

New Millennium Publishing
Editorial and Content Solutions

Publishing in the New Electronic Marketplace

Fresh Challenges, Fresh Opportunities

Bill Trippe, New Millennium Publishing

www.nmpub.com

Introduction

Are we at a New Juncture?

About this Paper

People are too often tempted to describe each new advance in technology as a watershed. Had all the breakthroughs been for real, we would have had thousands already this decade. But, the Internet, broadly speaking, is transforming businesses, perhaps none more radically than publishing. From book buying at Amazon.com to breaking news at CNN.com to boutique publications catering to the narrowest tastes, the Internet has blossomed into a large, instantly accessible marketplace for publishing.

Indeed, the recent decade has seen publishers, along with most other businesses, adapt to a new business model: highly targeted development of products to increasingly demanding and specific markets. The days of producing a single, monolithic publication are over; on the print side, short-run and on-demand are starting to dominate. Perhaps the Internet arrived exactly on time for publishers; it is a wonderful medium for targeted marketing, and there is no such thing as too much inventory.

Yet the challenge for publishers may well lie in the state of today's Internet and related publishing technology—it is a blizzard of constantly changing technology, standards and approaches. Many publishers have kept pace with HTML, but what of XML? Is CD-ROM giving way to DVD? The publisher who labored to bring a consumer CD to market three years ago now may now find a bigger opportunity on the Web.

This paper offers a brief exploration of this challenge for publishers, and then looks at several emerging approaches.

About the Author

Bill Trippe is President of New Millennium Publishing, a consulting practice focused on content management and editorial systems development. Clients include educational, professional, and reference publishers, as well as selected vendors and service organizations that market to the publishing business. Before forming New Millennium, Trippe held senior management and technical positions on both the publisher and vendor side of the business, including Houghton Mifflin Company, Ziff Davis, Xyvision, and Inso Corporation. He can be reached at 877-396-8466 or btrippe@nmpub.com.

The New Publishing Challenge

A Bigger, More Elusive Market

Electronic Publishing Then

As recently as five years ago, publishers were making their first forays into electronic publishing. As revenues for print products flattened and the cost of production and manufacturing continued to rise, publishers looked to electronic products as both incremental revenue and a means for cheaper distribution. What did this world look like?

- 1) First-generation CD-ROM publishing offered rudimentary search engines and rough-and-ready document formatting, most often leaving text in typewriter-style format and providing little in the way of graphic presentation or WYSIWYG output.
- 2) Online services that catered to the professional market emphasized depth and breadth of content over the quality of the user's experience. Even the pre-Web versions of services such as AOL lacked high-quality, high-fidelity presentation of commercial publications.

The Publishing Process and Tools Then

Publishers faced a challenge even then. Some had large repositories of published information, often in a proprietary format most suitable for print production. They needed processes and products that would automate electronic publishing in much of the same way they had successfully automated print publishing.

- Fortunately, while the need for electronic output might have been new, early CD-ROM and online formats were often not very complicated. Publishers who had developed workflow and production tools for large print publications adapted well to producing what amounted to a small number of simple electronic outputs for CD-ROM and online services.

If a second online service wanted a version of our database in a slightly different format, it was relatively simple to create a filter for that second format. We only needed to factor in this increased production cost. Indeed, the online services were eager to augment their content offerings, and were often willing to assume the cost.

Carol Ellerbeck, Web Development Manager at Lycos

- So, even a large publisher's need for electronic publishing, while extensive, often only amounted to a small number of processes and outputs.

Electronic Publishing Now

The advent of the Web, and the increasing power of personal computers, have had some surprising effects on publishers:

- Some of the traditional sources of licensing revenue, such as online services, have splintered.
- The retail CD-ROM market, originally a strong outlet for electronic publishing, is moving more toward media-intensive entertainment titles.
- The professional CD-ROM and electronic publishing market has grown substantially as libraries, corporations, and organizations devote more of their new spending to electronic products.
- The biggest potential revenue source on the Web is advertising. However, Web advertising is not a good fit for subscription-based publishing models, and has yet to generate significant revenues.
- Customers are expecting increasingly frequent updates of all kinds of information, forcing publishers to migrate to short-run printing and just-in-time print manufacturing while also shortening their editorial cycles.

