Toward a Web of Data and Programs

Steve Bratt
World Wide Web Consortium
steve@w3.org

Abstract

The explosive growth of the World Wide Web can be attributed largely to the power it has given to people to share and link documents. Growth of a much richer Web of the future could very well be driven by its ability to enable applications and people to share and link data and services over a broad range of devices. The World Wide Web Consortium (W3C) is developing new standards in the areas of Web Services and the Semantic Web that are fueling this enhancement of the Web. Implementation of these standards is growing rapidly.

1. Introduction

The continued globalization of science, commerce and society motivates the need for more effective ways to exchange data, information, knowledge, understanding and even friendly greetings within an expanding community of participants. The World Wide Web has revolutionized the way in which such exchanges can take place. Keys to the success of the Web has been its simple and elegant design, and the global interoperability offered by the standards upon which it is based.

Founded in 1994 by Web inventor Tim Berners-Lee, the 350 companies, academic institutions and organizations that comprise the World Wide Web Consortium (W3C) [1] have worked to develop the global standards (called W3C Recommendations) that have enabled the explosive growth of the Web. Perhaps the best known of W3C's standards are the Hypertext Markup Language (HTML) and the Extensible Markup Language (XML), but there are over 80 others at this time.

The early Web was largely a system of HTML documents shared among researchers. It was not long before technologies like HTML forms were applied to extend the Web to include interfaces between people and datastores.

The advent of XML [2] as the atomic unit of data exchange standards is making it easier for people and systems in different environments to share their programs and data. Emerging XML-based standards are enhancing graphics [3] and animation [4] on the Web, and aim to expand the Web to serve anyone (regardless of culture and ability), anywhere (regardless of geography) and using many modes of interaction (mouse, pen, voice, gesture, application-to-application) on any device (desktop, laptop, mobile, vehicle, appliance, RFID chips, etc.) [5, 6, 7, 8, 9, 10].

Work to develop and implement new standards in the areas of Web Services [11] and Semantic Web [12] should be of particular interest to the data management community. Web Services is a growing suite of XML standards to enable a Web of programs. The Semantic Web enables a Web of data with machine-processable meaning. These new standards promise to expand and facilitate communication and data exchange between people, and the systems that serve them.

This paper summarizes the work in W3C's XML, Web Services and Semantic Web Activities. Links to resources are provided for those wishing to explore particular topics in greater depth.

2. XML

The Extensible Markup Language (XML) is a simple, flexible text format derived from SGML [13]. Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data.

In terms of benefits, XML:

- Enables internationalized media-independent electronic publishing.
- Allows people to display information the way they want it, under style sheet control.
- Saves businesses money by enabling the use of inexpensive off-the-shelf tools to process data.
- Saves training and development costs by having a single format for a wide range of uses.
- Increases reliability, because user agents can automate more processing of documents they receive.

- Encourages industries to define platform-independent protocols for the exchange of data, including electronic commerce.
- Enables long-term reuse of data, with no lock-in to proprietary tools or undocumented formats.
- Provides the underpinnings of Web Services, the Semantic Web, and other technologies that enable a whole new level of interoperability and information interchange.

Though XML became a W3C Recommendation in 1998, a substantial amount of work continues within W3C's XML Activity [2]. Some of the most salient, recent work is highlighted here.

In February 2004, W3C released XML 1.1, which updates XML so that it no longer depends on the specific version of the Unicode character standard [14]. Work continues to develop XML Schema 1.1, the latest version of the language [15] to express shared vocabularies and allow machines to carry out rules. XML Query [16] is being developed to provide flexible query facilities to extract data from documents on the Web. This includes real documents authored in XML, and virtual documents where the contents of databases or other persistent storage that are viewed as XML via a mapping mechanism. W3C is updating XSL Transformations, XSL Formatting Objects (XSL-FO) and XPath [17]. These new additions enhance the existing family XML specifications, such as Namespaces, Infoset, DTDs, XInclude, XPointer, XLink and more.

While XML has been successful as a markup language for documents and data, the overhead associated with generating, parsing, transmitting, storing, or accessing XML data has hindered its employment in some environments where performance and storage are issues (e.g., high volume and/or low bandwidth communications). W3C has just completed an analysis of the issues surrounding the exchange of XML information in a binary format [18]. The conclusion was that the W3C should produce a standard for binary encoding of XML, and that this must be done in a manner that integrates with the existing XML stack and does not require changes to XML itself. W3C is currently considering options for moving forward.

3. Web Services

Web Services are based on a suite of XML standards which aim to enable a "Web of programs". Web Services support interoperable, extensible, platform-independent application-to-application interaction over a network. Thanks to the interoperability benefits XML, simple services can be combined in a loosely coupled way in order to achieve complex operations.

