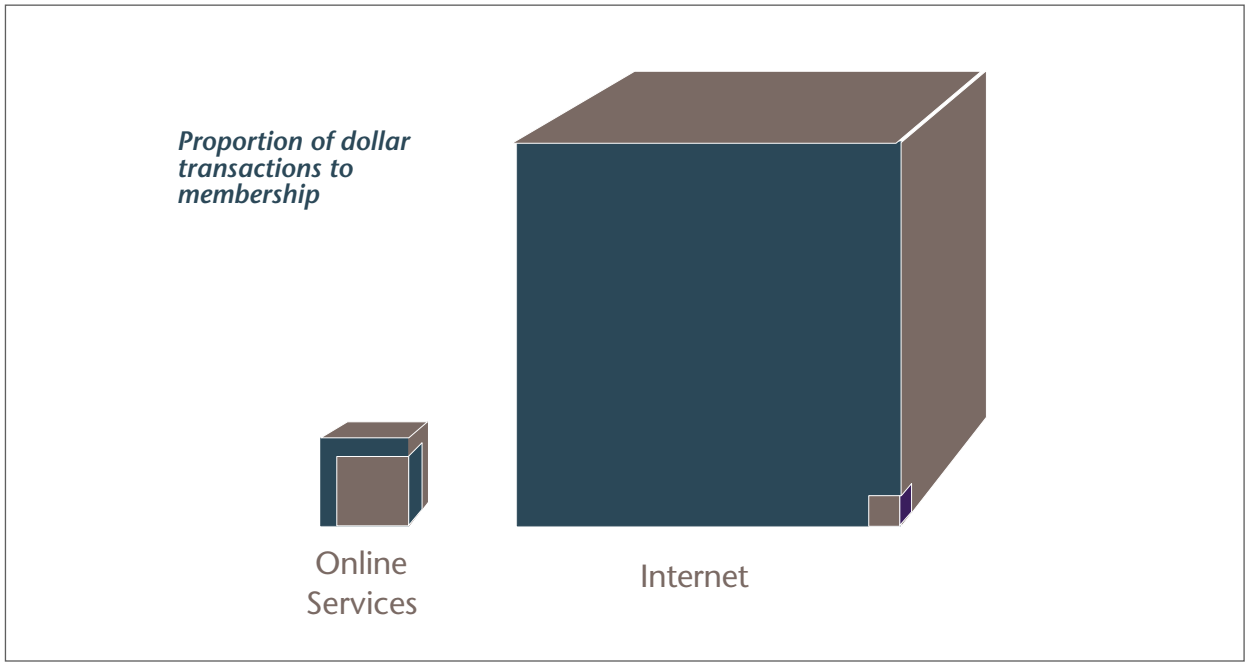



*Figure 1*

*Many more dollars are exchanged through online services with much smaller memberships than the Internet*

### *Immediacy*

What you publish on the Web, all of its millions of denizens can see and use (and potentially buy) immediately.

### *Low Cost*

Publishing on the Web has some appealing economies. There is no cost for delivery at this time. Let's repeat that: the Web offers you a way to publish content that is immediately available to 15 to 30 million potential buyers, with no cost to you for delivery. But – as you'll see – this does not mean that there are no costs. After all, the air is free, too, but flying isn't.

### *Market Intelligence*

The Web offers you an extremely effective and interactive channel for direct contact with your customers. You can provide a wealth of information and value for your business partners – going beyond simple marketing handouts and company data – to the essence and details of what you have learned about your markets and areas of expertise. You can design interactive forms that request information from visitors and funnel it straight into your company database. You can engage your customers directly in conversation about how you can provide better value for them. The market intelligence and the goodwill flowing from such activity cannot be overstated. For many of the pioneers who have already established Internet presence, this customer interaction is a large part of the value that they are reaping from their early investment.

### *The Reality*

Are these promises borne out in practice? Not yet.

First, only recently are the regulatory and technical barriers to transferring funds over the Internet being overcome. So people aren't buying much on the Internet yet. The number of dollars that changes hands using the Internet as a medium is 5 to 50 times smaller than the totals for other sorts of services, such as commercial-online, credit, or stockbroker networks, whose memberships are 5 to 50 times smaller (*Figure 1*).

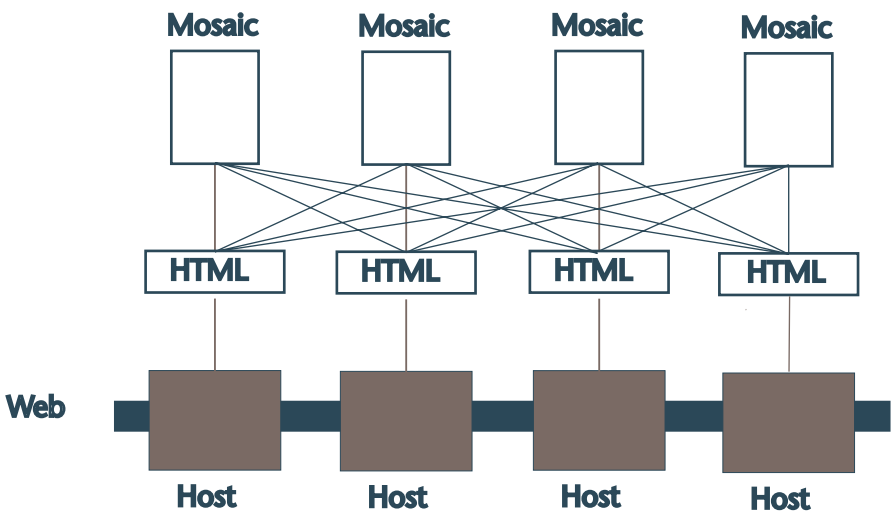



*Figure 2 The World Wide Web joins hosts who provide HTML documents rendered by Mosaic, or Mosaic-like browsers.*

Next, publishing on the Web isn't really free. The largest cost at the moment is the technical resource that must be devoted to acculturation, learning, understanding, and keeping up with a rapidly growing and shifting body of knowledge. Large uncertainties adhere to possible future charges for infrastructure services that are now either free, or simply unavailable.

