XML: Portable data

Executive Summary - Extensible Markup Language (XML) provides a nonproprietary way to label data objects. These objects may be elements of a database, pieces of a Web page (including links, text, numbers, meta data and graphics) or the contents of a spreadsheet -- almost any form of coded content. Combined with style sheets and Java® programs, XML allows sharing and use of data across applications, organizations and platforms. It supports authentication, advanced data handling (including realtime updates and superior searching) and personalization of output. XML is easily applied to any implementation that involves filling out forms and is expected to replace electronic data interchange (EDI), which is widely used in many industries.

You have just fired off an angry note to your PC vendor. The data on Weird Al Yancovich's "Running with Scissors" is accessible, but the CD drive can't find the tunes. Almost immediately, you get a note back from the vendor. This is an interactive note, and it first confirms that the model, serial number and other data you included in the complaint are what they appear to be. Other than its identification of some of your more colorful comments as the titles of music CDs, it gets these data elements right. It then asks for a few more pieces of data that seem to point to symptoms. All at once, a screen opens, explaining that an outdated driver is the likely problem. Would you like the new driver installed? You click "yes," and in no time, you are listening to Al's angry rap, "It's All about the Pentiums."

How this happens

HyperText Markup Language (HTML) is the common language of the Web, but it has important limits, particularly as the Web becomes more interactive, with applications such as e-commerce. One key limit is the lack of information HTML provides about content. Its focus is on format (font, layout, color, etc.) rather than meaning. This is one of the reasons why search engines, which rely on keywords, can be so inefficient. Ambiguities, such as a "Cutlass" being both a sword and a car, are inherent to HTML. The second limit is the inflexibility of HTML, which has one standard set of markups to which all users, in theory, must adhere. XML is an attempt to resolve these difficulties. First, its focus is data, rather than format. XML looks at a Web page -- and virtually any other electronic document -- as a collection of data objects. Cutlass can be unambiguously marked up in readable text as, for instance, <car>Cutlass</car>. Second, XML is extensible, so customized sets of markup tags (vocabularies) can be developed for specific business, industries or even individuals.

XML has a number of advantages that are leading to its broad adoption by business:

- **mature**: XML is based on Standard Generalized Markup Language (SGML), a markup language that predates the Web and is tested and familiar.
- **readable**: though XML isn't obvious to the lay person, its textual form makes it easy to debug. Additionally, the Internet is text-friendly, so compression and transmission of XML is no problem. XML also supports Unicode, making it an international language.
- **portable**: like HTML and Java®, XML is platform agnostic. In addition, its family resemblance to SGML facilitates interaction with legacy systems and applications.
Nonproprietary, open: there is no fee or license required for the use and development of XML. This encourages sharing and improvement. One area of high activity is the creation of business-specific vocabularies, document type definitions (DTDs), the set of rules for using XML to represent documents of a particular type, and parsers which process the data. This offers the advantage of reuse and supports sharing of data between organizations.

XML has inspired many clever applications -- exchanging information among component suppliers in the automotive industry; helping search engines and intelligent agents return more meaningful results; and storing and categorizing the résumés of employment applicants. But there are four anticipated applications that should find general implementation:

- **personalization**: since XML presentations are assemblies of data objects, they can be put together dynamically into a format that suits the specific user.
- **business integration**: the portability of XML makes it a good candidate for coding all or most of an organization's data. Content providers can automatically reformat a document to feed many different publishing media.
- **extranets**: XML allows not only uniform, secure sharing of data, it provides for processing of that data on the client side. This makes it an excellent choice as a cross-organizational standard. In retail, it is expected to subsume electronic data interchange (EDI) as the preferred means for handling sales data (catalog items, purchase orders, order confirmations, electronic payments and electronic receipts).

What this means to you
Almost every enterprise will be touched by XML. It is a key tool of e-business, central to Web-enabling common business applications. There are a number of large companies who are involved with the development of XML and are participating in the standards-setting community. One concern is that, while the XML standard is open, key vocabularies may become proprietary. This could have the largest effects if they are 1) widely used and 2) sufficiently nonstandard (i.e., they conflict with or even break standard XML). Proprietary vocabularies could become a means of hijacking XML so that its future is controlled by one company or a group of companies. Since many of the benefits of e-business and business integration are predicated on making data portable, large influential organizations should continue to support open use and development of XML. If open, standardized, XML-based vocabularies are critical to your success, participation in broad-based standards development organizations, which are supported and funded by a wide range of industry participants, should be a priority for your organization.

