Semantic Web and Other W3C Technologies to Watch

Steve Bratt
Chief Executive Officer
World Wide Web Consortium

October 2006

http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/
Overview

- Success of the Web and the Value of standards
- World Wide Web Consortium
- Evolution of Web technologies
- Focus on areas which could have an impact to your business
  - ... and in particular, Semantic Web
The Web Has Grown ...

Number of Web Servers (Dec 1990 - Dec 2005)

Source: http://www.zakon.org/robert/internet/timeline/

(see also: Growth in number of IP addresses [graphic, link])
... Growing Around the World ...

<table>
<thead>
<tr>
<th>World Regions</th>
<th>Population (2006 Est.)</th>
<th>Population % of World</th>
<th>Internet Usage, Latest Data</th>
<th>% Population (Penetration)</th>
<th>Usage % of World</th>
<th>Usage Growth 2000-2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>915,210,928</td>
<td>14.1 %</td>
<td>23,649,000</td>
<td>2.6 %</td>
<td>2.3 %</td>
<td>423.9 %</td>
</tr>
<tr>
<td>Asia</td>
<td>3,667,774,066</td>
<td>56.4 %</td>
<td>380,400,713</td>
<td>10.4 %</td>
<td>35.5 %</td>
<td>232.8 %</td>
</tr>
<tr>
<td>Europe</td>
<td>807,289,020</td>
<td>12.4 %</td>
<td>294,101,844</td>
<td>38.4 %</td>
<td>28.2 %</td>
<td>179.8 %</td>
</tr>
<tr>
<td>Middle East</td>
<td>190,044,161</td>
<td>3.0 %</td>
<td>18,203,500</td>
<td>9.6 %</td>
<td>1.7 %</td>
<td>454.2 %</td>
</tr>
<tr>
<td>North America</td>
<td>331,473,276</td>
<td>5.1 %</td>
<td>227,470,713</td>
<td>68.8 %</td>
<td>21.8 %</td>
<td>110.4 %</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>553,908,632</td>
<td>8.5 %</td>
<td>79,962,809</td>
<td>8.7 %</td>
<td>7.8 %</td>
<td>350.5 %</td>
</tr>
<tr>
<td>Oceania / Australia</td>
<td>33,856,977</td>
<td>0.5 %</td>
<td>17,872,707</td>
<td>52.6 %</td>
<td>1.7 %</td>
<td>134.6 %</td>
</tr>
<tr>
<td>WORLD TOTAL</td>
<td>6,498,897,060</td>
<td>100.0 %</td>
<td>1,043,104,886</td>
<td>16.0 %</td>
<td>100.0 %</td>
<td>189.0 %</td>
</tr>
</tbody>
</table>

NOTES: (1) Internet Usage and World Population Statistics were updated for June 30, 2006. (2) CLICK on each world region for detailed regional information. (3) Demographic (Population) numbers are based on data contained in the worldofstats website. (4) Internet usage information comes from data published by Nielsen/NetRating, by the International Telecommunications Union, by local NICs, and other reliable sources. (5) For definitions, disclaimer, and navigation help see the Site Surfing Guide. (6) Information from this site may be cited, giving due credit and establishing an active link back to www.internetworldstats.com. ©Copyright 2006, Minwatts Marketing Group. All rights reserved.

Source: http://www.internetworldstats.com/stats.htm

Note: in 1995, there were ~16,000,000 Internet users, or 0.4% of global population (see also: Top languages on the Internet [graphic, link], English 30% w/ 125% growth, Chinese 14% w/ 347% growth per year)
What Led to the Web's Success?

- Simple architecture - HTML, URI, HTTP
- Networked - value grows with data, services, users
- Extensible - from Web of documents to .. you'll see in a bit
- Tolerant - works with imperfect mark-up, data, links, SW
- Universal - independent of HW, OS, SW, language, ability
- Free / cheap - browsers, information, services
- Simple (and fun) for users - text, graphics, links
- Powerful - for people (and machines)
- Open standards ...
Why are Open Standards Important?

