Why Web Standard are Important: An overview of W3C, its operation and current technical directions

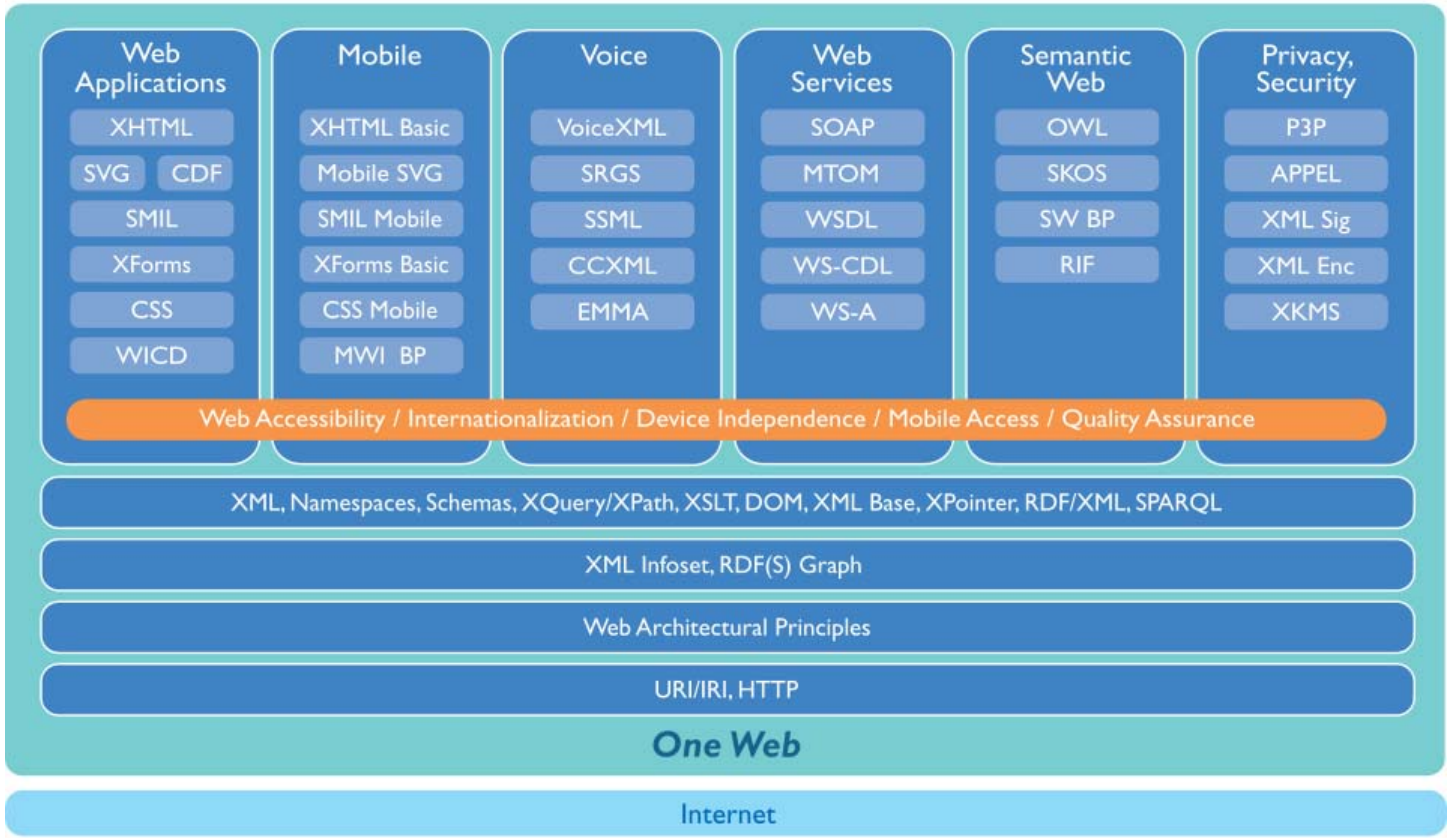
AusWeb 2006, Australia, 3rd July, 2006

World Wide Web Consortium (W3C)

