

XML for Content Management

Exploiting the Full Value of Enterprise Information

Abstract

In today's highly competitive market, companies do not implement content management systems without thinking carefully about the role XML should play. For these organizations looking towards XML, it is imperative to understand what this language is, how it works, and the specific business benefits it can bring. This article provides an overview of XML, its ever-expanding role in management of content, and describes real world examples of how today's leading organizations are exploiting XML to drive efficiencies and increased profitability.

What is XML

Enterprises today are information-centric, and they are finding themselves overwhelmed by an exponential growth in business content that is ever changing, complex, and multi-structured. With information requirements constantly evolving, organizations are finding it difficult to take advantage of the knowledge locked within their content, limiting business agility and flexibility. To address these issues, Extensible Markup Language (XML) is fast becoming the tool that organizations are using to structure their content, enabling process automation and radically altering traditional workplace content management applications.

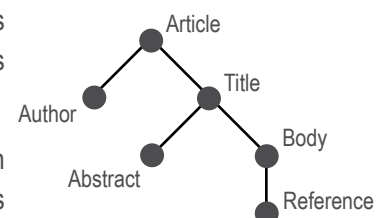
The history of XML is deeply integrated with the history of electronic text processing, document processing, and publishing. It is a subset of the Standard Generalized Markup Language (SGML) which was created over 10 years ago to facilitate the representation of data over the Internet. XML's phenomenal success is due to its acceptance as a way to encode and share all kinds of structured and unstructured data. The most important inherited feature is the definition of application specific semantic structure elements. Despite its roots, most XML development in recent years has targeted application integration.

XML, in its most basic definition, is simply a way of labeling information. It is an open standard that uses a human-readable tagging system to achieve two objectives:

1. Label Content Components: This defines exactly what is contained within each element of a document resulting in a file that is self-describing. For example: the following is tagged using XML to define the title of this document:

2. Label Content Structure: This defines the hierarchy of a document. Labeling this structure allows us to add context to our content. For example, the hierarchy of a news article may look like the following:

The semantic mark-up in the information gives us detailed access to the content. Each semantic content element provides a handle for that particular information unit and says something about the content, providing a more flexible and precise handling of information.



By identifying components and structure, various applications can now process these pieces in different ways, greatly increasing the sophistication of these applications and thus the usefulness and value of the content.

XML Schemas & Standards

XML tags have no pre-defined meaning. This means that virtually any word(s) can be used as an identifier within the tags. Human-readable XML tags provide a simple data format, but the intelligent defining of these tags and common adherence to their usage determines their value.

An XML schema is the definition of an XML document, which includes the XML tags and their interrelationships. Enterprises can use any number of schemas within their organization to meet specific functional, departmental, or enterprise-wide needs. Most companies choose to develop their own internal schemas to reflect the semantics of their unique content and implement widely-accepted industry-specific XML standards when similar information needs are distributed outside the organization to partners and customers.

Since XML first emerged from the World Wide Web Consortium (W3C) in 1998, the number of XML-associated standards has accelerated at a phenomenal rate. Standards, by their very nature, take years before they take hold and the relative immaturity of these standards means that companies must not only support multiple standards, but multiple versions of them. This issue of multiple standards contributes to the lack of critical mass of business partners behind a specific standard or standard version. **Table 1** provides examples of some of the most widely accepted XML standard schemas across the financial services industry today.

Table 1: Widely Accepted XML Standards within Financial Services

Standard	Description
FpML Financial Products Markup Language	FpML is a standard for the trade of derivatives and other financial instruments. It provides an XML-based specification that is focused specifically on over-the-counter (OTC) financial instruments directly traded between banks and other financial institutions.
IFX Interactive Financial Exchange	IFX allows financial institutions and associated service providers to access account information, download credit card statements, transfer funds, process consumer and business payments, enable bill presentment, and improve customer service.
MISMO Mortgage Industry Standards Maintenance Organization	MISMO standardizes information regarding loan data sent between two organizations. Key elements of functionality in the specification include credit reporting, loan boarding, applications, service orders, underwriting, and supporting services.
ACORD Association for Cooperative Operations Research and Development	ACORD specifies all the aspects of the insurance life cycle, from customer acquisition to claims fulfillment. These are divided along the lines of Property & Casualty, Life, and Surety insurance. ACORD specifies a very large and thick Document Type Definition (DTD) around the vocabulary and exchange mechanisms designed to meet these needs.
XBRL Extensible Business Reporting Language	XBRL uses XML to describe financial statements for both public and private companies. The specification facilitates the exchange of general as well as extremely detailed business reporting information such as financial statements, financial information, non-financial information, and general ledger transactions.
RIXML Research Information Exchange Markup	RIXML is an open, XML-based format that aims to provide tags for any piece of research content, in any form or media, with enough meta-data information for consumers to search, sort and filter through publisher research and quickly provide highly relevant information to their decision makers.

