The World Wide Web

By Richard W. Boss

The Web Wide Web (WWW or the Web) was first conceived in 1990 by scientists at CERN, a research organization in Geneva. It combined two established technologies: hypertext and the Internet. Hypertext, which has been in use on computers since the 1960's, makes it possible to link from one resource to another by clicking on hot spots. The Internet evolved from the networks of computers established by government agencies and academic institutions into a global network linking millions of computers in more than one hundred countries. The term came into common use in the 1970s when many previously separate networks were interconnected and commercial users were allowed on the global network. The term "web" was chosen to describe the interrelatedness and interconnection of resources on the Internet.

The hypertext language that became the foundation of the Web is HTTP, the HyperText Transfer protocol (HTTP). It is a standard for formatting documents so that there are links to other documents, including text, graphics, audio, and video files. Unlike older hypertext languages, HTTP required only unidirectional links. While that made it possible to link to another resource without action by the owner of that resource, it is the reason why there is "link rot," links that no longer work. The ability to jump from one document to another by clicking on hot spots has revolutionized searching on the Internet.

Other basic ideas underlying the Web are resource identifiers, client/server architecture, browsers, and markup language. The resource identifiers, commonly known as URLs, are used to locate a particular resource on the network. Client/server architecture puts the presentation level for information on the user's device, rather than a host. Using a browser, the client makes a request of the server to provide it with resources or services. The browser replaced the earlier gopher around 1993. The mark-up language (e.g. HTML or Hyper Text Markup Language)

consists of characters or codes embedded in text that indicate structure, semantic meaning, or advice on presentation.

The Web has changed dramatically in the past few years. It began as a way to search the Internet for resources and to link from a resource to other related resources. The emphasis was on finding, viewing, and downloading or printing information. In other words, the early Web was a publishing medium. The new direction, which began around 2003, and which became to be known as Web 2.0 in late 2004, emphasizes participation. Britannica Online is an example of Web 1.0, while Wikopedia is an example of Web 2.0. The former is read-only while the latter allows a large user community to edit the online content.

A common differentiation between Web 1.0 and Web 2.0 is that the first is about connecting computers and Web 2.0 is about connecting people. Another way of looking at it is that Web 2.0 software faces in two directions at the same time, one that consumes and remixes data from multiple sources while at the same time providing data and services in a way that allows remixing by others.

For more information about applications that utilize Web 2.0, consult the TechNote on Blogs and Wikis on this site.

Web Services

The term “Web services” describes a standardized way of integrating Web-based applications—including databases, spread sheets, word processing, and communications—using the XML, SOAP, WSDL, and UDDI open standards over the Internet.

XML is used to tag the data, SOAP is used to transfer it, UDDI is used for listing what services are available, and WSDL is used to describe the services that are available. Unlike the client/server models that have been widely used in the past few years, such as a Web server/Web

page system, Web services do not provide the users with a graphical user interface (GUI). Instead, Web services share information through a programmable interface. In other words, the applications interface, not the users. Developers can then add the Web services to a graphical users interface to offer specific functionality to users. The user interface can be tailored to the application. Web services do not require the use of HTML.

The Value of Web Services

Web services are the least expensive approach yet to interfacing heterogeneous systems and applications. The virtues of the Web are simplicity of access and ubiquity. Using the standards described it this TechNote, it is possible to interface systems without using custom coding or other proprietary solutions, without modifications to the applications that need to be interfaced, and without the need for users to acquire new skills.

Web services have great potential value for libraries. For example, Web services can be used to interface the acquisitions module of an automated library system and the system of a book jobber with minimum programming and minimum operator intervention at the library or book jobber. Thus, an operator in a library could enter and transmit an ISBN to a book jobber's system. The receiving system would automatically transfer bibliographic and pricing information to an acquisitions purchase order line record in the library's integrated library system. This approach is superior to the more labor-intensive BISAC and EDI-based interfaces that have been in use for the past few years.

Web services may also facilitate access to electronic databases. Several vendors have developed an XML gateway from a federated search system to the contents hosted on an electronic documents depository. This is particularly important when much of the information in a database is not in the MARC format. Anticipating the potential success of Web services for access to electronic databases, EBSCO introduces a Web services interface to EBSCOhost as early as 2004.

Web services may at some point in the future displace Z39.50, the standard used to link patron access catalogs, because Web services is easier to implement, broader in scope, and less expensive. However, that probably will not occur until the Web services technology matures somewhat.

The Roles of the Web Services Components

XML (Extensible Markup Language) was original developed as a language for defining new document formats for the Web. It was derived from SGML (Standard Generalized Markup Language). As early as 1983, the library at the University of California/Berkeley started developing a method for encoding archival materials in XML. The work led to the development of the Encoded Archival Description (EAD) standard.

The California Digital Library not only uses XML to store books in a standardized format, but it also uses it to allow users to define their own displays.

Oregon State University has been using XML since 1998 to automatically search interlibrary loan requests and print request forms sorted by location and call number, complete with availability information, scannable Ariel addresses, shipping labels (if needed), and billing data customized to the borrowing institution.

The Library of Congress announced MARC.XML in 2002, a specification for representing MARC data in an XML environment. That will facilitate the combining of search results from sources that use MARC with sources that do not.

More recently, the use of XML has gone beyond documents to structured data, including databases, spreadsheets, program configuration files, and network protocols. It is preferred to earlier formats because it can represent both tabular data such as data from a database or

spreadsheet and semi-structured data such as a Web page or business document. XML is independent of hardware platforms, operating systems, and programming languages. It is superior to HTML in that it does not have predefined elements and attributes that must be used in all applications. Instead, it is possible to create an XML vocabulary that is specific to an application.

XML can be used without SOAP, UDDI, and WDDL, but when used with these standards the result is Web services.

SOAP (Simple Object Access protocol) is an XML-based protocol specification that defines a uniform way of passing XML-encoded data, whether a request or a response. SOAP messages are independent of any operating system or protocol and may be transported using SMTP, MIME, or HTTP. When SOAP uses HTTP as the transport mechanism, it is not blocked by most firewalls unless system administrators have configured the firewall to selectively block out SOAP requests using SOAP-specific HTTP headers. Without SOAP, it is necessary to use platform-specific solutions.

UDDI (Universal Description, Discovery, and Integration Service) provides a mechanism for clients to find other Web services. It can be thought of as a DNS service. A ULDDI registry has two kinds of clients: organizations that want to publish a service and its usage interfaces, and clients who want to obtain services of a certain kind.

WSDL (Web Services Definition Language) provides a way for service provides to describe the basic format of Web services requests. It describes what a Web service can do, where it resides, and how to invoke it.

The leading source of technical information is W3C (www.w3.org). Its site maintains nearly 700,000 Web pages about its activities and the standards that deal with the Web, including Web services. The home page features a section entitled “Where do I find what I

need?”

The Web sites of vendors of integrated library systems also contain information about Web services. However, the information tends to be buried in product descriptions, rather than searchable by that term.

Prepared by Richard W. Boss, March 26, 2007, revised May 7, 2007