- Cross-application integration
  - Friedman: "The best companies are the best collaborators ..." ([The World is Flat](http://www.amazon.com/The-World-Flat-Michael-Friedman/dp/0307352678))
- Avoids vendor lock-in
  - … need a particular operating system and/or browser to file taxes, access health insurance data, …
- Access for all
  - Developers, systems, users know what to expect
- Unexpected reuse
  - e.g., Web on phones, Google Pagerank, your data is useful to others
- Open, RF standards = good business sense ...
  - Enable unencumbered implementation of foundational standards
  - Enable commerce and communication to thrive on top of standards
- Customers / gov't regulations will mandate them
W3C's Mission: Leading the Web to its Full Potential

*Founded by Web inventor Tim Berners-Lee in 1994, W3C is:*

- **Unique International Standards Organization**
  - Providing the **Vision** to Lead
  - **Engineering** the **Open Standards** that Make the Web Work
- **19 Hosts/Offices:** [MIT](#) | [ERCIM](#) | [Keio](#) | Australia | Benelux/Bénélux | 中国 | Suomi | Deutschland und Österreich | Ελλάδα | 香港 | Magyarország | भारत | ישראל | Italia | 한국 | المغرب | España | Sverige | United Kingdom and Ireland | (map)

- **425 Members** ([history map](#), [largest](#))
- **800 Technologists** developing standards in **60 Groups**
- **65 Member-neutral Technical Staff**
- **Accountable to the Global Public**

*(Membership / Benefits / "At a Glance" brochure)*
Web Usage and Technologies are Evolving ...

- From a Web of **Documents** ...
- Toward **One Web** ...
  - ... of **Data and Services**
  - ... on **Everything**
  - ... for **Everyone**
- -- think **Web 2.0+**, **Web as a platform**, etc.
Engineering an Interoperable Foundation of the Web

- Close to 100 Web Standards to date (list/svg-by-yr/translations/~60 groups svg)
  - Including: HTML, XML, CSS, SOAP, SVG, Voice XML, RDF and more
The Real Question

- Q: Which of these are relevant to you?
- A: All of them… 😊
- But let's look at W3C's *emerging* technologies, and focus on one
Leading Edge: Web for Everyone

- Expanded Internationalization Activity
- Expanding global participation in W3C (global Offices, translations, etc.)
- New Web security and trust work under consideration
- Web Accessibility Initiative (charters)
Leading Edge: Web on Everything

- Mobile Web Initiative and [Device Independence](http://en.wikipedia.org/wiki/Device_Independence)
- Ubiquitous Web (workshop, [slides](http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/))
Leading Edge: Web of Data & Services

- **XML**: Binary, Processing Model
- **Web Services**: Performance, Addressing, SemWeb Services, Policy
- *Web of Data* - Semantic Web: Deployment, Query, Rules, Health Care/Life Sciences, Content Labeling, Geospatial, Multimedia Semantics, perhaps eGov, etc.
Why Do We Need the Semantic Web?

- Tasks often require *combining* data on the Web, e.g.:
  - Hotel, transport, meeting, personal info come from different sites
  - Cross-referencing disparate digital libraries
  - Mining data from biochemical, genetic, pharmaceutical, patient databases
  - Integrating data across the enterprise
- Humans understand how to combine this information ...
  - Not always easy (different vocabularies, languages, formats)
- Machines aren't smart enough to understand :-)

http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/
Why Can't Machines "Understand"?

Analogy:

*What We Say to Dogs*

- from Gary Larson cartoon ([local link](https)) ...
- "Stay out of the garbage! Understand, Ginger? Stay out of the garbage!"

*What Dogs Understand*

- "... blah blah blah blah GINGER blah blah blah blah blah ..."
What Machines "Understand"

"... blah blah `<a href=http://www.xwz.com/foo.html>.text-link.</a>` blah blah . . . ."*

- Computers must "understand" more:
  - Not human concept of "understanding"
  - Just useful machine processing, for example, capturing the nature of
    - the "link-text" object
    - the thing at the other end of the link ("http://www.xwz.com/foo.html")
    - the relationship between the two (why are they linked?)