One promise that is being realized is the flood of high-quality market data flowing into the organizations who are using the Internet as a new channel for direct interaction with their customers.

One group that clearly is making money off the growth of the Internet is the providers of access, presence, and consulting services.

## What Makes Internet Publishing Possible?

Two factors have come together recently that highlight the potential of the Internet as a publishing venue:

- World Wide Web (the infrastructure)
- Mosaic (the killer app)

Much of the excitement about the Internet is still about its potential, because broadly compelling content has yet to appear.

### *The Killer Infrastructure – the World Wide Web*

What makes the Web possible are three inventions recently introduced by CERN, the European nuclear research institute: a client-server architecture, a way of uniquely identifying resources anywhere on the Internet, and a simple hypertext markup language.

*Servers and clients:* A Web server is a machine on the Internet running software that enables it to respond to requests from elsewhere on the network. A client is a machine running software that issues such requests – such as Mosaic. In the Web a client and server do not maintain a connection over long periods; rather they exchange a request and a response and break off immediately.

*Universal Resource Locators*: a URL uniquely describes something on the network, for example one file on a particular machine in Italy. (In fact the URL can get more specific still, and reference a location tag within the file.) A URL also contains the name of the "protocol" that the client and server will use to deal with the details of the request, allowing Web servers and clients, as they evolve, to subsume entire existing Internet services. All Web browsers today support standard file-transfer (FTP) and Gopher requests. Some also support the protocols for email and/or Usenet news.

*Hypertext Markup Language, HTML:* This deliberately limited, simple markup language makes it easy for people not specialized in publishing technology to put their documents online. But no one wants to read a pure markup language (see sidebar); it needs to be interpreted and rendered by an application. That's where Mosaic comes in.

### *The Killer App – Mosaic*

The application that touched the fire to CERN's kindling was NCSA Mosaic, developed at the National Center for Supercomputer Applications, University of Illinois at Urbana-Champaign (called the Silicon Prairie). The guiding intelligence behind Mosaic was Marc Andreessen. The first early version seen outside of UIUC ran on Unix/X machines and was introduced in February of 1993. The general release in September 1993 included Mosaic clients for X, Macintosh, and Windows machines. The software was made available for free over the Internet.

Mosaic caused such a stir in the Internet community that it was soon noticed by the broader world. By December 1993 Mosaic had been written up in the *New York Times* business section and in the *Guardian* of London.

When demonstrated on a computer that enjoys a fast connection to the Internet, Mosaic is indeed compelling – breathtaking would not be too strong a word. The interface is simple and uncluttered, and it is child's play to skim around the world, jumping from page to page, sampling colorful and imaginative offerings from the collective mind of humankind.

### *The Killer Content – Missing So Far*

Browsing is the usual metaphor for exploring the Web. Browsing is what the current crop of Web client programs is good at. As the medium and the tools encourage random walks, they make other kinds of intellectual activities more difficult – such as investigating in a directed fashion and finding the exact bit you're looking for. Many Web pages, when

### *How to Abbreviate the World Wide Web*

Debate has raged about the proper way to abbreviate the "World Wide Web." Should it be "WWW" or "W3" or "W$^3$"?

The Gilbane Report has never shied from controversy, and so we are weighing in on this weighty topic. Each position has its positives and negatives. "WWW" is the fastest to type since you don't have to move your fingers out of place, but it is the kerner's nightmare. "W3" takes up less room on the page, but is slower to type.; how many of even the touch typists among us have mastered the row of numbers? "W3"takes up the least room and is aesthetically the most pleasing, but is virtually impossible to type and cannot be represented in most of today's email systems.

Now, it might seem that human typists should use "WWW" and publications such as this one ought to use "W$^3$". But The Gilbane Report supports standards above all. We do not need two ways of abbreviating "World Wide Web." "WWW" it shall be. (Except when we call it "the Web".)

*— David Weinberger*

*"For all the hoopla, there is still not much "there" there."*

you come upon them, have an unfinished feel (and, often, literal "Under Construction" signs). Many are innocent of any real content, being simple collections of pointers to other things. For all the hoopla, there is still not much "there" there.

However, many specialist communities are well served by the Web in its current form. Researchers worldwide in numerous fields find the Web a good match for their needs. One example of the near-instantaneous worldwide collaboration that the Web enables was seen last July when Comet Schumaker-Levy collided with Jupiter. Photographs from telescopes around the world – and far out of it – were made available on the Web sometimes within minutes of capture, along with the the researcher's initial impressions. The experience was unprecedented for the community of astronomers (and for the larger number of amateurs and interested laymen).

## Mosaic Basics

### *Services a Web "Page" Can Offer*

By taking full advantage of the generality of Universal Resource Locators and the capabilities of the most advanced client Web browsers, here are some of the things a Mosaic document can do with its HTML content:

- hypertext: jump to any point in any other Web page anywhere
- download: return to the client formatted hypertext, graphics, sound, video...
- launch local: start a local application to display returned data
- graphical navigation: *e.g.*, click on a map to zoom in
- search: search a server-local database at client's request
- fill-in forms: take structured data from client
- other: perform any action local to the server and return any result to the client

### *What HTML Can and Can't Express*

Here are the constructs that you can use in HTML 1.0 today. All Web browsers support this minimal level of formatting; some support more advanced options (such as forms) that have not yet been standardized.

