

# Frequently Asked Questions About XML

---

Last updated: November 4, 1998

## The XML Family

---

### XML, the language

#### What is XML?

Extensible Markup Language (XML) is the universal language for data on the Web. It gives developers the power to deliver structured data from a wide variety of applications to the desktop for local computation and presentation. XML allows the creation of unique data formats for specific applications. It is also an ideal format for server-to-server transfer of structured data.

#### Does XML replace HTML?

Microsoft expects many authors and developers to use XML and HTML in tandem, for example by using XSL to generate HTML.

#### What are the benefits of adding XML to HTML?

There are many benefits to using XML on the Web:

- **It delivers data for local computation.** Data delivered to the desktop is available for local computation. The data can be read by the XML parser, then delivered to a local application such as a browser for further viewing or processing. Or the data can be manipulated through script or other programming languages using the XML Object Model.
- **It gives users an appropriate view of structured data.** Data delivered to the desktop can be presented in multiple ways. A local data set can be presented in the view that is right for the user, dynamically, based on factors such as user preference and configuration.
- **It enables the integration of structured data from multiple sources.** Typically, agents will be used to integrate data from back-end databases and other applications on a middle-tier server, making this data available for delivery to the desktop or to other servers for further aggregation, processing, and distribution.
- **It describes data from a wide variety of applications.** Because XML is extensible, it can be used to describe data contained in a wide variety of applications, from describing collections of Web pages to data records. Because the data is self-describing, data can be received and processed without the need for a built-in description of the data.
- **It improves performance through granular updates.** XML enables granular updating. Developers do not have to send the entire structured data set each time there is a change. With granular updating, only the changed element must be sent from the server to the client. The changed data can be presented without the need to refresh the entire page or table.

## How does XML fit into the Microsoft Windows® Distributed interNet Applications (Windows DNA) strategy for building three-tier, Web-enabled applications?

XML is quickly becoming the vehicle for delivering structured data from the middle tier to the desktop. XML-based data can be integrated from multiple back-end (database) sources, using agents on the middle tier. Schemas (see the XML-Data section) can improve this process, as developers can describe and exchange data more precisely.

## Where will XML be used on the Web?

Because XML describes data in a consistent, self-describing, open format, XML could potentially be used anywhere there is a need for data interchange and delivery. Microsoft expects that initially XML will be used to describe information about HTML pages, as is the case today with the channel definition format (CDF) for building Active Channel™ content, as well as future applications such as searching and distributed printing.

More important, because XML can describe data itself, it will be useful for delivering any kind of data — such as financial transactions, news updates, weather information, patient records, and legal libraries — to the desktop. Once on the desktop, applications can compute with the data and dynamically present the data.

## Does Microsoft Internet Explorer 4.0 support XML?

Yes, Internet Explorer 4.0 supports XML. It supports the following features:

- **A generalized XML parser**, which reads XML files and hands them off for processing to applications such as viewers. Microsoft has two parsers, the Microsoft XML parser in C++ ([www.microsoft.com/xml/cparser-f.htm](http://www.microsoft.com/xml/cparser-f.htm)), a high-performance, nonvalidating parser written in C++ that ships with Internet Explorer 4.0, and the Microsoft XML parser in Java ([www.microsoft.com/xml/parser/xmlparse-f.htm](http://www.microsoft.com/xml/parser/xmlparse-f.htm)), available for download from this site, for use by application developers.
- **The XML Object Model (XML OM)** uses the World Wide Web Consortium (W3C) standard Document Object Model (DOM) to allow programmatic access to the structured data, through the XML parsers, giving developers the power to interact and compute on the data. For more information on the DOM, see <http://www.w3.org/DOM>.
- **The XML Data Source Object (XML DSO)** allows developers to connect to structured XML data and supply it to the HTML page using Dynamic HTML's data binding facility.

## What is the level of XML support in Internet Explorer 5.0 Beta?

Internet Explorer 5.0 Beta has the following XML support:

- **Direct viewing of XML.** The Microsoft XML implementation lets users view XML using XSL or Cascading Style Sheets with their Web browser, just as they view HTML documents.
- **High-performance, validating XML engine.** The XML engine familiar to Internet Explorer 4.0 developers has been substantially enhanced and fully supports W3C XML 1.0 and XML namespaces, which let developers qualify element names uniquely on the Web and thus avoid conflicts between elements with the same name. Native XML support in Windows means that developers can count on the full XML processing

capabilities being present to read and manipulate the data they move between their applications and components.