XML Saturation: Data vs. Content

One of the early and big surprises in the development of XML is that although it comes out of the publishing and document processing world, it has been most embraced for the integration of highly structured data as opposed to content.

XML represents data in a platform-neutral, open, and extensible manner. It creates a common foundation for unifying diverse platforms, applications and communities without displacing the underlying heterogeneous systems and architectures. As a result, the most prevalent use of XML to date has been for the implementation of XML-based messages to allow the exchange of information from these multiple, distributed systems.

That is not to say that XML has only a singular use. XML is replacing HTML as the underlying language of the World Wide Web. There are hundreds of XML industry standards, and there are thousands of software products that claim XML support. XML truly is used everywhere. However, the full impact of XML has not been felt because it is not being used to its full potential. No where is this more prevalent than in the area of content.

Applying XML to Content

Content today is frequently out of context, difficult to reuse, and constantly changing with multiple versions in multiple languages. Content solutions that leverage XML promise to improve the economics of working with content through:

- **Reuse:** XML allows you to break content down into granular elements that can be reused and repurposed in many forms and in many contexts. This reuse capability also allows the same content to be simultaneously linked in multiple documents. When a change is made to one element, it is automatically reflected in all instances that reference it.
- **Multi-channel Publishing:** XML is channel-independent and as such, allows you to deliver content real time in a variety of formats to worldwide audiences. Content maintains its quality, integrity and security no matter how it is delivered – in print, PDF, online or wireless.
- **Intelligent Content Retrieval:** XML provides the opportunity to use the structure of documents to write complex queries. Results reflect both the terms or words searched and the context in which the terms or words appear in the content. This means that queries are more focused and results are more precise. Results can return only the XML element(s) that meet the query parameters, rather than entire documents, to answer precise questions.
- **Customization:** XML gives you the ability to create virtual documents by assembling, on-demand, multiple pieces of content from various sources into a single document. This allows you to “slice and dice” content to create new products and services and deliver information customized to individual requirements.

To what extent these benefits can be derived is a function of two items: the level of structural, semantic, and metadata granularity of the XML schemas employed, and the level of XML support provided by your content management platform.

Fortunately XML is flexible so that it can be extended or reduced to describe content of any size, as the content or the structure of the content changes. So, using XML to manage content and achieve reuse can be an iterative process.

Let's look at an example of XML content in action. **Figure 1** provides an example of a restaurant review featured in local magazine.

Figure 1: Magazine Restaurant Review (Courtesy of Time Out Group Ltd.)


At first glance, this resembles what most people call an unstructured document. However, if you examine the article, a structure begins to appear and an initial XML schema – a template of sorts that reflects the article’s metadata and basic structure – can be developed as shown in Figure 2.

Paternoster Chop House

[print this page](#)
[rate this venue](#)

Info:

Area: City
Category: British
Address: Warwick Court, Paternoster Square, EC4M 7DX [MAP](#)
Phone: 020 7029 9400
Travel: St Paul's tube
Hours: Open Mon-Fri 10.30am-11pm
Prices: Meals served Mon-Fri 12noon-3pm; 5.30-10.30pm. Meal for two with drinks and service: around £80



Review:

City restaurants are rarely good value. With only five days to make their money - and most of this at lunchtime - prices need to be much higher than in any other area of London. Catering for expense-account diners is the only way to make a healthy profit, and this tends to narrow horizons, not broaden them.

Paternoster is Terence Conran's latest City restaurant (of several), and is utterly predictable: City prices, tasteful but bland design, a rather manufactured atmosphere not helped by its location in a newly built development.

The menu is based around 'simple British food', and good suppliers have been found for everything from the steaks to the cheeses. Also in its favour was unstintingly charming service.

Despite an evident commitment to quality ingredients, only the puddings shone in our meal. Cambridge Burnt Cream (ie crème brûlée) was faultless, as it should be at £5.50; sticky toffee pudding and Bakewell tart were also textbook perfect. But the starters and main courses were merely average.