- Title, header levels
- Body text
- Indented text
- Bold and italic
- Lists – bulleted, numbered, glossary-style
- Hypertext jumps and destinations
- Literal (preformatted and/or monospaced) text
- A few miscellaneous items such as author, block quote, etc.

The following rather basic formatting features are not available in HTML 1.0. Again, some browsers support some of these features today.

- Single-column formatting only
- No centered, flush-right, or justified text
- No tables
- No text flow around images

- No forms

### *The Appearance of HTML*

The author of an HTML document has little control over its final formatted appearance: many formatting choices are left to the reader to decide. Some browsers, such as MacWeb/WinWeb, provide easy-to-use graphical tools to help the end-user (*i.e.*, the document reader) customize the appearance of each and every HTML element. The user decides font, size, color, and indents for each level of header, type of list element, style of emphasized text, and so on. Other browsers, such as Netscape, allow the reader a less detailed range of choices.

## HTML at Work

```
<title>SEDS.LPL.Arizona.EDU WWW Site</title>
<a href="/seds/chapters/UASEDS/UASEDS.html"><IMG ALIGN=TOP ALT="UA SEDS LOGO" SRC="uaseds1.gif"></a> U of A SEDS Chapter Homepage<br>
 <h2>Students for the Exploration and Development of Space</H2>
<address>Lunar and Planetary Laboratory, University of Arizona, Tucson</address>
<hr>
<dl>
<dt><IMG ALT="" SRC="sedslogo-NEW.gif"> <a href="seds/seds.html"><b>SEDS National</b> Home Page</A>
<p>
<dt><IMG ALT="" SRC="sl9a.gif"> Here's the latest info on <a href="sl9/sl9.html">Comet P/Shoemaker-Levy 9</a>.
<br>
<dt><img ALT="" src="galaxicoA.gif"> A <i>galaxy</i> of <a href="other.html">images and information</a>
<br>
<dt><img ALT="" src="/nineplanets/gifs/tnp.gif"> <A HREF="/nineplanets/nineplanets/nineplanets.html">The Nine Planets</A>, a virtual tour of the Solar System
<br>
<dt><IMG ALT="" SRC="arecibicoA.gif"> The <a href="../pub/faq/astroftp.html">Astronomy FTP
List</a> <br>
<dt><dt><IMG ALT="" SRC="stationA.gif"> Student Space Action's <a href="/ssa/ssahome.html">homepage</a> and <a href="ssa/petition.html">on-line <BSpace Station</B> Petition</a>
<p>
<dt><IMG ALT="" SRC="pinkball.gif">Try out our <a href="messier/Messier.html">Messier Catalog</a>
</dl>
<hr>
<B>Check out the <A HREF="ftp://SEDS.LPL.Arizona.EDU/pub/">FTP</A> and <A
HREF="gopher://SEDS.LPL.Arizona.EDU">GOPHER</A> services at
<A HREF="gopher://SEDS.LPL.Arizona.EDU/00/About%20this%20Server"><B>SEDS.LPL.Arizona.EDU!</B></A><br>
<hr>
<ADDRESS>Maintained by the University of Arizona SEDS | seds@SEDS.LPL.Arizona.EDU</ADDRESS>
Please leave your <A href="comments.html"><I>comments</I></A> here.<br>.
```

*Figure 3 Portion of the HTML behind a Web page with hyperlinks to photos of the solar system.*

All browsers interpret HTML at display time; therefore text is formatted (wrapped) to whatever size the window happens to be.

Browsers implement different sub-and supersets of HTML 1.0, +, 2.0, and other functionality; and they differ in the strictness with which they enforce HTML syntax. Since you, the author, don't know which browser your reader will use, you have even less detailed control over how your information is formatted for the user.

*"Browsers implement different sub-and supersets of HTML 1.0, +, 2.0, and other functionality; and they differ in the strictness with which they enforce HTML syntax."*

### *Relation to SGML*

HTML as defined is an application of SGML. But initially there was no formal definition of HTML (1.0) syntax. For HTML 2.0, a complete document type definition (DTD) has been developed and is near standardization. When this DTD has been standardized, users will be able to check the validity of HTML syntax using SGML parsers. HTML client-side browsers do not incorporate SGML-compliant parsers. Available Web browsers differ in the strictness or laxness they bring to the task of validating HTML. (See "SGML" sidebar for further details.)

## Document Publishing with the Worldwide Web

There are, obviously, limitations to delivering information over the WWW.

- As noted above, the formatting capability provided by HTML is relatively crude; and you the publisher do not control the format the reader will see,
- Online monetary transactions are just now becoming possible.
- The ease of copying and republishing makes ownership of intellectual content hard to enforce, without some form of encryption.

Despite these limitations many organizations are reaping benefits from being on the Internet (in addition to the general sense of being "with it" that Internet savvy conveys).

There are three basic uses of web publishing. The first, which may exist only on a LAN and not be connected to the Internet at all, is for the internal distribution of information. Second, companies can use the Web to distribute information and materials they traditionally would have put out on paper or over the telephone. Third, the Internet represents a new channel for people in the publishing business itself.

### *Inhouse information dissemination*

A number of companies are using Web technology internally for information dissemination, corporate database applications, groupware, etc. No one knows how widespread this practice is, because for the most part the internal networks of such companies are screened behind "firewalls."

For example, Sun Microsystems, which hosts one of the largest collections of external content anywhere on the Web, is rumored to have 800 inhouse servers dedicated solely to its internal Web.

Genentech Corporation serves 2,700 inhouse users from a repository based on an internally visible Web with Mosaic browsers. It offers company information, an employee database with photographs, company news, etc.

### *Corporate publishing*

The WWW can provide a highly beneficial new channel for distributing information to your market, supplementing if not yet replacing traditional channels.