W3C's Web Services Architecture Working Group [19] produced a Note [20] which explains how the key Web Service components fit together. This document provides the following definition:

> *A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.*

The Architecture Note also provides Figure 1, which illustrates the general process for and elements involved in engaging a Web Service.

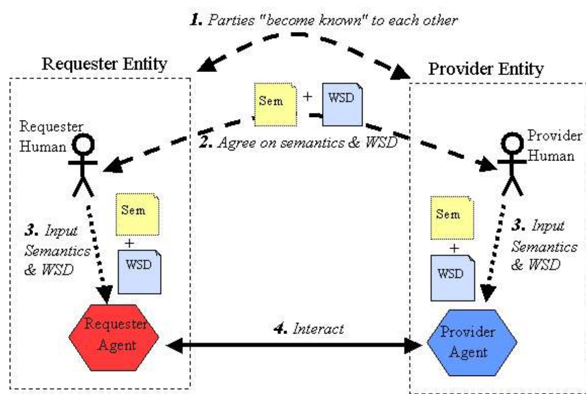

Figure 1. Engaging a Web Service

First (1), the requester and provider entities become known to each other (or at least one becomes know to the other). Second (2), the requester and provider agree on the service description and semantics that will govern the interaction between the requester and provider agents. This description is composed using the Web Services Description Language (WSDL). An open Web Service standard to exchange semantics does not yet exist. Third (3), the service description and semantics are realized by the requester and provider agents. Fourth (4), the requester and provider agents exchange messages, thus performing some task on behalf of the requester and provider entities. The protocol for realizing this exchange is SOAP.

Though this may sound straightforward, there are many tens of specifications on the Web Services landscape. Most of these began as proprietary specifications, developed by vendors. A subset of these have been submitted to open standards organizations. W3C is developing many of the core standards,. The Organization for the Advancement of Structured Information Standards [21] is working on other specifications in this space. Work to build profiles of and test suites for Web Services is underway within the Web Services Interoperability Organization [22]. More detail regarding the development of core Web Services standards at W3C is provided here.

W3C chartered the development of an XML-based messaging framework in September 2000 [23], before the term “Web Services” became common-place. SOAP Version 1.2 was released as a W3C Recommendation in June 2003. In January 2005, a set of specifications was released to optimize the transmission of SOAP messages. The SOAP Message Transmission Optimization Mechanism (MTOM) provides an efficient way to handle binary data in SOAP messages, in the form of attachments. This is based on XML-binary Optimized Packaging (XOP), a standard method for applications to include binary data, in conjunction with an XML document, in a package. Along with MTOM, the Resource Representation SOAP Header Block allows SOAP message recipients to access cached representations of external resources.

Another key standard is the Web Services Description Language (WSDL), which specifies how to describe interfaces to and interaction with Web Services. Included are descriptions of the format of messages to and from the service, protocols for exchanging messages, and the address of the endpoints providing this service. Work on WSDL 2.0 is nearing an end at this time.

W3C is also working on standards for choreography and addressing. The Web Services Choreography Description Language (CDL) [25] specifies, from a global perspective, the composition and description of relationships and interactions between Web Services (including sequencing, state management, etc.), to enable combinations of simple services to provide complex functionality. Web Services Addressing [26] specifies defines constructs for message addressing properties and endpoint references in a way that is independent of any particular transport or messaging system.

Web Services are in widespread and expanding use around the world. Amazon provides a Web Services toolset that provides direct access to their technology platform and product data [27]. Surveys of other Web Services implementations and tools are available on the Web [e.g., 28, 29]. Though many current implementations are based on proprietary versions of the specifications, a growing number are based on open standards from W3C and other standards bodies.

4. Semantic Web

The Semantic Web enables a "Web of data" – data with machine-processable meaning. This extension of the original Web provides a common framework for data sharing and reuse across different datastores, programs (including Web Services), enterprises, and communities. The Semantic Web does not require that all communities agree on a single vocabulary or ontology (e.g., definitions of and relationships between terms and data elements). The Semantic Web provides a standard syntax that communities large and small can use to describe their own data, vocabularies, ontologies, etc. Communities that employ Semantic Web standards greatly increase the level of data interoperability within their community, and between other communities employing these standards.