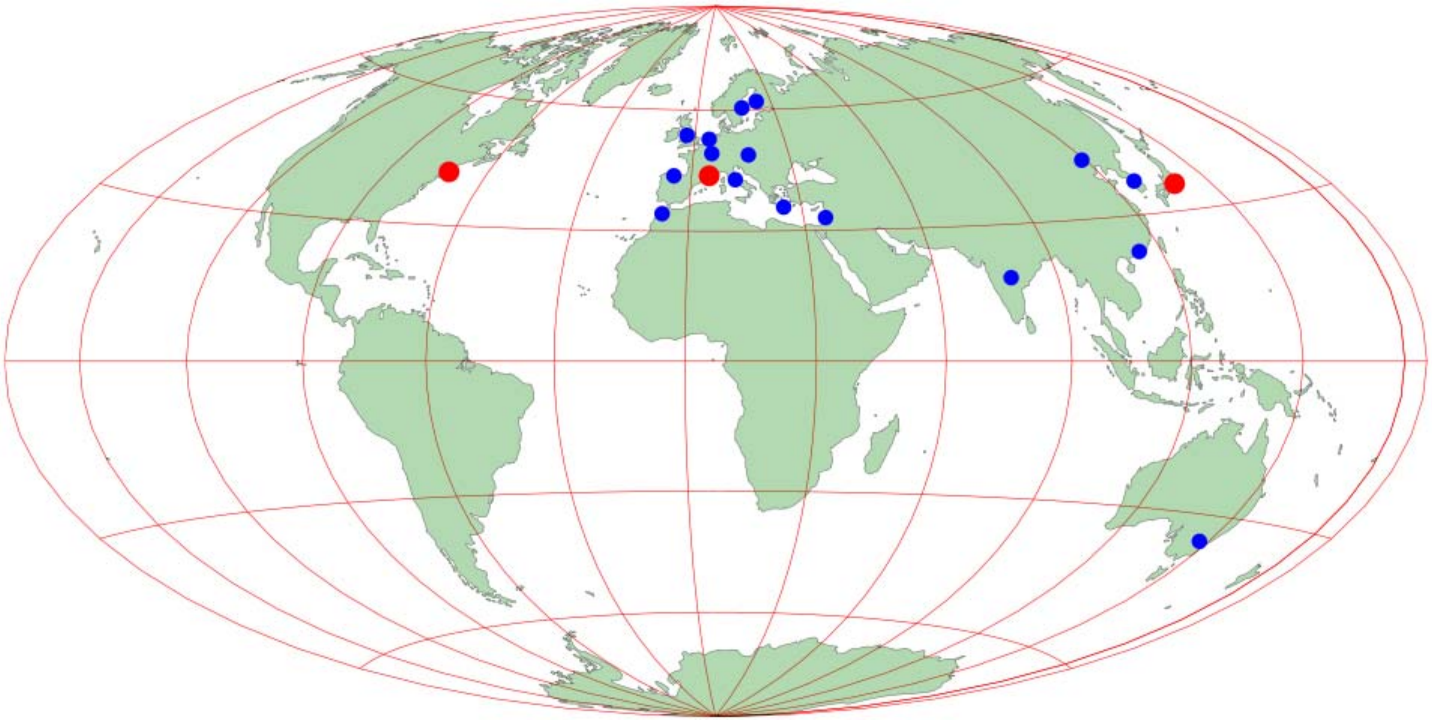
"To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web"

- Founded by Tim Berners-Lee in 1994
- Develops open *Recommendations* (Web Standards)
  - *over eighty so far…*
- Engages in education, outreach, develops guidelines…
- *A neutral forum for building consensus around Web standards*

Just a glimpse (we will come back to this later)…



W3C is international…



W3C Hosts (in red) and W3C Offices (in blue) around the Globe

Some guiding principles at W3C

- Web Technologies should be interoperable
  - *the Web is based on a large palette of technologies*
  - *no technology can pretend to cover all needs on the Web*
  - *hence the interoperability of technologies necessary*
- Web Standards should be open, i.e., non-proprietary
- The Web should be accessible to *all*