From a home page hosted by some provider of Internet access, you can distribute high-

level corporate and other overview information, and offer data sheets, white papers, and press releases.

If your end product is software the Web provides an excellent medium to convey software documentation. The technique offers the usual benefits of online publishing, along with the ability to be extremely responsive in updating the information.

Web publishing is also an excellent medium for a customer-support problem database. It enables users to find information quickly and ensures the information is up to date. Clicking on Mosaic links is a lot more satisfying to users than struggling with phone mail ("If your problem is with subsystem C of version 12 of product 54, then press 666 followed by the pound sign..."). It is also well-suited for a for marketing-hosted forum for discussions among customers.

### *For publishing companies*

The Internet is providing a new channel for existing publishers as well as spawning its own generation of Internet-specific publishers. As of October, 1994, according to a Web page that traces such magazines (URL is http://www.acns.nwu.edu/ezines/), there were already eight academic magazines, 24 independent magazines, six commercial, and one award winner edited by a 16 year old high school student in Vancouver.

Some printed magazines are online, notably *Wired, Mother Jones*, and *Boardwatch. Wired* is developing a made-for-the-Web publication called *HotWired,* which has just made its first appearance. It is not an online version of *Wired,* but a separate new magazine published in a different medium; the magazines may share some content between them.

An example of an Internet-only general-interest magazine is O'Reilly's *Global Network Navigato*r. Such a publication in this medium has unprecedented flexibility to mold itself to meet the needs of customers. The entire creative output of the Web community worldwide can be used, coopted, and continuously mined for ideas. Such a magazine, once established, becomes a magnet for externally produced content, and a channel for internally generated content.

*"Web publishing is also an excellent medium for a customer-support problem database."*

Booksellers are also going onto the Net. For example, Duthie Books in Vancouver, BC uses the net to offer browsing of inventory and electronic ordering, as well as value-added free services such as "The Reader" book reviews. The total effort invested so far is just two people for one year.

## Publishing Strategies

So now you want to publish over the Internet. How to do it? How to you write documents in HTML? How do you convert legacy documents into HTML? Isn't HTML just SGML in disguise? Won't it therefore swallow every conceivable resource

Producing HTML need not be a terribly painful process. As you investigate techniques, tools and methodologies, keep the following in mind:

- You want minimum disruption of your current publishing practices
- You need to manage the documents throughout their life cycle, including updating them and creating new documents based upon them
- You probably want maximum flexibility in both input and output to guard against the inevitable shifts and changes in your environment and market needs

The following strategies are not exclusive; some organizations may make use of more than one of them in exploiting the WWW as a publishing medium.

### *Author in HTML*

This strategy is simple in concept but hard to execute. It puts the burden on authors or content originators to become expert in HTML coding. It demands the researching and creation of an HTML house style guide.

A number of tools are available to aid in HTML authoring, ranging from simple macro packages for various text editors, to standalone editors, to fully SGML-compliant parsing editors. At one end is the use of text editors such as vi and emacs to create ASCII, leaving it up to the author to insert the HTML markup; this is known as the brute force method and it requires your authors to become experts in HTML. At the other end is HoTMetaL from SoftQuad, one of the first HTML-specific editors (available in a shareware version for free). It is a word processor that enables authors to create HTML documents and embed HTML features without actually knowing HTML.

### *Author in your current word processors and translate to HTML*

This strategy requires the least dislocation on the part of content originators, but may reduce the flexibility with which you can deal with your Web-targeted content. Translation between markup languages is rarely a 100% proposition; it resists complete automation. Even where translators exist in both directions, information is usually lost in each pass. Consistent results may require imposing unwelcome standards for input coding on the content originators when using their original word processing applications.

Look for Microsoft Word and WordPerfect to offer some form of built-in support for HTML so that you can create documents ready for the Web from within your current word processing environment.

A variation of this approach focuses on publishing rather than point-to-point translation. Interleaf's new Cyberleaf product is the first to embody this approach. It, too, assumes you have written your documents using a standard word processor (it accepts documents

### *HTML and SGML*

Thus far, documents available on the World Wide Web for browsing and display are all encoded in a "tagging language" called HTML. HTML defines a simple set of information structures (such as "P" for paragraph and "TITLE" for the document title) and establishes ways in which "markup" distinguishes those structures in a text file. HTML is good for browsing data, but is not well-suited to the wealth of activities associated with doing any more complex work.

On the other hand,HTML is an application of the international standard SGML (Standard Generalized Markup Language) that is well-suited to complex tasks. SGML's supports the specific needs — especially needs for re-use of data in automated ways — of any class of documents. It is far more general. Comparing HTML and SGML is somewhat like comparing apples and the entire category of fruit.

An architecture for SGML capability will open the networks up to all those groups that count on the semantic richness of their own markup to be able to automate a great range of production and other processes and for whom losing that richness to publish in HTML on the Web is too high a price to pay.

For example, in order to automate computerized use of information, it is useful to distinguish tasks and subtasks in maintenance manuals, part numbers and dimensions for manufacturing documents, place of origin and pronunciation in dictionaries. Markup is used to carry this weight. Each

from Word, WordPerfect, Frame and Interleaf) and translates them into HTML, including graphics as well as text. Within Cyberleaf you point-and-click to create what Interleaf calls "persistent links" — when the source document is changed in its original word processor, the web is be updated without the links breaking. This means the document can be maintained throughout its life cycle.

### *Generate HTML out of databases using custom applications*

With custom programming, you can generate valid HTML files on-the-fly from document databases. Many of the Web search services operate in this manner from databases of Web resources. In this model HTML becomes essentially a target for on-demand printing. This strategy requires the most technical sophistication and custom software of all the choices outlined here, but offers all the benefits of automation.