--

* Where `<a href=...>` is HTML for a "link"
Toward Processable Search Semantics

**Google:**

- Web links are machine processable
- Minimal semantics assumed: "This" refers to "that" = popularity
- Amazing results from minimal semantics

*What if Web pages had more semantics?*

*(See Article by Bijan Parsia)*
Ways to Enable Machine Processing

Smarter Machines

- Teach computers to infer the meaning of Web data
  - Natural language, image recognition, etc.

- ... this is the Artificial Intelligence approach
  - *WARNING*: Hard problem!

Smarter Data

- Make data easier for machines to find, access and process
  - Express data and meaning in standard machine-readable format
  - Support decentralized definition and management, across the network

- ... this is the *Semantic Web* approach
What is the Semantic Web?

"The Semantic Web is an

... extension of the current web in which

... information is given well-defined meaning,

... better enabling computers and people to work in cooperation."

The Semantic Web

Tim Berners-Lee, James Hendler and Ora Lassila

Scientific American, May 2001
Semantic Web: Data on the Web

Machine-processable, global Web standards:

- Assigning unambiguous names (URI)
- Expressing data, including metadata (RDF)
- Capturing ontologies (OWL)

- Query, rules, transformations, deployment, application spaces, logic, proofs, trust (in progress)

- Semantic Web Activity. (Overview presentation)
Uniform Resource Identifiers (URI)

- **URIs** have two different uses:
  1. Unambiguous name for something
  2. Location of a document

- URIs can be used to identify definitions for concepts
  - Especially useful for ontologies & metadata

- Examples:
  - [http://www.w3.org/People/Berners-Lee/](http://www.w3.org/People/Berners-Lee/)
  - [mailto:steve@w3.org](mailto:steve@w3.org)
  - See also Internationalized Resource Identifier standard ([RFC 3987](https://tools.ietf.org/html/rfc3987))
Resource Description Framework (RDF)

- Semantic Web's Resource Description Framework: a W3C standard ([Primer](http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/))
  - Statements linking data so as to describe things (concepts, objects, etc.)
  - RDF: Data :: HTML link : Documents

- Descriptive statements expressed as triples:
  - (Subject, Predicate, Object) or (Subject, Property, Value)

![Subject Property Value Diagram]

- Most useful, Web-wise, when these are URIs, e.g. for this presentation:
  - [http://www.w3.org/2006/Talks/0811-sb-W3Cemergingtech/Overview.html](http://www.w3.org/2006/Talks/0811-sb-W3Cemergingtech/Overview.html) as the subject
  - [http://purl.org/dc/elements/1.1/creator](http://purl.org/dc/elements/1.1/creator) as the property
  - [http://www.w3.org/People/Bratt/stevenrbratt.rdf#SB](http://www.w3.org/People/Bratt/stevenrbratt.rdf#SB) as the value
RDF: Flexible Representation

Linking table and tree data ...
Web Ontology Language (OWL) +

- W3C standard ([Guide, Overview](http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/))
  - Defining concepts & relationships within area of knowledge

- Like DBMS data dictionary or schema, *but*..
  - Global, standard syntax based on RDF
  - Can define more complex, graph relationships
  - Using URIs as keys
  - On the Web

- Includes standard vocabulary for describing properties and classes:
  - Datatypes (e.g., integer, date, string ..)
  - Relations (e.g. subclass of, property of ..)
  - Cardinality (e.g. exactly one ..)
  - Characteristics of properties (e.g. symmetry)
  - ... and more
Standards En Route

- **Query**

- **Rules**

- **Transformations**

- **Deployment** and application spaces
  - Health Care and Life Sciences
  - thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies'
  - Content Labeling, Multimedia, geospatial (just starting)

- Logic, proof, trust, etc. in the future?
Most of the Current Web

- Minimal machine-processable information -- dumb links
Semantic Web: "Smarter" Resources and Links