The Result: A Bigger, More Elusive Market

The contemporary publisher is in the business of creating content to sell in many forms.

Dennis La Beau, Technical Director for the publisher *Facts on File*, said “We publish the *Facts On File World News Digest* in print, CD-ROM, soon on the Web, and in a variety of other electronic versions for delivery via various vendors and resellers of data to the library marketplace.”

Carswell does short-run printing, demand printing, CD-ROM, Intranet and Internet (Web) as well as “push” services via e-mail. Customers want information in many different forms. Each media format has its benefits and its drawbacks. For example, CDs are portable, fairly responsive and contain lots of information. But, they aren't timely in comparison to online and put pressure on internal IT resources if they are to be networked. Web services are easy to use, easily updated, can be personalized, but performance is an issue due to the vagaries of the Internet.

Huw Morgan, Chief Technology Office, Carswell Publishing

It has become clear that the electronic publishing model of a few years ago, in which a handful of formats are easily teased from even a proprietary repository, has given way to a new one. A new **electronic publishing model** in which the publisher must preside over a flexible database that not only can continuously generate **many different forms of the content**, but also **generate subtly or significantly different versions of the content**.

Meeting the Challenge

“Capture Once, Use Multiple Times”

The Long Road to Here

The past several years have been a period of investment and development for publishers. The goal: To aggregate content in an element or document-based repository that can be used in generating multiple print and electronic products. Having structured repositories give publishers the flexibility to market highly targeted products.

The proper solution is to establish an environment where the structure of the data is captured in a granularity that supports all possible outputs and reuses

Dennis La Beau, Technical Director, *Facts on File*

Publishers can structure their data in many ways: using SGML/XML, relational databases, strict use of styling conventions. As *Facts on File*'s La Beau points out, “The importance of identifying structure as early as possible is the key to lowering costs of production by reducing the amount of manual intervention. The trick is to provide inexpensive tools to make it possible to move the ‘tagging’ as close as possible to the author.”

SGML, Standard Generalized Markup Language, now a mature technology, offers publishers the ability to capture data at an extremely fine grain, and to amplify the data elements with metadata. For example, a medical publisher could store detailed information about a given drug, say, its dosage, then indicate it is available in tablet and capsule form. Given this level of granularity, the publisher could then produce multiple published views of this same information. Not every publisher has adopted SGML, which brings with it certain thresholds of cost and technical complexity. Even publishers that do use SGML do not use it for all content, particularly simple and one-off publications.

And what of XML?

The eXtensible Markup Language has quickly emerged as a significant new technology for the Web. A “slimmed-down” version of SGML, XML will be supported in Microsoft's Internet Explorer 5.0. Along with product adoption, efforts are underway to bring XML into day-to-day use for publishers, including efforts such as the Information and Content Exchange initiative backed by the Tribune Company, CNET, Preview Travel and others.

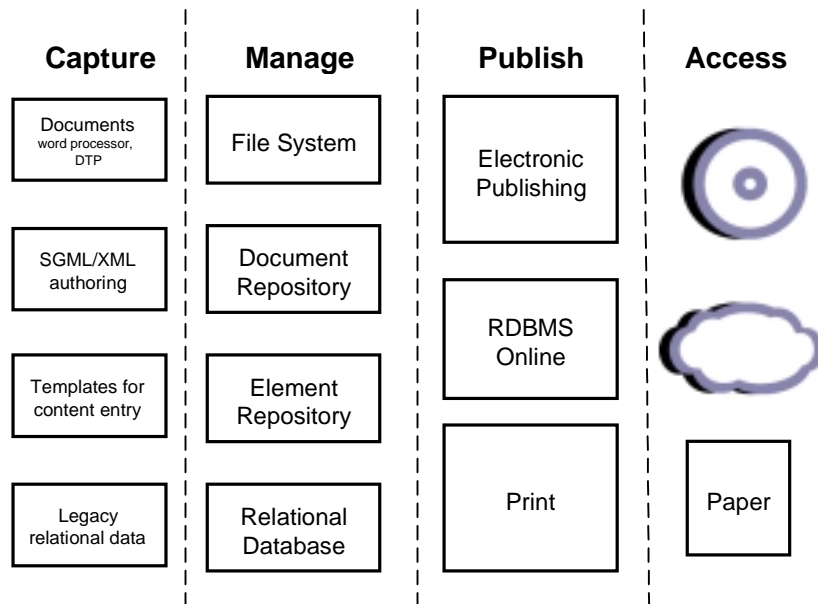
XML support should be an essential criterion in evaluating new publishing technology.