### *Use a Portable-Document Tool*

If a portable-documents tool such as Adobe Acrobat or No Hands Software's Common Ground is central to your publishing strategy, you can extend its use to the Web with little additional effort. All Web browsers can be configured to launch a particular application upon encountering downloaded content of a given type. This strategy requires that your end-user content consumers configure their browsers appropriately to launch, for example, Acrobat Reader or Common Ground.

A Web publishing strategy based solely on portable documents would almost certainly prove insufficiently rich when compared to alternate strategies. The documents viewed by the user, however visually rich in their own context, will appear "dead" and static from the point of view of Web interactivity. No portable-document format today offers more than half-integration with HTML's hypertext links: that is, you can launch a portable document application from HTML, but you cannot trigger an HTML link from within a portable document.

SGML application is actually a little language of markup optimized for one class of documents or circumstances or uses. SGML encoding lets us tell a "task" from a "part number" in a computer processable way.

From this point of view, HTML is just another little language. Because its real goal is display and linking, it doesn't have enough subtlety to distinguish the information structures for any particular set of documents or requirements. SGML does.

The most HTML can do is give guidance on how to format text for screen display and attach links. These are very useful capabilities for relating pieces of content across a large pool of information, but not enough for the management of that information.

So the real secret is to have both: the simplicity of HTML for consistent cross-platform browsing and display, and the richness of SGML for retrieval of meaningful chunks of information.

Architectures and mechanisms are being designed to allow them to co-exist (including SoftQuad's own Panorama, a World Wide Web browser that can read any SGML file). SGML's greater sophistication and capability means that it runs the risk of greater complexity. It will take the best work of Web software designers to harness that power without sacrificing the simplicity which has allowed so many people to become Web-capable so quickly.

*Yuri Rubinsky*

*"The growth of content on the Internet will outstrip the ability of software to tame it."*

### *Use SGML*

If SGML is important in your overall publishing strategy, it may soon become easier for you to address the potentially large Web audience with SGML content.

Any of the popular Web browsers can be set up to launch an SGML parse-and-display application upon encountering an SGML file. The problem is that such full-blown SGML parsers are not widely distributed (although they are available), so you as a publisher cannot count on the ability of a Web customer to make use of SGML.

SoftQuad and NCSA announced at the 2nd WWW Conference in in October an agreement for NCSA to distribute – free – a full SGML viewer called SoftQuad Panorama along with an upcoming beta distribution of NCSA Mosaic. (SoftQuad will also sell a more fully featured version of Panorama.)

More than a million users so far have downloaded Mosaic from NCSA. The SoftQuad-NCSA agreement points toward a time when a large Web audience may be equipped to make effective use of SGML content.

### *Where Document Management Fits In*

One way to think of the Web is as a target for publishing, on the same level as print or CD-ROM.

Document management will be useful in this sphere in the same ways it is useful in the rest of your publishing operation: for source-code control, document building, and configuration management to manage the "builds" and versions of documents. For example, your document management system may inform you that now that a price has changed somewhere on a spreadsheet, some set of published documents needs to be updated, approved, rebuilt and republished.

In short, you should consider published webs to be one more output from your document management system.

## Futures

First, expect HTML to become a more robust publishing standard. HTML+ and HTML 2.0 are being developed in parallel are jointly defining the next level of HTML support. This will include greater layout capabilities (centered text, justified text, text flowing around graphics, tables, etc.). It will also result in Mosaic documents with greater interactivity — more buttons to push, more events to trigger.

Mosaic will begin to provide ways to complete financial transactions. You will be able to sell products over the Internet.

Desktop access to the Internet — and to the wider range of online capabilities — will become commonplace; this is already on the way with support built into the next version of Microsoft Windows. Computers will feel increasingly like communication devices, not information processors.

The growth of content on the Internet will outstrip the ability of software to tame it. For network surfers, the surf will always be up. But for those who need to conduct business, there will be enough available resources to let you find what you need (if not absolutely everything there is).

The Internet will become an important publishing channel, but will cease being free and eventually will be subsumed by the Information Highway. There will be enough investment in Internet publications, however, to ensure that the Information Highway will preserve that investment.

## Risks and Costs

If your use of the Internet goes beyond simply putting information "out there," to include interaction with customers, then you must consider the cost of developing – or licensing or hiring the development of – custom software that works with your product database, billing system, customer records, etc.

If you host Web information inhouse, you will have to pay for:

- network connectivity
- dedicated machines
- staff expertise

If you rent space on the Web, you will have to pay:

- consulting costs
- design and development costs
- periodic fees to an access provider and/or a "presence" provider.

### *Licensing and/or service fees*

As the infrastructure of Web commerce becomes better defined, suppliers will emerge and start charging money for services that used to be either free (such as transport of bits) or unavailable at any price (such as secure/authenticated transaction processing and electronic cash). Some of the costs your organization already bears in the business of publishing – such as research, credit information, market intelligence, even banking – may move into the column of Internet infrastructure, as the suppliers shift their delivery mode to suit an emerging market.

### *A cultural guide to the Internet*

Quote of note from *The Economist*: "First Eastern Europe, then Russia, now the Internet. The latest place to venture from socialism to the free market is not a country at all: it is the largest computer network in the world."

The roots of the Internet go back 25 years in academic, government, and military networks, but the culture that dominates is that of the academic, specifically the computer room culture: anti-commercial sentiment, a reverence for grassroots democracy, tolerance, and anarchy.

The Internet may represent the purest form of democracy there is. All voices can be heard; the outcome of arguments is determined by consensus forged around the best logical arguments. Sometimes consensus never emerges.

The idea of "free" permeates the Internet. (As *Wired* magazine keeps reminding us, information wants to be free.) Anything that looks capitalist is suspect.