To appreciate the potential of Semantic Web requires a mental paradigm shift. The best way to understand the shift is to compare the Semantic Web to the HTML Web. Today's (largely) HTML Web is a network of *linked documents.* The Semantic Web augments the HTML Web by providing standards for *linked data.* This is especially important given that the volume of data potentially accessible over the Internet is growing rapidly. However, most of these data are trapped within non-standard datastores -- including proprietary database management systems, spreadsheets and flat files -- in a wide variety of formats. Most of these data are trapped within community stovepipes -- including stovepipes within companies, scientific disciplines, and cultures -- based on a variety of vocabularies and taxonomies. Today, HTML Web users follow links from one document to find other documents in a simple and largely human-directed process. Tomorrow, Semantic Web standards promise to release data from their traps, and enable humans, system-assisted humans, and systems themselves to navigate data-to-data links to find, aggregate and process data across different datastores.

A number of specifications, both completed and in progress, are envisioned to realize a Trusted Semantic Web (Figure 2). Three cornerstones of the Semantic Web, which have not yet been discussed here, are the Uniform Resource Identifier (URI), Resource Description Framework (RDF), and Web Ontology Language (OWL).

URIs could be considered the “primary keys” for the Web in general, and are key to the Semantic Web concept, as well. URIs provide unique and unambiguous identifiers for resources on the Web. They also provide unambiguous pointers to the location of those resources. URIs and International Resource Identifiers (IRIs) are standardized at the Internet Engineering Task Force [30], with the close cooperation of the W3C community [31].

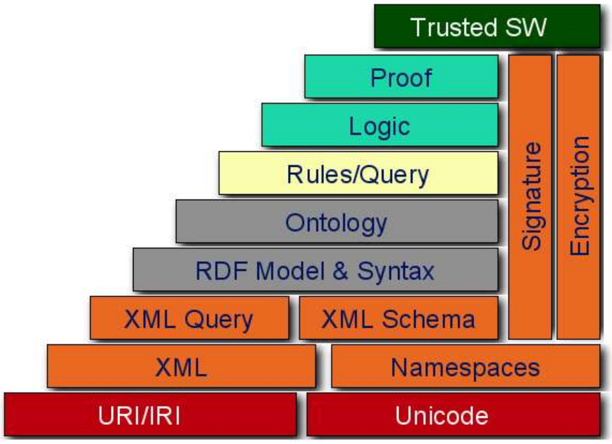

Figure 2. Semantic Web stack.

Much the same way that HTML is the framework that helped drive the original Web, RDF [32] provides a framework for supporting relationships between data on the Web. The RDF standard provides an XML syntax for descriptive statements expressed as "triples" (Figure 3). For example, a triple might be constructed to say that a particular person, "Steve Bratt" (subject) "has an employer" (property) that is "W3C" (value). Triples notation offers a very flexible mechanism for relating anything to anything else, including data housed within tabular and tree structures.

Figure 3. RDF triple.

If only literals are used, such as in the person-employer example above, the triple would not be very useful beyond the local environment. However, a major step toward the full potential of the Semantic Web is realized when URIs are used to identify the subject, the property and (possibly) the value. For instance, when a URI is used to identify the subject, statements about this same resource can be made anywhere on the Web. Further, any application or service on the Web may then *merge* these statements using the URI as a kind of primary key. In other words, data described using RDF with URIs can be more easily found, shared, reused and aggregated.

The Web Ontology Language [33] builds upon RDF to define concepts and relationships that describe an area of knowledge. A result of cooperation between the Web and knowledge representation communities, OWL has a number of distinct advantages over, say, a data dictionary for a relational database management system. OWL is open Web standard. OWL, together with RDF Schema [32], offers a richer vocabulary and formal semantics for describing properties and classes of things, such as relations (e.g. subPropertyOf), cardinality (e.g. Exactly 0 or 1), characteristics of properties (e.g. Symmetry), versioning (e.g., priorVersion) and others. These feature make it easier to do complex inferencing within and between datastores.

The of W3C's Semantic Web Activity over the next year will be on further development of the infrastructure stack (Figure 2), the application and deployment of these standards, and facilitating the sharing of data across distributed collections on the Web. This focus is reflected in the efforts of the RDF Data Access and Semantic Web Best Practices and Deployment Working Groups [12]. W3C is holding workshops on Semantic Web for Life Sciences [34] and Semantic Web Rules for Interoperability [35] as well.

Communities are already using RDF to publish their data on the Web. Early community adopters include the Dublin Core Metadata Initiative [36] and the RDF Site Summary (RSS 1.0) [37] for supporting news syndication. Early commercial adopters include Adobe, whose eXtensible Metadata Platform (XMP) [38] leverages RDF to enable more interoperable creation, processing, and interchange of document metadata across publishing workflows. The recent Semantic Technology Conference [39] highlighted the growing importance of this technology across an expanding breadth of applications.

