Semantic Web, and Other W3C Technologies to Watch

Steve Bratt
Chief Executive Officer
World Wide Web Consortium

January 2007


Evolution of the Web
and the
Role of Standards
"Over 1 Billion Served"

<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 2000-2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>915,210,928</td>
<td>14.1 %</td>
<td>32,765,700</td>
<td>3.6 %</td>
<td>3.0 %</td>
<td>625.8 %</td>
</tr>
<tr>
<td>Asia</td>
<td>3,667,774,066</td>
<td>56.4 %</td>
<td>378,693,457</td>
<td>10.3 %</td>
<td>35.2 %</td>
<td>231.2 %</td>
</tr>
<tr>
<td>Europe</td>
<td>807,289,020</td>
<td>12.4 %</td>
<td>311,406,751</td>
<td>38.6 %</td>
<td>28.9 %</td>
<td>196.3 %</td>
</tr>
<tr>
<td>Middle East</td>
<td>190,084,161</td>
<td>2.9 %</td>
<td>19,028,400</td>
<td>10.0 %</td>
<td>1.8 %</td>
<td>479.3 %</td>
</tr>
<tr>
<td>North America</td>
<td>331,473,276</td>
<td>5.1 %</td>
<td>231,001,921</td>
<td>69.7 %</td>
<td>21.5 %</td>
<td>113.7 %</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>553,908,632</td>
<td>8.5 %</td>
<td>86,042,986</td>
<td>15.4 %</td>
<td>7.9 %</td>
<td>370.7 %</td>
</tr>
<tr>
<td>Oceania / Australia</td>
<td>33,956,977</td>
<td>0.5 %</td>
<td>18,364,772</td>
<td>54.1 %</td>
<td>1.7 %</td>
<td>141.0 %</td>
</tr>
<tr>
<td>WORLD TOTAL</td>
<td>8,499,697,080</td>
<td>100.0 %</td>
<td>1,076,203,987</td>
<td>16.6 %</td>
<td>100.0 %</td>
<td>198.1 %</td>
</tr>
</tbody>
</table>

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/ 135% growth, Chinese 14% w/ 347% growth per year)
"Over 100 Million Serving"

Number of Web Sites (domain names and content)


Potential to Break the Digital Divide


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 ...
- Tolerant - even w/ imperfect mark-up, data, links, software
- Universal - independent of systems and people
- Free / cheap - browsers, information, services
- Simple / powerful / fun for users - text, graphics, links
- Open standards ...
Why are Open Standards Important?

- Broad industry agreement (if done right)
- Interoperability ... cross-application, -organization, -data
- Avoids vendor lock-in ... for providers and users
- Open access = no black boxes
- Mandated ... by customers, government
- Open, royalty-free standards = good business sense
Founded by Tim Berners-Lee in 1994, W3C is:

- Providing the **Vision** to Lead
- **Engineering** the **Open Standards** that Make the Web Work ...
  - expanding ...
  - From a Web of **Documents** ...
  - Toward **One Web** ...
    - ... of **Data and Services**
    - ... on **Everything**
    - ... for **Everyone**
  - -- think Web 3.0
Engineering an Interoperable Foundation of the Web

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

Applications

- Interaction
  - XHTML
  - CSS
  - SVG
  - SMIL
  - CDF
  - XForms
  - MathML
  - InkML

- Mobile Web
  - XHTML Basic
  - Mobile SVG
  - SMIL Mobile
  - XForms Basic
  - CC/PP
  - DD

- Voice
  - VoiceXML
  - SRGS
  - SML
  - CCXML
  - EMMA

- Web Services
  - SOAP
  - MTOM
  - WSDL
  - WS-CDL
  - Addressing

- Semantic Web
  - OWL
  - SKOS

- Privacy, Security
  - P3P
  - APPEL
  - XML Sig
  - XML Enc
  - XKMS

Web Accessibility / Internationalization / Device Independence / Quality Assurance

- XML, Namespaces, Schemas, XQuery/XPath, XSLT, DOM, XML Base, XPointer, RDF/XML, SPARQL
- XML Infoset, RDF Graph
- Web Architectural Principles
- URI/IRI, HTTP

The Web Advancing to its Full Potential

Internet
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
Emerging W3C Technologies
Web for Everyone

Universal Access ...