'Value' is not the goal of gastronomy, but when you're paying £18 for a Bamsley chop, you don't expect it to be mostly bone, fat and gristle. And the accompanying onion sauce 'tasted like Knorr' according to the American critic I was dining with (though it obviously wasn't). Another main course of wild mushroom pie (£12.50) resembled pub food from the pre-gastropub days, swimming in liquid with the crust soggy. Neither dish came with veg: a small bowl of chips costs an additional £3.

At £35 per head for a largely indifferent lunch without any alcohol, we couldn't help feeling that this is not one of Conran's better-value restaurants from his ever-growing portfolio. *Guy Dimond*

Figure 2: Basic XML Schema (Courtesy of Time Out Group Ltd.)

As this publisher’s journalists, customers, and syndication partners introduce new requirements, the publisher can extend the XML schema to include deeper semantic tagging within the body of the article. This lets content users develop more intelligent, precise queries to retrieve content as well as lets the content producers and publishers get more granular when selecting content for reuse. Figure 3 reflects this same piece of content with the enhanced schema.

```

<?xml version="1.0" encoding="ISO-8859-15" ?>
<article>
  <name>Paternoster Chop House</name>
  <listing>
    <city>London</city>
    <area>City</area>
    <category>British</category>
    <address>Warwick Court, Paternoster Square</address>
    <postcode>EC4M 7DX</postcode>
    <country>UK</country>
    <telephone>020 7029 9400</telephone>
    <travel>St Paul's tube</travel>
  </listing>
  <open>Open Mon-Fri 10.30am-11pm</open>
  <meals>
    Meals served Mon-Fri 12noon-3pm; 5.30-10.30pm.
    <averagecost>Meal for two with drinks and service: around £80</averagecost>
  </meals>
  </venue>
</listing>
  <review>
    <body>City restaurants are rarely good value. With only five days to make their money - and most of this at lunchtime - prices need to be much higher than in any other area of London. Catering for expense-account diners is the only way to make a healthy profit, and this tends to narrow horizons, not broaden them. Paternoster is Terence Conran's latest City restaurant (of several), and is utterly predictable: City prices, tasteful but bland design, a rather manufactured atmosphere not helped by its location in a newly built development. The menu is based around 'simple British food', and good suppliers have been found for everything from the steaks to the cheeses. Also in its favour was unstintingly charming service. Despite an evident commitment to quality ingredients, only the puddings shone in our meal. Cambridge Burnt Cream (ie crème brûlée) was faultless, as it should be at £5.50; sticky toffee pudding and Bakewell tart were also textbook perfect. But the starters and main courses were merely average. 'Value' is not the goal of gastronomy, but when you're paying £18 for a Bamsley chop, you don't expect it to be mostly bone, fat and gristle. And the accompanying onion sauce 'tasted like Knorr' according to the American critic I was dining with (though it obviously wasn't). Another main course of wild mushroom pie (£12.50) resembled pub food from the pre-gastropub days, swimming in liquid with the crust soggy. Neither dish came with veg: a small bowl of chips costs an additional £3. At £35 per head for a largely indifferent lunch without any alcohol, we couldn't help feeling that this is not one of Conran's better-value restaurants from his ever-growing portfolio.</body>
    <author>Guy Dimond</author>
  </review>
</article>
  