5. Conclusions

First, one might ask, "How do Web Services and Semantic Web technologies fit together?" In reality, these technologies arose from largely different communities. Though Web Services is focused primarily on programmatic interoperability, and Semantic Web is focused primarily on data interoperability, the fact is that each supports a little bit of the primary function of the other. W3C is hosting a workshop entitled, "Framework for Semantics for Web Services" [40], which will foster discussion on how Web semantics and services technologies could complement each other, while enhancing the overall functionality of the World Wide Web.

Second, one might ask, "What can all of this new Web Services and Semantic Web technology do for the world?". To answer that, consider two additional questions: "What if these standards allowed applications to access not only services and data, but also structured descriptions of the services and data (e.g., the definition, units, origin, relationships)?" And, "What if individual communities in academia, industry and government employed the same underlying standards to describe and make available their data and services? “

Perhaps one of the most important reasons to take seriously the use and potential of Web Services and Semantic Web standards is because they are, well, Web standards. They are open, global, interoperable and firmly grounded within the architecture of the World Wide Web [41].

One could, with some confidence, extrapolate the global benefits that have arisen from the standardized HTML Web. There would be obvious benefits from improved data interoperability and networking within any particular community that chose to adopt Web Services and the Semantic Web. As more people employ these standards, linking and mapping between services, vocabularies, and ontologies from different communities would be facilitated. There would be expanded opportunities for sharing tools for authoring, sharing, finding, aggregating and analyzing data and services within a larger Web of communities. The “network effect” of an expanding number of semantic data and service nodes on the Internet could foster an explosive growth in communication, collaboration, creativity, and directed and serendipitous discovery. Thus, the most exciting thing about the Web Services and Semantic Web is not what we can imagine doing with them, but what we can't yet imagine they could do for us.

Acknowledgments

Thanks to the W3C staff and Working Group members for their many technical contributions and documentation, from which most of the contents of this paper have been drawn.

References

[1] W3C. http://www.w3.org
[2] XML. http://www.w3.org/XML/Activity
[3] Scalable Vector Graphics. http://www.w3.org/Graphics/SVG/
[4] Synchronized Multimedia Integration Language. http://www.w3.org/AudioVideo/
[5] Multimodal Interaction. http://www.w3.org/2002/mmi/
[6] Device Independence. http://www.w3.org/2001/di/
[7] Voice Browsing. http://www.w3.org/Voice/
[8] Web Accessibility. http://www.w3.org/WAI/
[9] Internationalization. http://www.w3.org/International/
[10] Mobile Web Initiative. http://www.w3.org/2005/02/FlyerWebMobile.pdf
[11] Web Services. http://www.w3.org/2002/ws/Activity
[12] Semantic Web. http://www.w3.org/2001/sw/Activity
[13] SGML. http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387
[14] Unicode. http://www.unicode.org/unicode/standard/versions/
[15] XML Schema. http://www.w3.org/XML/Schema
[16] XQuery. http://www.w3.org/XML/Query/
[17] XSL. http://www.w3.org/Style/XSL/
[18] Binary XML. http://www.w3.org/XML/Binary/
[19] Web Services Architecture. http://www.w3.org/2002/ws/arch/
[20] Web Services Architecture Note. http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/
[21] OASIS. http://www.oasis-open.org/home/index.php
[22] WS-I. http://www.ws-i.org/
[23] XML Protocol. http://www.w3.org/2000/xp/Group/
[24] Web Services Description. http://www.w3.org/2002/ws/desc/
[25] Web Services Choreography. http://www.w3.org/2002/ws/addr/
[26] Web Services Addressing. http://www.w3.org/2002/ws/chor/
[27] Amazon Web Services. www.amazon.com/gp/aws/landing.html
[28] XMethods Web Services survey. http://www.xmethods.com/
[29] Google Web Services survey. http://directory.google.com/Top/Computers/Programming/Internet/Web_Services/
[30] Internet Engineering Task Force. http://www.ietf.org/
[31] URI Activity. http://www.w3.org/Addressing/
[32] Resource Description Framework. http://www.w3.org/2001/sw/RDFCore/
[33] Web Ontology Language. http://www.w3.org/2001/sw/WebOnt/
[34] Semantic Web for Life Sciences Workshop. http://www.w3.org/2004/07/swls-cfp.html
[35] Rules Languages for Interoperability Workshop. http://www.w3.org/2004/12/rules-ws/cfp
[36] Dublin Core. http://www.dublincore.org/
[37] RSS. http://web.resource.org/rss/1.0/
[38] Adobe XMP. http://www.adobe.com/products/xmp/main.html
[39] Semantic Technology Conference, 2005. http://semantic-technology.com/
[40] Framework for Semantics for Web Services Workshop. http://www.w3.org/2005/01/ws-swsf-cfp.html
[41] Architecture of the World Wide Web, Volume One. http://www.w3.org/TR/webarch/