- Global participation (Offices, translations, etc.)
- Internationalization Activity (overview talk)
- Web Accessibility Initiative
- New Web security work
- Future objective: "Trust"
Web on Everything

*The* User Interface, everywhere ...

- Interaction Technologies: HTML, XForms, CSS, MathML, Voice, Graphics, Multimedia, Multimodal
- Web 2.0 ≤ Rich Web Clients: Compound Doc Formats, Web Apps APIs (AJAX) and Formats
- Mobile Web Initiative
- Device Independence
- Ubiquitous Web (workshop, slides)
Web of Data & Services

- Interoperable information and programs ...
  - XML: Binary, Processing Model
  - Web of Services: Performance, Addressing, SemWeb Services, Policy (overview slides)
  - Web of Data = Semantic Web: Deployment, Query, Rules, Health Care/Life Sciences, Content Labeling, Geospatial, Multimedia Semantics
Introduction to the Semantic Web
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: Why?

- Tasks often require combining data across the Internet, e.g:
  - Integrating data across the enterprise
  - Hotel, transport, meeting, personal info come from different sites
  - Mining data from biochemical, genetic, pharmaceutical, patient databases
  - Cross-referencing disparate digital libraries

- Humans understand how to combine this information ...
  - Not always easy (different vocabularies, languages, formats)

- Machines aren't smart enough :-(


Why Can't Machines "Understand"?

Analogy:

- **What We Say to Dogs**
  - "Stay out of the garbage! Understand, Ginger? Stay out of the garbage!"

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

- 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
Semantic Web: Linked Data on the Web

Machine-processable, global Web standards:
- Assigning unambiguous names (URI)
- Expressing and linking data, including metadata (RDF)
- Capturing ontologies (OWL)
- Query, rules, transformations, deployment, application spaces (in progress) logic, proofs, trust

Semantic Web = Web 3.0? (Markoff, NYT, Nov 2006)

(W3C's Semantic Web Activity / Semantic Web overview slides / example)
Uniform Resource Identifiers (URI)

- **URIs** have two, very powerful functions:
  1. Unique, unambiguous name for something
  2. Where to find that something

- 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)
  - 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)

- Most useful, Web-wise, when these are URIs, e.g. for this presentation:
  - http://www.w3.org/2006/Talks/0811-sb-W3Cemergingtech/Overview.html as the subject
  - http://purl.org/dc/elements/1.1/creator as the property
  - http://www.w3.org/People/Bratt/stevenrbratt.rdf#SB as the value
Linking table and tree data...
Web Ontology Language (OWL) +

- W3C standard (Guide, Overview)
  - 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
- 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

[diagram with labeled nodes and relationships]

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

(see also more detailed example related to book searching and selling)
Enterprise Integration Today

- Issue tracking
- Human Resource Mgt
- order processing
- Stock control
- Parts Catalog
- ext
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
Life Science Ontologies Linked via Semantic Web
### Challenges Ahead

**Classic "chicken and egg problem" en route to "network effect"**

- Providing information in Semantic-Web-friendly ways
  - Exposing existing data stores as RDF
  - Automated and human-assisted tools to create RDF stores, ontologies, mappings

- 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

- Addressing broader operational and social needs
  - Universality, quality, provenance, versioning, safety, privacy, access control, authorization, trust
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:

- IngentaConnect bibliographic metadata storage: over 200 million triplets; RDF version of Wikipedia: more than 47 million triplets; Tracking the US Congress: data stored in RDF (around 25 million triplets); RDFS/OWL Representation of Wordnet: also downloadable as 150MB of RDF/XML, “Département/canton/commune” structure of France published by the French Statistical Institute

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/
Discussion
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 | •••• | España | Sverige | United Kingdom and Ireland | (map)

- **435 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)
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, workshops, Incubator Groups
- Influence standards through Working Group participation, review, implementation

Early insight into market trends
- Access world’s foremost Web technologists from Member & Team
- Plan for emerging technologies & markets through Member-confidential access

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

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