Despite the costs and overheads of structured data repositories, the need for structured content will continue to grow. Customers are demanding more tailored products, and publishers move toward automated production. Tools, such as XML, that lower capture and output costs of structured data will shape the new electronic marketplace.

New Technologies, New Processes

Along with the adoption of standards such as SGML/XML, publishers have made significant investments in technology and tools. Publishers have invested in authoring tools, text and media databases, and a variety of mechanisms for publishing and output.

A typical production environment looks like this:



Currently most publishers have multiple production processes tailored around specific source data or output media. Emerging publishing technologies are aiming to integrate the publisher's production environment.

Key Requirements of New Production Systems

- They must support more dynamic publishing: automated production, tailored products and frequent updates.
- They must (potentially) support print, Web and CD-ROM distribution.
- They must be a good fit with the technical infrastructure and workflow of current and planned publishing operation.
- They must not create a large programming, database administration or system administration burden.
- They must support emerging document models such as XML.
- They must provide support for standard authoring environments, ranging from word processing to structured editing tools.
- They must support integrated searching.

Emerging Approaches to the Challenge

New technologies for Web and automated publishing

Over the last two years, a number of products have reached the market to support Web-based and automated publishing. These include:

- Extensions to mature technologies that now add, for example, Web distribution to a product that already supported CD-ROM output.
- Web-focused technologies featuring on-the-fly output of HTML pages from a database.
- Tailored push technologies for article-length material.
- Authoring and creative tools for media-rich sites.

Two main approaches for commercial publishers

For the commercial publisher interested in the development and support of medium-to-large-size publishing operations, one useful way to divide the market is in two:

A set of technologies I will call ***RDBMS-Backed Repositories***. These include StoryServer from Vignette, Cold Fusion from Allaire, and Active Server Pages from Microsoft.

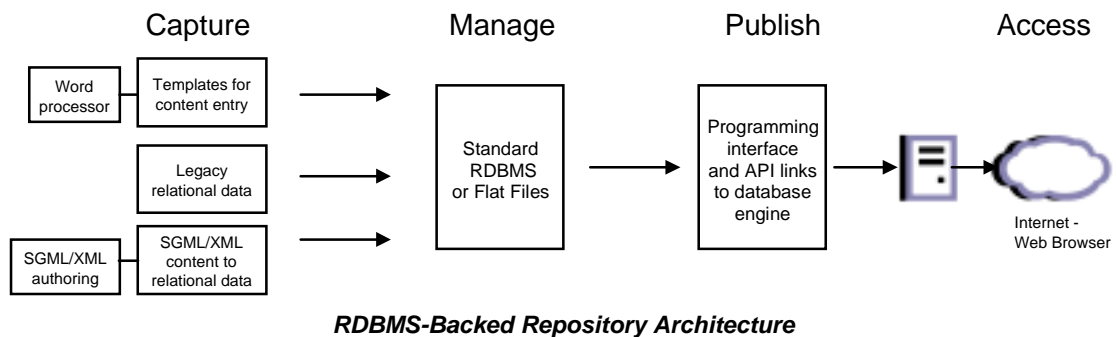
A second set of technologies I will call ***Electronic Publishing Engines***. These include DynaText from Inso, Folio from Open Market, and INSIGHT from Enigma.