XML is still in the beginning stages of adoption, but it is on the to-do lists of many -- if not most -- major IT organizations. There is no important competing technology and there are no technical breakthroughs needed for business applications. Thanks to its relationship to SGML, XML can claim stability and reliability. Barriers to swift adoption include a shortage of developers, development tools and business-specific vocabularies and DTDs. Conceptually, XML is not revolutionary. Many of the benefits it brings are already assumed to be part of the new economy, including:

- transparency of data, regardless of systems and applications
- searchability
- availability of content in familiar formats, such as browsers and my spreadsheet
- dynamic assembly of documents, depending on role or purpose

Though it's not obvious to the user, these benefits are often the result of an intense back office investment. XML facilitates them by making them part of the way data is coded. Virtually every industry is actively involved in
XML, with finance, retail and healthcare leading the way.

XML will create discontinuous change because of its ubiquity, not because of specific value shifting or structural changes. Essentially, it is about creating and filling out forms. This limits its effect on individuals and communities, which often find value “between the lines.” A breakthrough that would automate the creation of new vocabularies and classification of data, or that would allow the creation of “fuzzy forms” (getting into the white space of forms without losing meaning) might make XML discontinuous in a more profound way. There is no evidence of such a breakthrough on the horizon.

Two Industries: Retail and Healthcare
For the Retail Industry, XML is a way to reach further and more quickly. It takes the advantages of EDI (such as querying current stock) to the whole supply chain and out to the customer. In particular, it will bring smaller businesses into the world of electronic transactions by lowering costs and complexity. XML also makes messages and forms updatable and actionable in realtime. It has the potential to make products more easily available and to help improve customer service. The ease with which XML data can be shared will make profiling of customer groups significantly easier, which may raise privacy issues.

The Healthcare Industry has been a hotbed of activity for XML-based initiatives for several years. Much of the data of healthcare is text based, and an increasing percentage of this data is being captured and stored in electronic formats. According to Eliot Kimber, there are four challenges which XML can address for this industry:
- Storing documents electronically in a manner that will enable them to survive changes in storage and processing technologies
- Representation of documents in a data format that can enable reuse or reprocessing for different purposes, including purposes not yet known
- Enabling effective search and retrieval of complex documents and information, and
- Enabling the interchange of electronically encoded documents between different applications and computer systems

There are currently efforts underway within a variety of industry standards organizations, including Health Level 7 (HL7), Digital Imaging and Communications in Medicine (DICOM), Object Management Group’s (OMG) healthcare domain task force (CORBAMed), and American Society for Testing and Materials (ASTM), to develop XML-based document and message exchange protocols. Within HL7 in particular, there are two major initiatives. The first is development of a multi-level Patient Record Architecture, which enables development of documents which are readable by standard XML-aware browsers, but support multiple levels of technical sophistication between systems exchanging information. The second major initiative, known as HL7 Version 3, will include a common Reference Information Model and message definition process, resulting in XML-based, industry-wide messaging capabilities.

Tek to watch
Java®
Scripting languages
XSL (Extensible Stylesheet Language)
Xlink and Xpointer (ways to locate XML data)
XQL (XML Query Language)
Data mining
References for this article


Yancovich, Al. It's All about the Pentiums. [Song title].

Yancovich, Al. Running with Scissors. [Song title].
About this publication

Executive Tek Report is a semimonthly publication intended as a heads-up on emerging technologies and business ideas. All the technological initiatives covered in Executive Tek Report have been extensively analyzed using a proprietary IBM methodology. This involves not only rating the technologies based on their functions and maturity, but doing quantitative analysis of the social, user and business factors that are just as important to its ultimate adoption. From these data, the timing and importance of emerging technologies are determined. Barriers to adoption and hidden value are often revealed, and what is learned is viewed within the context of five technical themes that are driving change:

Knowledge Management: capturing a company's collective expertise wherever it resides -- databases on paper, in people's heads -- and distributing it to where it can produce the big payoffs

Pervasive Computing: combining communications technologies and an array of computing devices (including PDAs, laptops, pagers and servers) to allow users continual access to the data, communications and information services

Realtime: "a sense of ultracompressed time and foreshortened horizons, [a result of technology] compressing to zero the time it takes to get and use information, to learn, to make decisions, to initiate action, to deploy resources, to innovate" (Regis McKenna, Real Time, Harvard Business School Publishing, 1997.)

Ease-of-Use: using user-centric design to make the experience with IT intuitive, less painful and possibly fun

Deep Computing: using unprecedented processing power, advanced software and sophisticated algorithms to solve complex problems, and derive knowledge from vast amounts of data

This analysis is used to form the explanations, projections and discussions in each Executive Tek Report issue so that you not only find out what technologies are emerging, but how and why they'll make a difference to your business. If you would like to explore how IBM can help you take advantage of these new concepts and ideas, please contact us at enVision@us.ibm.com. Visit the Centers for IBM e-business Innovation at www.ibm.com/services/innovation.

Executive Tek Report is written by Peter Andrews, an IBM emerging technology analyst, and is published as a service of IBM Corporation.