- **Extensible Style Language (XSL) support.** With the Microsoft XSL processor, which is based on the latest W3C Working Draft, developers can apply style sheets to XML data and display the data in a dynamic and flexible way that can be easily customized. The querying capabilities of the Microsoft XSL processor also allow developers to programmatically find and extract information within an XML data set on the client or the server.
- **XML Schemas.** Schemas define the rules of an XML document, including element names and rich data types, which elements can appear in combination, and which attributes are available for each element. In order to enable multi-tier applications, Microsoft will be releasing a technology preview for XML Schema based on the Schema submissions to the W3C XML working group.
- **Server-side XML.** Server-side XML processing allows XML to be used as a standard means of passing data between multiple distributed application servers -- even across operating system boundaries.
- **XML document object model (DOM).** The DOM is a standard object application programming interface that gives developers programmatic control of XML document content, structure, formats, and more. The Microsoft XML implementation includes full support for the W3C XML DOM recommendation and is accessible from script, the Visual Basic development system, C++, and other languages.
- **C++ XML Data Source Object.** This XML DSO allows you to bind HTML elements directly to an XML data island. In addition, it has increased performance, has a greater ability to bind to various XML nodes, and takes advantage of all the new data binding features within Internet Explorer 5.0 Beta.

## What is the difference between SGML and XML?

The Standard Generalized Markup Language, or SGML (ISO 8879), is the international standard for defining descriptions of structure and content in electronic documents. XML is a simplified version of SGML; XML was designed to maintain the most useful parts of SGML. While SGML requires that structured documents reference a document type definition (DTD) to be valid, XML allows for "well-formed" data and can be delivered without a DTD. XML was designed so that SGML can be delivered — as XML — over the Web.

## How are HTML, Dynamic HTML, and XML related?

HTML is used in conjunction with CSS to format and present hyperlinked pages. Dynamic HTML, through the Document Object Model, makes all elements in HTML accessible through language-independent scripting and other programming languages, thus dramatically increasing client-side interactivity without additional requests to the server. The page's object model allows any aspect of its content (including additions, deletions, and movement) to be changed dynamically.

By adding XML for structured data, developers have the technologies they need to build the next generation of rich, flexible Web applications. With XML, they can deliver structured data

to the desktop and compute on the data via the XML Object Model. Today developers can display XML-based data in a browser, such as Microsoft Internet Explorer 4.0 and Microsoft Internet Explorer 5.0 Beta, or in other applications through scripting. In addition, they can also apply formatting rules to the data without complex scripting using XSL stylesheets, which essentially transform the XML-based data into display. These two methods of displaying XML-based data make it possible to generate multiple views of complex data.

### **Will it be necessary to compress XML for transmission over the Web?**

In general, the need to compress XML data will be application-dependent and largely a function of the amount of data being moved between the server and the client. XML compresses extremely well because of the repetitive nature of the tags used to describe the structure of the data. Benchmarks will be provided in the future to assist with determining whether compression is necessary. It is worth noting that compression is standard to HTTP 1.1 servers and clients, and XML will automatically benefit from this.

### **How secure is XML as a data format? Are there plans to add security to XML?**

XML is as secure as HTML. Just as secure HTTP (HTTPS) can be used to add encryption to HTTP, thereby protecting HTML, it can also be used to protect XML. XML is a text-based format for representing structured data. This maximizes simplicity and interoperability with the data. A number of steps can be taken to add security and authentication to the XML format. First, XML can be encrypted on the server before transmission to the client, then decrypted on the client. XML can also be authenticated by digital signatures applied to the data itself.

### **How will XML be generated from existing databases?**

In general, this will be handled using a three-tier architecture. Agents will be built to run on the middle tier to access multiple existing DBMSs and output XML. These agents will also support the ability to generate XML “updategrams” bidirectionally, to inform the client of changes made to the data on the middle tier or database server, and vice versa. Consequently, the agents will be able to receive updategrams from the client and send updates to the DBMS.

### **What is a DTD? What is it used for?**

The document type definition (DTD) defines the valid syntax of a class of XML documents. That is, it lists a number of element names, which elements can appear in combination with which other ones, what attributes are available for each element type, etc. A DTD uses a different syntax from that used by XML documents.