Both types of products are in extensive use today at major publishing operations and are solid technologies. They do serve very different needs, however, and should be evaluated in terms of organizational, technical and business requirements.

RDBMS-backed repositories serve the “news” customer well

RDBMS-backed repositories have gained rapid adoption in the Web environment.

- As the simplified block diagram below shows, these systems produce HTML pages typically by extracting content from an RDBMS backend and converting it on-the-fly to a fully realized HTML page. Some or even all of the HTML markup can be stored in the database, or created on-the-fly. Many organizations are using this model to create different, sometimes personalized, versions of HTML pages.
- As the diagram shows, text capture and editorial workflow is somewhat limited in this model. Most typically, editorial updates are done from browser-based forms, though other formats can be integrated with some programming effort.

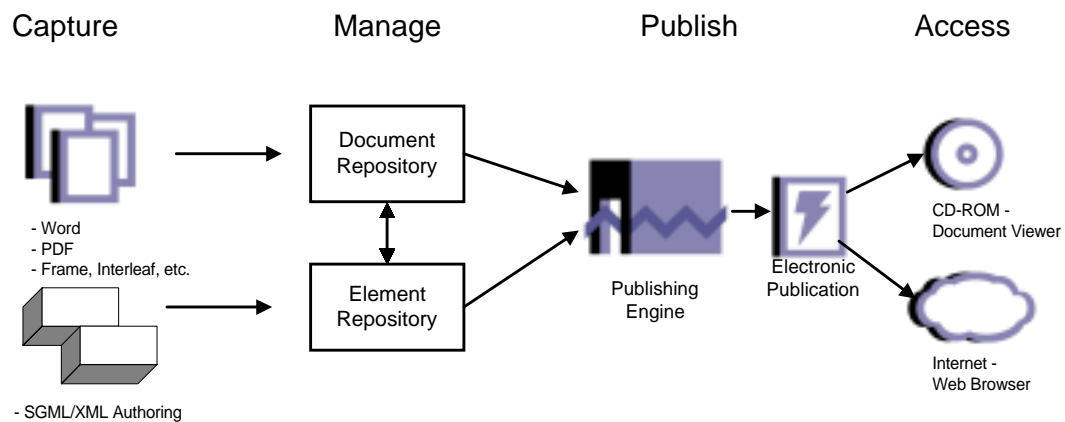


- Publishing output is typically to the Web only.

Electronic publishing solutions better support the complex document

Electronic publishing engines are a mature technology with widespread adoption on the Web and in CD-ROM applications.

- As the simplified block diagram below shows, these systems produce HTML pages typically by extracting content from the electronic publishing engine and converting it on-the-fly to a fully realized HTML page. The engine typically stores data as SGML, XML or a proprietary format.
- SGML storage offers the benefit of supporting a more standard document model, which makes the process of creating HTML (and eventually XML) easier.
- As the diagram shows, the repository in this model is usually a document management system or an element management system.



Electronic Publishing Engine Architecture

- Text capture and editorial workflow is much more flexible in this model, and is often supported by rich workflow and applications control capabilities built into the repository.
- Publishing output is typically to CD-ROM, the Web and print.

Evaluating Different Offerings

Source Formats, Output Flexibility, Integration Are Key

Which is for you?

The two approaches discussed here solve different publishing problems, mainly differentiated by the size and complexity of the material you publish.

This table illustrates how *RDBMS-Backed Repositories* and *Electronic Publishing Engines* handled a range of common publishing requirements.

Requirement	RDBMS-Backed Repository	Electronic Publishing Engine
Document Length	Shorter, article length	Long, multi-volume
Document Complexity	Relatively simple documents	Complex or especially hierarchical documents
Output Types	Primarily Web	Mixture of print, Web, optical media
Real-time Content	Designed for online data and delivery	Designed for delivering distributed databases. New update capabilities make content dynamic*
Authoring Support	Best supports forms interface	Works well with all standard authoring platforms
Document Shelf Life	Best for quick creation, shorter life	Long shelf life and multiple versions
Time to Market	New products typically require upfront programming	Supports short development cycles (typically)

*New developments: While electronic publishing tools have generally been well suited to multiple output media, standard authoring tools and document/element repositories, the inability to deliver the most up-to-date content has been its weak point.