As a result, to succeed in the Internet, you must understand the culture and bow to it. Make sure that your materials are created (or at least edited) by people who have hung out on the WWW because they like it.

At least for now. Although the Internet culture is hostile to the idea of a volume-based fee for common carriage, such a charge is likely to come, as the NSF-owned backbone transitions to ownership by commercial interests. But until then, think of the Internet as making communications "emailish" — fast, friendly, instantaneous. And wearing Birkenstocks wouldn't hurt either.

*Keith Dawson*

### *Users' costs*

The user's network connectivity is assumed, by definition of the target market.

Today bits move over the Internet at no charge to the end user. This is almost certain to change – the NSF-owned backbone is going away by mid-1995. See the brief article "Pricing Internet" in Scientific American, 10/94, subtitled "Tolls may prevent traffic jam on the data super-highway." It reports on researchers employing simulation and modeling to determine the impact of various bucks-for-bits strategies; the finding is that some kind of fee for common carriage could ease or prevent traffic bottlenecks that would otherwise occur.

Depending on how the infrastructure services shake out and how the pricing models develop, users may bear some of these costs.

## Conclusions and Recommendations

Set someone to looking into the Internet and the Web, if you don't already have people doing it.

Buy some consulting help if necessary to educate management on the possibilities that Web publishing opens up for your company.

Look around the Web at what others are doing.

Follow the progress on issues that matter to your business – *e.g.*, the development of eCash and NetCash if you plan to be selling online.

Investigate the options to establishing a Web presence for your company. How much service and consulting will you need? Full-service or bare-bones? Can you host Web information from your own corporate machines or will you need to rent space?

## Sources

Send an email request to ptm@world.std.com and we will reply with an HTML file containing URLs for all of the companies, services, and organizations mentioned in this article.

**Keith Dawson**

# Industry News

## Instant Trends

If you were to project trends based solely on the news covered below, you would notice the following "trends":

1. Continuing fragmentation in the electronic distribution market. There is no coalescence on a single viewing format.
2. The "everyone-but-Microsoft" gangs continue to form as standards are being marketed sometimes seemingly only to give users a reason to avoid going with Microsoft.
3. Companies with like names buy one another. Intergraph has bought InterCAP and Caere has bought Calera.

## Lotus Licenses Common Ground for Notes Application

Lotus Development Corp. has licensed No Hands Software's Common Ground portable document viewing technology for integration into a Lotus Notes application for Compaq Computer to be hosted on AT&T Network Notes. InfoPaq, the name of this pilot Compaq program, will be used to electronically distribute sales, marketing and technical information to its marketing partners and corporate customers via AT&T Network Notes.

## Simon & Schuster Goes Digital

Simon & Schuster has announced a new imprint, Macmillan Digital USA, that will produce and distribute electronic books on CD-ROM and the Internet. The books will be either reference volumes or about computers. The new imprint will include Frommer's Travel Guide in American Express's upcoming online travel service, which will make available travel information on 160 major cities in the U.S. and Europe

## OpenDoc Does Windows, Windows, Windows

Novell, IBM and Apple have reaffirmed their commitment to making OpenDoc component software available on all the major operating systems. Novell has committed to developing three versions of OpenDoc for Windows (for NT, Windows 95 and Windows 3.1), where it will compete directly with Microsoft's OLE standard.

In addition, Component Integration Laboratories and Taligent (The Apple/IBM joint venture) demonstrated "the first stages of their joint commitment" (presumably this is something less than shipping software) to interoperability between OpenDoc and Taligent.

## Microsoft Word for Internet Authoring

Microsoft has announced Microsoft Internet Assistant for Microsoft Word 6.0 for Windows. Bill Gates himself demonstrated the tool at COMDEX. It's a free add-on that allows users to generate HTML files from Word; HTML is the format that Mosaic, the Internet browser, expects. (See all of this issue.) The add-on also enables users to create hypertext links among Word documents.

## OMG & CI Labs Join Forces

Component Laboratories, Inc. and the Object Management Group will work together to get the OpenDoc standard to conform to the OMG's Common Object Request Broker (CORBA) specifications 1.2 and 2.0. CI's founding sponsors are Apple, IBM, WordPerfect and the Novell Applications Group. An alliance of OpenDoc and CORBA would further increase the reach of the everyone-but-Microsoft attempt to forestall OLE's imminent adoption as the *de facto* standard.

## Oracle Makes Moves In Document Management On Many Fronts

The sleeping giant of the document management industry is waking — and is making announcements.

Oracle has introduced Oracle Documents, their document management offering. It is based on Oracle Media Server and can, according to Oracle, integrate structured data from database tables with text and multimedia data. It consists of a suite of application services and "electronic agents."

A flurry of announcements has followed:

- Oracle is licensing Hewlett-Packard's WorkManager workflow engine technology, integrating it into Oracle Documents.
- Oracle 7 will be available as a strategic database for ViewStar's client/server business process automation software.
- Odesta is shipping its document management system (ODMS) integrated with the Oracle Documents architecture, allowing Oracle users to plug Odesta into their system. In a not-quite-unrelated announcement, Oracle Government said that it is offering the Odesta system as part of an integrated document automation solution for government, education and aerospace and defense

## Interleaf Cyberleaf for Internet Publishing

Interleaf's new Cyberleaf product enables organizations to produce Internet webs by using their current word processors. Cyberleaf transforms the files to HTML and inserts hypertext links. It also imports graphics and will present them as thumbnails or full images. Interleaf claims a key benefit is the ability to revise documents in their originating word processor and update them for the Web without losing any of the hypertext links.

## Odesta Embeds Verity's Text Retrieval in ODMS

Odesta Systems Corporation will put Verity, Inc.'s TOPIC search technology into its ODMS Software product line.