```

Figure 3: Enhanced XML Schema (Courtesy of Time Out Group Ltd.)

```

<?xml version="1.0" encoding="ISO-8859-15" ?>
- <article>
  <name>Paternoster Chop House</name>
  - <listing>
    <city>London</city>
    <area>City</area>
    <category>British</category>
    <address>Warwick Court, Paternoster Square</address>
    <postcode>EC4M 7DX</postcode>
    <country>UK</country>
    <telephone>020 7029 9400</telephone>
    <travel>St Paul's tube</travel>
  - <venue>
    <open>Open Mon-Fri 10.30am-11pm</open>
  - <meals>
    Meals served Mon-Fri 12noon-3pm; 5.30-10.30pm.
    <averagecost>Meal for two with drinks and service: around £80</averagecost>
  </meals>
</venue>
</listing>
- <review>
  - <body>
    City restaurants are rarely good value. With only five days to make their money - and most of this at
    lunchtime - prices need to be much higher than in any other area of London. Catering for expense-account
    diners is the only way to make a healthy profit, and this tends to narrow horizons, not broaden them.
    Paternoster is
    <owner>Terence Conran</owner>
    's latest City restaurant (of several), and is utterly predictable:
    <summary>City prices, tasteful but bland design, a rather manufactured atmosphere not helped by its location
    in a newly built development.</summary>
    <foodtype>The menu is based around 'simple British food', and good suppliers have been found for everything
    from the steaks to the cheeses.</foodtype>
    <service>Also in its favour was unstintingly charming service.</service>
    <foodreview>Despite an evident commitment to quality ingredients, only the puddings shone in our meal.
    Cambridge Burnt Cream (ie crème brûlée) was faultless, as it should be at £5.50; sticky toffee pudding and
    Bakewell tart were also textbook perfect. But the starters and main courses were merely
    average.</foodreview>
    'Value' is not the goal of gastronomy, but when you're paying £18 for a Barnsley chop, you don't expect it to
    be mostly bone, fat and gristle. And the accompanying onion sauce 'tasted like Knorr' according to the
    American critic I was dining with (though it obviously wasn't). Another main course of wild mushroom pie
    (£12.50) resembled pub food from the pre-gastropub days, swimming in liquid with the crust soggy. Neither
    dish came with veg; a small bowl of chips costs an additional £3.
    <value>At £35 per head for a largely indifferent lunch without any alcohol, we couldn't help feeling that this is
    not one of Conran's better-value restaurants from his ever-growing portfolio.</value>
  </body>
  <author>Guy Dimond</author>
</review>
</article>

```

It is the level of XML functionality within your content management architecture that dictates how granular and how specific an XML schema can be. Today, almost all content management vendors have some level of XML support, from how XML supports content and metadata for search, workflow and distribution, to its use for integration and for application design and deployment. Unfortunately, the level of support is as varied as the content within your enterprise.

To illustrate, many content management systems do not support multiple or evolving schemas, or do so in a way so that the costs of administering the system outweigh the benefits. In these cases, the exercise of developing multiple extensible schemas is a wasted effort. In addition, an organization that spends valuable time and resources implementing a detailed XML schema may be surprised to find that their XML-enabled content management system provides them few search options beyond classical free-text searching capabilities, meaning that it is difficult, if not impossible to perform contextual queries. In these cases, the XML schemas add little value in helping you retrieve content more intelligently.

To realize the full potential of XML, your organization should implement the technology with an eye towards the company's strategic objectives. Organizations must have a good understanding of the type of content they wish to store, the type of information retrieval required, and the desired level of flexibility regarding re-use and distribution. Once this foundation is established, your organization can develop their XML schemas accordingly and determine how much XML support is required from their content management vendor.

The Business Benefits of Exploiting XML Content

Today, XML is primarily used for data exchange but using XML for content is quickly becoming the norm. The following use cases are clear examples of how forward-thinking, competitively-focused companies are using XML to exploit their valuable information assets. While the applications vary, the goals remain the same: use XML and XML-enabled technologies to derive maximum value from high-value content to drive profitability through better informational representation, faster response to emerging market opportunities, and automation of content processes.

Media/Publishing Companies Repurpose & Syndicate High-Value Content

The need to move from a print-centric architecture to one that is content-centric, i.e., where information is dynamic and on demand, is essential for today's top publishers. These organizations look to XML to help them create new revenue streams by repurposing content to create specialized information products, delivered in any format and through any channel, while ensuring the integrity of content as it is shared across products.

To achieve these goals, some publishers have created flexible repositories of rich, multi-structured XML content. Using intelligent content retrieval that allows for structural queries on XML content, these publishers are "slicing", "dicing", and reassembling content to create specialized information services for customers and partners. XML's channel-independent nature allows for the automatic distribution of this information to print, Web and wireless syndication partners. **Figure 4** illustrates a structured query on XML content.

Figure 4: Structured Query on XML Content

The screenshot shows a search interface with the following sections:

- Search:** A dropdown menu set to "with Reviews in All publications". Below it are input fields for "Edition", "Year", "Review", and "Comment". There are also small icons for filtering (a star, a square, and a star).
- Listing information:** A form with fields for "Name", "Address", "Postcode", "District", "Dialing code", "Phone", and "Travel". To the right is a grid for "Area" with checkboxes for NW, N, NE, W, E, SW, S, SE, and an "OUTER LONDON" checkbox.
- Credit & Open:** A section with two columns of checkboxes. The "Credit" column includes American Express, Visa, JCB, UK TC, UK US TC, Diners Club, MasterCard, TC, US TC, and LV. The "Open" column has a table with meal types (brunch, lunch, dinner, breakfast) and days of the week (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday).
- Services:** A dropdown menu with a list of services: child, karaoke, kosher, no smoking, and outdoor tables.
- Comment:** A text input field at the bottom.