- More machine-processable information: data connected by relationships

[ellipses = resources; color = one data source; x: = one ontology]

(see also more detailed example related to book searching and selling)
Enterprise Integration Today
Enterprise Integration on the "RDF Bus"
Clients on the "RDF Bus"

New data applications can be built on top of RDF bus, for example:

See Semantic Web interface concepts for other types of data sources (Tim Berners-Lee)
Adapting SQL/RDBMS Databases

Keep your existing systems running, and add RDF interfaces:

Examples of DBMS interfaces: 1. life sciences, 2. conference information

See Semantic Web interface concepts for other types of data sources (Tim Berners-Lee)
Applications Linked via Semantic Web

[Diagram showing various applications linked via Semantic Web technologies.]
Life Science Onotologies Linked via Semantic Web
Challenges Ahead

• Providing information in Semantic-Web-friendly ways
  ○ Exposing existing data stores as RDF
  ○ Automated and human-assisted tools to create RDF stores, ontologies, mappings
  ○ "Chicken and egg problem" en route to "network effect"

• Addressing operational and social needs
  ○ "Worldwide" Web
  ○ Quality, provenance, versioning, explanation, safety
  ○ Privacy, access control, authorization
  ○ Trust

• Making data accessible to people and programs
  ○ Usable interfaces to masses of semantic data
    ■ Search, filtering, aggregation, processing, graphics
  ○ Access via multiple modes, multiple devices
Getting Familiar with the Semantic Web

- See more detailed tutorial*
- Research groups (e.g., MIT/DIG, UMD, UBristol, DARPA/DAML, etc.)
- Commercial initiatives (HP, IBM, Oracle, Nokia, etc.)
- Conferences (e.g., ISWC, SemTech, etc.)
- Follow W3C's work to drive incubation and standardization
- Large datasets are accumulating. E.g.:
- Other portal examples:
  - Sun’s White Paper Collections and System Handbook collections; Nokia’s S60 support portal; Harper’s Online magazine linking items via an internal ontology; Oracle’s virtual press room; Opera’s community site,…
- Development tools
Timing Strawman

- 2006. Strategic planning. Model your data.
- 2007. Test by adding value to your organization's data web:
  - Build SemWeb wrappers around a few databases
  - Integrate disparate datasets to solve a few unsolved problems
  - Offer filtered SemWeb data to partners, customers

Killer apps (or even failures) could radically change the timeline.
Summary

- Evolution toward one Web ...
  - of Data and Services, on Everything, for Everyone
- Strong business case for
  - Understanding emerging standards
  - Implementing standards as they appear likely
  - Participating in standards orgs supports both of these
- A good resource ...

http://www.w3.org/
Extra Slides

- W3C Membership
Who are W3C's Members?

"Third-class companies make products; second-class companies develop technology; first-class companies set standards."

- **W3C's Members includes ...**
  - most of the world's leading IT companies
  - other large and small companies
  - academic and research institutions
  - government agencies
  - non-profit organizations

- **... which are ...**
  - developing Web-based products
  - using Web technologies
  - conducting research on the Web
  - developing specifications built upon W3C's work


[Full-fee Members (Jul 2006)]
Why Do People Participate in W3C?

- **Leadership**
  - Introduce ideas through [submissions](http://www.w3.org), [workshops](http://www.w3.org), [Incubator Groups](http://www.w3.org)
  - Influence standards through [Working Group](http://www.w3.org) participation, review, implementation

- **Early insight into market trends**
  - Access world's foremost Web technologists from [Member & Team](http://www.w3.org)
  - Plan for emerging technologies & markets through Member-confidential access

- **Promoting image as innovator**
  - Participate in international media activities, [press releases](http://www.w3.org), testimonials (e.g., [MWI](http://www.w3.org))
  - Display your logo on [W3C site](http://www.w3.org) (250K visits/day) and [W3C logo](http://www.w3.org) on your site

*Membership / Benefits / How to join W3C / "At a Glance" brochure*