### **Do Web developers have to include a DTD when they use XML to describe data?**

No. XML can be used to describe data with or without a DTD. The term “valid” XML refers to XML data that references a DTD, while “well-formed” XML refers to XML that does not use a DTD. The addition of well-formed XML is one of the fundamental differences between XML and SGML. Clearly, in both cases, the XML itself must conform to the standards for the language (so, for example, all tags must be closed and tags may not overlap).

## What are XML schemas? How are they different from DTDs?

[From the W3C XML Activity Page at <http://www.w3.org/XML/Activity.html> ] While XML 1.0 supplies a mechanism, the Document Type Definition (DTD), for declaring constraints on the use of markup, automated processing of XML documents requires more rigorous and comprehensive facilities in this area. Requirements are for constraints on how the component parts of an application fit together, the document structure, attributes, datatyping, and so on. The W3C XML Schema Working Group is addressing means for defining the structure, content and semantics of XML documents.

In Internet Explorer 5 Beta, Microsoft is providing a release of XML Schema as a technology preview that may be useful for developers interested in building prototypes and gaining experience with schema. This technology preview is based on the XML-Data note submitted to the W3C. XML Schema, as implemented in this technology preview, can be thought of as the subset of the XML-Data submission that corresponds to the feature set proposed for Document Content Description (DCD) . Microsoft is actively involved in defining the emerging W3C XML schema standard and will track this effort. Developers should note that the version of XML Schema released with Internet Explorer 5 Beta is subject to change.

## What are namespaces? Why are they important?

Namespaces are another advanced feature of XML, outlined in a W3C note as part of the XML 1.0 specification. They allow developers to qualify uniquely the element names and relationships and make these names recognizable to avoid name collisions on elements that have the same name but are defined in different vocabularies. They allow tags from multiple name spaces to be mixed, which is essential if data is coming from multiple sources.

For example, a bookstore may define the <TITLE> tag to mean the title of a book, contained only within the <BOOK> element. A directory of people, however, might define <TITLE> to indicate a person's position, for instance: <TITLE>President</TITLE>. Namespaces help define this distinction clearly.

## Extensible Stylesheet Language (XSL)

### What is XSL? What can you do with XSL today?

The World Wide Web Consortium (W3C) working draft for XSL divides the language into two main parts: transformation and formatting semantics. This release supports the transformation part of the W3C XSL specification . Microsoft is tracking the W3C working draft and will be updating this implementation to match the final W3C recommendation.

XSL is defined as an XML grammar that consists of a set of XSL elements. This grammar can be used to transform XML documents into HTML or XML documents.

You can use XSL for direct browsing of XML files and from the XML DOM. The XML DOM **transformNode** method supports the use of XSL Elements to perform transformations. The DOM **selectNodes** and **selectSingleNode** methods support the XSL pattern-matching syntax

that enables sophisticated queries for nodes within a particular context of the overall tree structure.

## **XML-Data**

### **What is XML-Data?**

XML-Data, a specification that has been submitted to the W3C for review, makes XML even more powerful and extensible. It outlines a richer method of describing and validating data, making XML even more powerful for integrating data from multiple disparate sources and building three-tier Web applications.

In January 1998, the W3C acknowledged the Extensible Markup Language XML-Data submission from Microsoft, ArborText Inc., DataChannel Inc., and Inso Corp. The specification is available for public review at <http://www.w3.org/TR/1998/NOTE-XML-data-0105/> or <http://www.microsoft.com/standards/XML/>.

## **Standards**

---

### **What is the relationship between XML and the World Wide Web Consortium?**

The W3C has an active XML Working Group. Microsoft was one of the co-founders of this group in June 1996, and since then numerous industry players have joined, including Netscape Communications Corp, IBM and Oracle. For more information on the XML standards process, see <http://www.w3.org/>.

### **What is the status of XML with the W3C?**

XML version 1.0 recently moved from the proposed recommendation phase to the recommendation phase, which is the last step in the approval process at the W3C, and is a very stable standard. For more information on the current XML specification, and on the submission and review process within the W3C, please refer to <http://www.w3.org/>.