XML helps publishers extend their market leadership and drive tremendous efficiencies in the production process. A flexible XML architecture for content reuse allows publishers to:

- Respond immediately to emerging sub-segment market opportunities,
- Eliminate the unnecessary re-creation of content across different products and,
- Format and automate previous manual processes for search and distribution.

Publication archives are another area within Publishing where XML is adding great value. Archives are much more than just old news, they serve as a valuable source of information to a broad range of readers such as journalists, students, and researchers. As such, they serve as important revenue source for many organizations. However the process of searching through numerous years of archives is, at most times, an arduous process at best.

XML is helping to change this. By using XML to organize, classify, and add intelligence to these documents, organizations are creating rich archives with elegant search capabilities that provide significantly higher levels of precision, improving the experience of the end user and enhancing the value of archived information.

Eliminating Distribution Latency of Short Shelf-Life Financial Information

In today's intensely competitive environment, the timeliness and customization of high-value financial research plays a major role in terms of an organization's profitability. Financial services organizations need to be resilient. As research requirements change and impact the content structures that contain financial information and analyses, financial services organizations must immediately react and deliver new or updated content in "near real-time" to customers in the format and through the medium that each customer wants.

As early adopters of XML, financial service providers are implementing flexible XML architectures that allow them to adapt on-the-fly to structural changes in their research documents. Furthermore, cutting-edge financial firms use automated alerts driven by complex XML queries. In this way, new content can be delivered near real-time and customized, with notifications sent to thousands of applications, portals, and end users.

XML is helping banks increase the value of financial information while eliminating latency in the delivery of short shelf-life information. Intelligent retrieval of XML elements through highly structured, yet ad-hoc queries, ensures a truly customized environment where information distribution is one-to-one and not one-to-many.

On-Demand Assembly of Customized Insurance Policies

Insurance companies offer a vast array of personal insurance products to millions of policy holders. Today's environment dictates that these companies comply with new and constantly changing legislation and improve the service provided to customers through online support and delivery of customized products, all while driving down the costs of creating, distributing, and managing these policies.

Certificates of Coverage and Physician/Provider Contracts are examples of content that are normally developed and managed in MS Word. Many of these contract clauses are standard and easily repurposed, but the effort to customize contracts and, in particular, manage clause changes is time-consuming, costly, and prone to error. If a legislation requirement results in a changed clause for example, you need a way to easily assimilate the change and renew the affected contracts. Alternatively, if there is a change initiated by your sales representative, you need to easily notify relevant internal parties so that the impact of changes can be analyzed prior to renewal.

The insurance industry is fast turning to XML to help them address these issues. XML allows companies to define a policy by its individual clauses and the relationships between them. These clauses are authored once and stored in a flexible XML repository that allows for the on-demand virtual assembly of multiple clauses to create customized policies and send real-time notifications when updates are required. **Figure 5** illustrates this process.

XML and XML-enabled technologies are helping insurance companies drive end-to-end automation through the life of a policy; from compilation, to delivery, to updates based on legislative changes. The elimination of redundant content and errors due to manual processes are driving tremendous productivity gains while drastically improving service to customers.