## Wang Files Suit, Wings FileNet

Wang Laboratories, Inc. has filed a suit alleging that FileNet is infringing five of its patents. The suit doesn't say which products are at issue.

## Novell Forms GroupWare Division

Novell has formed the Novell GroupWare Division, responsible for providing, well, groupware. The current product line of the division includes SoftSolutions for document management, InForms for electronic forms, and two messaging products (GroupWise and MHS).

The division's first order of business — on the day of the announcement of its creation — was to roll out Novell's "long-term groupware vision," the Collaborative Computing Environment (CCE), an environment in which customer groupware solutions can be developed.

## EBT Signs Japanese Distribution Deal

Gakken Co. Ltd., one of the largest publishing firms in Japan, will resell Electronic Book Technologies, Inc.'s Dyna-family of products in Japan. EBT has opened a Technical Center in Japan as well.

## XSoft Enhances InContext & Adds Acrobat Support

Xerox's XSoft division has announced version 2.0 of XSoft InContext, its SGML authoring program for Windows. The new version provides greater control over document formatting, and will include a copy of Adobe Acrobat Reader so that users can view and print any document in Adobe's PDF format.

## Fulcrum Adds Full Text to PowerBuilder

Fulcrum Technologies, Inc. has announced a toolkit that will enable developers to incorporate full text retrieval into the applications they build with PowerBuilder.

In a move potentially more important to its stockholders, Powersoft was acquired by Sybase, making it the seventh largest software company.

## Dataware Technologies Releases Client/Server Versions of BRS/Search

Dataware Technologies, Inc., announced the release of BRS/SEARCH Client/Server 1.1 and a new Windows client interface. BRS/SEARCH is its full-text, online information management and retrieval system designed to manage large collections of unstructured information.

## InTEXT Announces CD-ROM Tools

InTEXT Systems has announced CD-ROM SmartPublisher, a software development toolkit for developing information bases put on CD-ROM. The software provides search, retrieval and summarization tools that can massage an information base so that it can be published on a CD-ROM.

## Caere buys Calera and Intergraph buys InterCAP

In a move welcome by those with poor memories, Caere is buying Calera, and Intergraph is buying InterCAP.

Caere is acquiring Calera Recognition Systems, Inc., through a merger transaction. The companies hope that Caere's strengths in the shrink-wrapped market and Calera's strength with the VAR channel will result in very strong presence in the OCR market. They also plan to "address the emerging market of document management."

Intergraph Corporation is acquiring InterCAP Graphics Systems, Inc. InterCAP makes technical illustration software. Intergraph makes CAD workstations and software, InterCAP will become a wholly owned subsidiary.

## Frame and Saros Announce Deal

Frame Technology Corporation is licensing Saros Corporation's Mezzanine library services software for incorporation into Frame's product line. They will also cooperate on major accounts and on developing VAR and system integrator relationships. Frame previously had acquired Curo, a document management application that is also based on Saros technology.

## Vendors Support DEN

Documentum, Kodak, IDI, Oracle, PC DOCS and Verity have joined Novell and Xerox in supporting the DEN (Document Enabled Networking) document management framework. DEN joins Shamrock and ODMA as Everyone-But-Microsoft attempts to define a common document management interface to enable divergent applications to work together.

## Sherpa to Ship PDM

Sherpa Corporation will ship its product data management system running on the Oracle7 database in December.

## Verity Provides "Information Agents" for the World Wide Web

Verity's TOPIC Information Server for the World Wide Web will combine information agent technology with Web document services. The agents can search for information across a network using various techniques, while "watcher agents" keep an eye out for new documents that match a user's profile. The Server will also provide document conversion and security features.

## Documentum Raises $4.5M to Fund Growth

Documentum, Inc., has completed a third round of financing, totaling $4.5M. The brings the total capital invested in the company to $13.5M. The funds will be used to grow sales, support and service, and research and development. In addition, the company has opened new offices in the United States, France and Germany.

## Buzzer of the Month

This month's award for the most buzz-word-compliant blurb in a press release goes to Component Integration Laboratories, Inc., announcing the Novell Application Group's plan to develop three Windows versions of OpenDoc, "to support the needs of developers, end users and corporate decision markers whose enterprise commitments include plans to support a heterogeneous Windows environment. Novell., Inc. also committed to delivering OpenDoc interoperability in all future Microsoft Windows operating systems." This was supplemented by a Novell announcement that this software will "take full advantage of the more robust, 32-bit operating system, utilizing advanced features including multitasking, multithreading, a flat memory model and a document-centric user interface."

# People News

**Ed Koepfler** , formerly of SSA,has joined Interleaf, Inc. as president and CEO. In other Interleaf news, **Bill Freeland**, former manager of WorldView, is RR Donnelley's Manager of Advanced Technologies; and **David Weinberger**, former Marketing Fellow, has founded Evident Marketing, Inc., a strategic marketing firm (and is acting editor of *The Gilbane Report)*. **Linda Burman** has left SoftQuad to become a marketing consultant. **Mike Maziar**ka has joined Xyvision as Director of PDM Product Management. Mike comes from the Datalogics division of Frame. **Pam Gennusa** has been named President of SGML Open and **Kent Summers** has been given her former position as chief marketing officer.

# 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.

**PTM Document Management & Electronic Delivery Seminars**. November 28-29, London, UK. These two day seminars are conducted Gilbane Report staff and are managed by Technology Appraisals. Call +44 81 893 3986 or (617) 576-5700, Fax +44 81 744 1149 or (617) 576-5708.

**Document Computing**. November 30-December 1, London, UK. This two day conference focuses on document computing technology. Produced by Technology Appraisals. Call +44 81 893 3986, Fax +44 81 744 1149.