Now electronic publishing vendors are aggressively moving to provide dynamic update capabilities. Dynamic update capabilities significantly enhance the end-user experience and positively affect publishers' subscription models. At least one such solution is commercially available, in Enigma's INSIGHT.

A closer look at Electronic Publishing Engines

This table offers a comparative evaluation of three electronic publishing engines: Inso Corporation's DynaText, OpenMarket's Folio, and Enigma's INSIGHT.

Note: Every attempt has been to compare the latest commercially available versions of the products/family of products.

	Folio Views	INSO Dynatext	Enigma's INSIGHT
Product Summary	Electronic Publishing software for the creation of a navigable document collection. Distribution on CD-ROM, intranet, internet		
Most common usage	Reference publications	Technical Documentation	Reference Publications and Technical Documentation
Search Index / Database Features	<ul style="list-style-type: none"> • Very fast engine • Good full-text search features • Limited topic-specific searching • No relational numeric or date searching 	<ul style="list-style-type: none"> • Relatively slower engine • Good full-text search features • Topic-specific searching supported • No drop-down search fields • No relational numeric or date searching 	<ul style="list-style-type: none"> • Fast engine • Good full-text search features • Topic-specific searching supported • Drop-down search fields supported • Relational numeric or date searching supported
End-user Navigation and Markings	<ul style="list-style-type: none"> • Strong feature set • Good end-user markings 	<ul style="list-style-type: none"> • Good feature set • Weak annotation and customization capabilities 	<ul style="list-style-type: none"> • Strong feature set • Good search result sorting
Presentation, Print, Export	<ul style="list-style-type: none"> • Rewrap to window but no page fidelity • Strong printing - record oriented • Weak compound document display • Weak multi-document view 	<ul style="list-style-type: none"> • Rewrap to window but no page fidelity • Good printing - element oriented • Weak compound document display • Weak multi-document view 	<ul style="list-style-type: none"> • Page fidelity but no rewrap to window • Strong printing - document oriented • Strong document exporting • Good compound document display • Good multi-document view
End-user application creation	<ul style="list-style-type: none"> • Good default end-user interface • Weak customization of end-user interface 	<ul style="list-style-type: none"> • Weak default end-user interface • Weak customization of end-user interface 	<ul style="list-style-type: none"> • Strong default end-user interface • Strong customization of end-user interface

	Folio Views	INSO Dynatext	Enigma's INSIGHT
Dynamic Updates	<ul style="list-style-type: none"> • Can add new content to Infobase on CD-ROM by Appending • Cannot update "delivered intranets" 	<ul style="list-style-type: none"> • Cannot update CD-ROM • Cannot update "delivered intranets" 	<ul style="list-style-type: none"> • Can add to, modify, or delete content on CD-ROM with Dynamic Updates • Wizard-driven updating and distributing of updates • Push or pull delivery mechanism • Can update "delivered intranets"
Source Data	<ul style="list-style-type: none"> • Good MS-Word/RTF support • Weak Frame support • No PDF support • No SGML support • Strong HTML support • Strong common image support • No intelligent graphics support 	<ul style="list-style-type: none"> • Convert MS-Word/RTF to SGML • Good Frame support • No PDF support • Strong SGML support • Good HTML support • Strong common image support • Strong intelligent graphics support 	<ul style="list-style-type: none"> • Strong MS-Word/RTF support • Weak Frame support • Good PDF support • Good SGML support • Good HTML support • Strong common image support • Strong intelligent graphics support
Fit with Infrastructure	<ul style="list-style-type: none"> • Thin client web access and CD-ROM client • No integration with data repositories • No XML support 	<ul style="list-style-type: none"> • Thin client web access and CD-ROM client • Tight integration with INSO Dynabase repository • Strong support of XML 	<ul style="list-style-type: none"> • Thin client web access and CDROM client • Tight integration with FileNet, PC DOCS and Documentum repositories • Strong support of XML