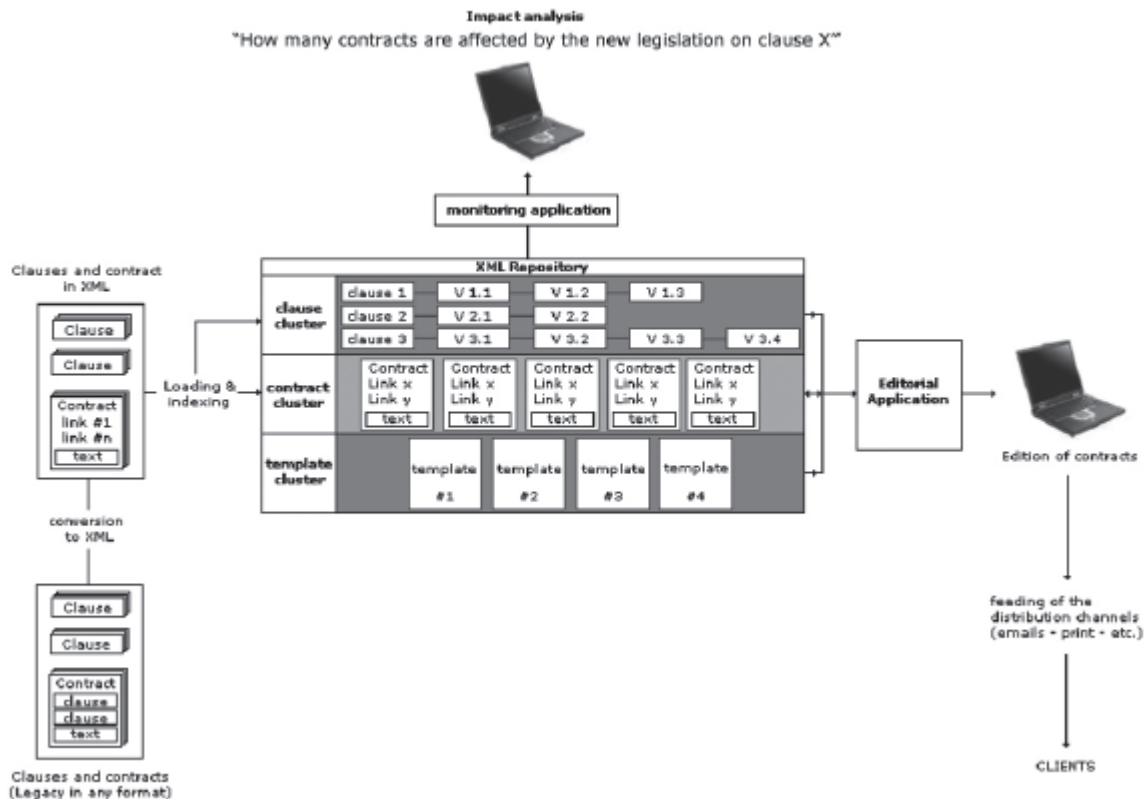


Figure 5: Contract Clause Update

Real-Time Auditing of Financial Transactions

While we don't think of financial transactions as the same type of content that we discussed in the above examples, they are much more than simple numerical data that can be easily managed in a relational database. Like the content described above, financial transactions are complex, hierarchical, and contain much textual information.

When a mission-critical financial transaction fails to complete, there is precious little time to rectify the situation before it costs an organization millions of dollars in lost revenue. The flexibility of XML combined with power tools that support highly granular XML storage and queries, can help these organizations gain immediate visibility into the state of these transactions. By archiving and persisting transaction messages, an organization has a complete audit trail and improved visibility into the state of mission-critical business transactions. In this way, an organization can provide proof that a business transaction is complete, on-time, and in accordance with company policy and regulatory requirements.

Intelligent, ad-hoc, content retrieval and virtual document creation produce a near real-time reproduction of events surrounding a transaction, allowing faster reaction time to potential issues. In addition, on-demand, ad-hoc auditing can also be conducted on message archives that receive millions of multi-formatted messages per day.

References

- [1] "The Role of XML in Content Management", October 2002, The Gilbane Report.
- [2] "Use XML to Structure and Manage Content", December 16, 2002, Gartner Group.
- [3] "What's Next for XML and Enterprise Content Management?", November/December 2003, The Gilbane Report
- [4] "XML: Chance and Challenge for Online Information Providers", Hans Holger Rath, Ph.D., at www.XML.org
- [5] "XML Databases: An Idea Whose Time has Finally Come", June 1, 2003, The Gilbane Report. [6] "XML in Financial Services", March, 2002, ZapThink Research.

About Xyleme, Inc.

Xyleme, Inc. is the industry's leading provider of open and standards-based learning content management solutions that enable single-source publishing of training content. Xyleme is 100% XML-based to create massive efficiencies through modularity and reuse. With Xyleme, training organizations improve collaboration in content development, reduce time-to-market for customized training products, and minimize the cost of supporting a blended learning strategy for print, online and mobile delivery.

With Xyleme LCMS, the entire content lifecycle, from authoring to storage to publishing, is pure open XML. Therefore, content is effortlessly repurposed across print, eBook, online courses and mobile applications and automatically synchronized across these learning products. Xyleme LCMS takes full advantage of user-generated content to support enterprise content development processes and can publish training content to both formal and social learning environments. The industry's leader in standards support, Xyleme complies with key industry standards including SCORM 2004 and 1.2, Common Cartridge, QTI and ePub.

For more information about Xyleme, Inc., visit www.xyleme.com.