### **What is the status of DOM in the W3C?**

The XML DOM recently moved from the proposed recommendation phase to the recommendation phase, which is the last step in the approval process at the W3C, and is a very stable standard. For more information on the current DOM specification, and on the submission and review process within the W3C, please refer to <http://www.w3.org/>.

### **Where does XSL stand in the W3C?**

XSL is currently in the working draft stage in the W3C. It was submitted by ArborText, Inso, and Microsoft in September 1997. Microsoft plans to update its XSL code to track changes as it moves forward in the standard-development process.

# XML Vocabularies and Data Formats

---

## What are XML vocabularies?

XML vocabularies are the elements used in particular applications or data formats — the definitions of the meanings of those formats. For example, in CDF, element names such as <Schedule>, <Channel>, and <Item> make up the vocabulary for describing collections of pages, when these pages should be downloaded, etc. Vocabularies, along with the structural relationships between the elements, are defined in XML DTDs today, and soon with XML-Data schemas.

## What is CDF?

The channel definition format is an XML-based data format used in Microsoft Internet Explorer 4.0, for describing Active Channel content and Active Desktop™ components. It is used by thousands of content developers and millions of end users to describe collections of pages and data about pages, such as channel bar display, download behavior, Web page usage, and page-hit logging. For more information on CDF, see <http://www.microsoft.com/standards/cdf-f.htm>.

## What is OSD?

The open software description (OSD) is an XML-based data format, fully supported in Microsoft Internet Explorer 4.01, for advertising and installing software components over the Internet. When new versions of software become available, OSD provides a mechanism to notify the user (a process referred to as publishing). In addition, OSD provides the functionality to describe in great detail how to install ActiveX® Controls and Java packages and class files, adding functionality to the use of .INFs for setup. Microsoft and Marimba Inc. submitted this specification to the W3C in August 1997. For more information, see <http://www.microsoft.com/standards/osd/>.

## What is OFX?

The open financial exchange (OFX) is a data format that Microsoft Money and Intuit Quicken personal finance applications use to communicate with financial institutions over the Web. Although it is currently described using SGML, OFX will soon be based on XML.

## What is RDF?

The resource description framework (RDF), is an XML-based application being developed in the W3C. It brings together ideas from the meta content format, or MCF (technology acquired by Netscape from Apple Computer Inc.) and XML-Data (defined in a proposal recently submitted to the W3C by Microsoft, ArborText, DataChannel, and Inso).

RDF allows for generalized searching of information without application-specific rules, such as those defined in DTDs. RDF allows a complementary view of data through graphs and nodes, rather than through a structured tree, which the current XML technology enables. RDF,

together with XML schemas, will provide a standard way for developers to write these relationships down for broad classes of XML elements.

The crucial technologies that will deliver value this year and next are XML for structured data, XML namespaces to make names unique and recognizable, and new XML tags that add meaning to data so smarter search engines can perform better searches.

## Competition

---

### Does Netscape Navigator support XML?

No. Although Netscape has talked about support for XML, and the company recently joined the XML Working Group in the W3C, it has referred to XML as a “futures” technology for release later in 1998. Microsoft supports XML today in Internet Explorer 4.0.

## Tools Support

---

### What tools support XML today?

Many of the top SGML vendors have made generalized XML versions of their products available. These include ArborText ADEPT 7 (<http://www.arbortext.com/>) and Inso DynaBase (<http://www.inso.com/>), Chrystal Software Astoria (<http://www.chrystal.com/>), and POET Object Server (<http://www.poet.com/>) for authoring, editing and database publishing. Other vendors, such as DataChannel (<http://www.datachannel.com/>), have products based on XML for data management. See <http://www.microsoft.com/xml/> for a listing of some of these tools.

### What companies have promised XML support in their products in the near future?

Allaire Corp., ExperTelligence Inc., InterMax Solutions Inc., Pictorius Inc., Powersoft Corp., and SoftQuad Inc. recently committed to providing XML support in their products by March 1998.

### Where will the tools come from in the future?

Microsoft expects a wide variety of applications to be developed in the coming months that convert information currently stored in documents and databases into XML for delivery to the desktop. In addition, Microsoft expects XML-centric databases, rich authoring and application developer tools, and data format-specific tools such as wizards to be developed as new vocabularies are defined.

© 1997, 1998 Microsoft Corporation. All rights reserved.