**CALS Expo '94**. December 5-8, Long Beach, CA. The annual expo and conference covering CALS activity in the U.S. and internationally. Heavy defense industry emphasis. Call (202) 775-1440, Fax (202) 775-1309.

**Documation France Conference & Workshops.** December 6-8, Paris, France. Sponsored by Techno-Forum SARL, These two day seminars are moderated by Frank Gilbane, Yves Stern, and Guy Fermon and cover compound document management and electronic delivery technology and trends. Call +33 1 43 48 57 92 or (617) 576-5700, Fax +33 1 43 48 55 43 or (617) 576-5708.

**PTM Document Management & Electronic Delivery Seminars**. January 23-24, Amsterdam, The Netherlands. These two day seminars are conducted by Gilbane Report staff and are managed by TA Ltd. Call +44 81 893 3986 or (617) 576-5700, Fax +44 81 744 1149 or (617) 576-5708.

**Documation '95**. March 7-9, Long Beach, CA. The international conference and exposition of the year for the document management and document computing industry. Covering all aspects of enterprise document management applications. Co-sponsored by PTM, The Gilbane Report, the GCA, and the GCA Research Institute. Call (703) 519-8160 or (617) 576-5700, Fax (703) 548-2867, or (617) 576-5708.

**Seybold Seminars '95**. March 27-30, Boston, MA. The annual conference where the publishing technology elite gather. Focus is on pre-press, color, newspaper, and magazine applications with some corporate application coverage. Call (415) 578-6990, Fax (415) 525-0183.

**AIIM**. April 10-13, San Francisco, CA. The large annual storage and retrieval and document imaging trade show. Lots of hardware. Call (301) 587-8202, Fax (301) 587-2711.

**SGML Europe**. May 16-19, Gmunden, Austria. The annual European gathering of SGML experts and novices. Call (703) 519-8160 or +44 0793 512515, Fax (703) 548-2867, or +44 0793 512516.

**On-Demand Printing**. July 27-29, New York City, NY. The only commercial conference and expo devoted to on-demand printing technology and applications. Sponsored by CAPV. Call (617) 837-7200, Fax (617) 837-8856.

# Topics Covered In Previous Issues

**Vol. 1, No. 1.**

**What The Report Will Cover & Why** — An Introduction To "Open Document Systems", And A Description Of The Report's Objectives.

**Imaging, Document & Information Management Systems** — What's The Difference, And How Do You Know What You Need?

**Vol. 1, No. 2.**

**SGML Open** — Why SGML And Why A Consortium?

**Document Query Languages** — Why Is It So Hard To Ask A Simple Question?

**Vol. 1, No. 3.**

**Document Management & Databases** — What's The Relationship?

**Vol. 1, No. 4.**

**Electronic Delivery** — What Are The Implementation Issues For Corporate Applications?

**Vol. 1, No. 5.**

**Multimedia Rights & Wrongs** — What IS Managers Should Know About Copyrights In The Age Of Multimedia.

**Vol. 1, No. 6.**

**Document-Centered Interfaces & Object-Oriented Programming** — How Will They Affect You?

**Vol. 2, No. 1.**

**State Of Wisconsin Legislature TEXT 2000** — Reengineering For Document Management.

**Vol. 2, No. 2.**

**Document Management Industry Update** — Documation '94 & Other Spring Industry Events.

**Vol. 2, No. 3.**

**Document Formatting Interchange** — Why Don't We Have A Solution?

# Topics To Be Covered In Upcoming Issues

**Vol. 2, No 5.**

**Computer Graphics Metafile — SGML for Graphics?** — What Is This Graphic Interchange Standard & When Should Companies Use It?

## Order Form

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

U.S.A.: $225 Canada: $232 Foreign: $242

Additional copies and site licenses are available at reduced rates. Call for information.

Please send me additional information on:
o Strategic Consulting Services
o Documation '95
o Document Management & Electronic Delivery Seminars

o My check for $ _____ is enclosed o Please bill me

o Please charge my credit card o MasterCard o Visa o 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_____

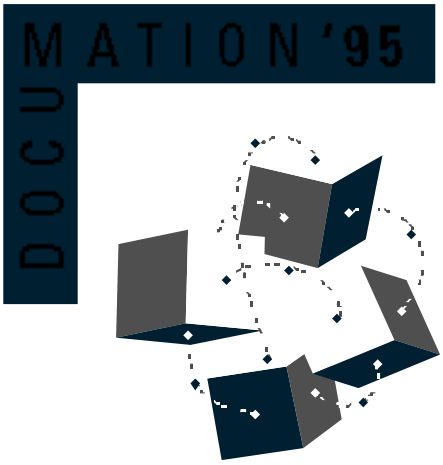
Address_____

City _____ State_____ Zip _____ Country _____

Telephone _____ Fax _____ Email _____

**Mail or fax this form to: Publishing Technology Management, Inc.**
**University Place, 124 Mt. Auburn St., Suite 200N , Cambridge, MA 02138**
**Fax: (617) 837-8856 • To order by phone call: (617) 837-7200**

# Mark Your Calendar!




*March 7 - 9, 1995 — Long Beach, CA*
The program brochure included with this issue has every- thing you need to plan for the Document Management industry's big event of the year.

*January 23 - 24, 1995 — Amsterdam*
Document Management & Electronic Delivery Seminars.

*April 3 - 4, 1995 — London*
Document Management & Electronic Delivery Seminars.

*June 27 - 29, 1995 — New York*
On-Demand Printing Conference & Trade Show.

## Call, Fax, Or Email Us To Find Out More About Events & Companies Mentioned In This Issue



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., University Place, 124 Mt. Auburn Street, Suite 200N, Cambridge, MA 02138, or ptm@world.std.com or APPLELINK:PTM

ISSN 1067-8719