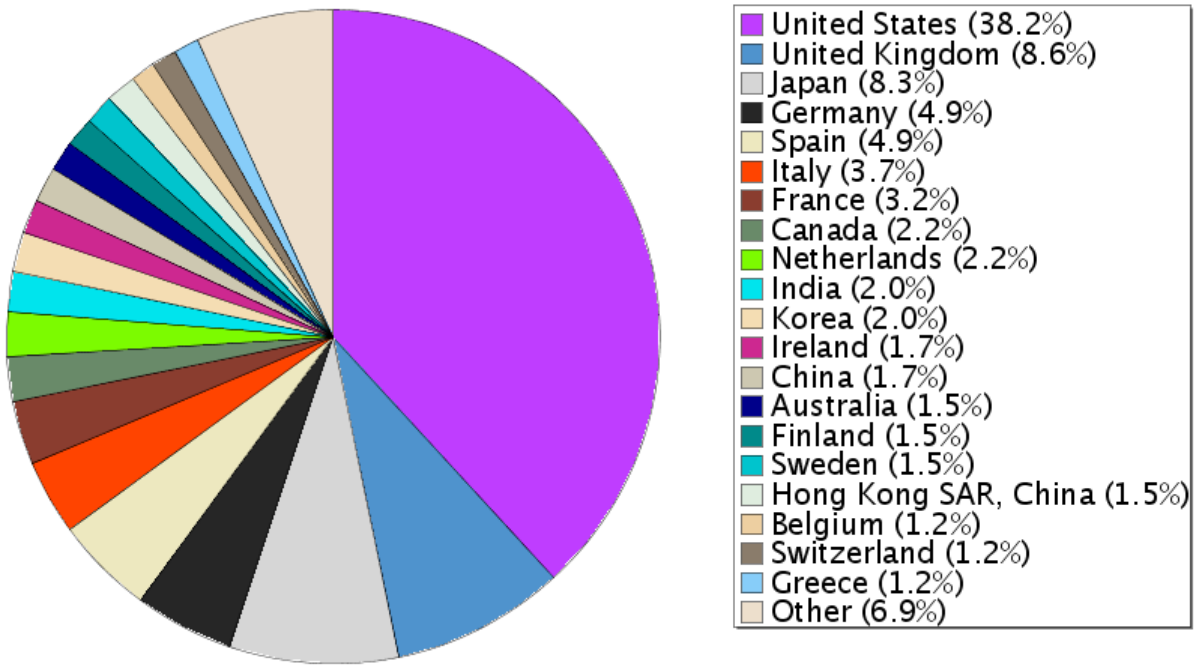
W3C's long term goals

- Web for Everyone
  - *regardless of language, user capabilities, geographical location, device used for access,…*
- Web on Everything
  - *not only PC-s, but Phones, PDA-s, Television,…*
- Knowledge Base, Advanced data searching and sharing
  - *information for both human and machine processing*
- Trust and Confidence
  - *technologies for collaborative environment*
  - *a Web with accountability, security, confidence, and confidentiality*

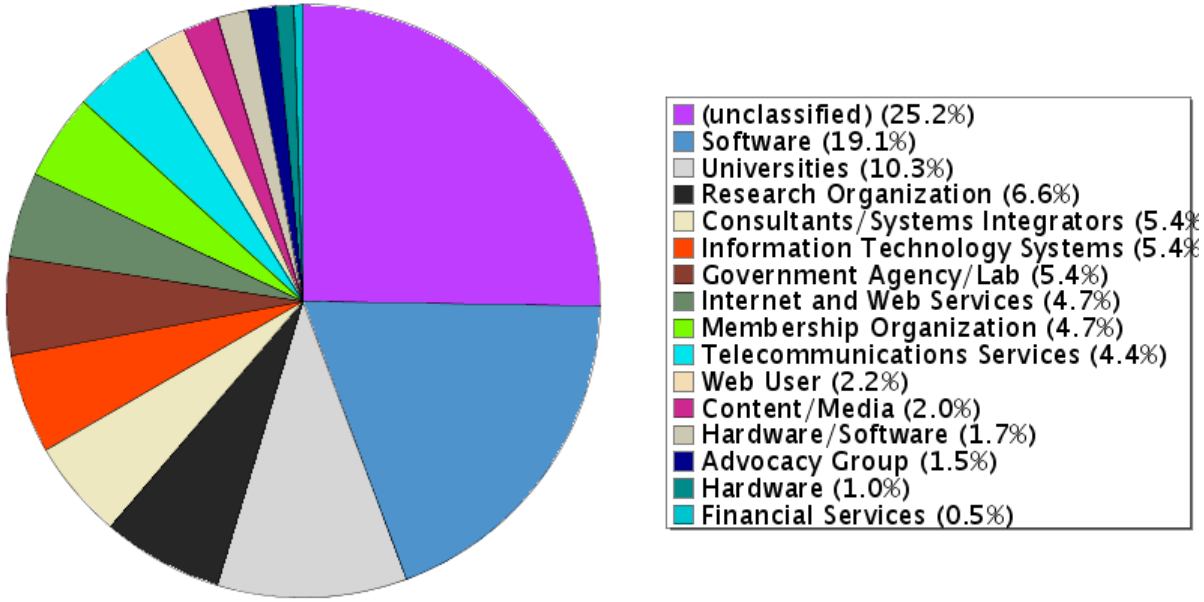
W3C members

- W3C Members ensure the strength of W3C
  - *they influence the strategic direction of Web Standard Development*
    - each member is represented in the *Advisory Committee* (AC)
    - the AC has regular meetings (twice a year) where issues are discussed
  - *the community of key players on the Web*
- Recommendations are developed by the Members' experts
  - *documents are developed in Working Groups staffed by the Members' representatives*
  - *altogether, they form a community of more than 600 experts*
  - *the keyword is consensus building*

Around 400 members from more than 28 countries…



… and with a wide activity spectrum



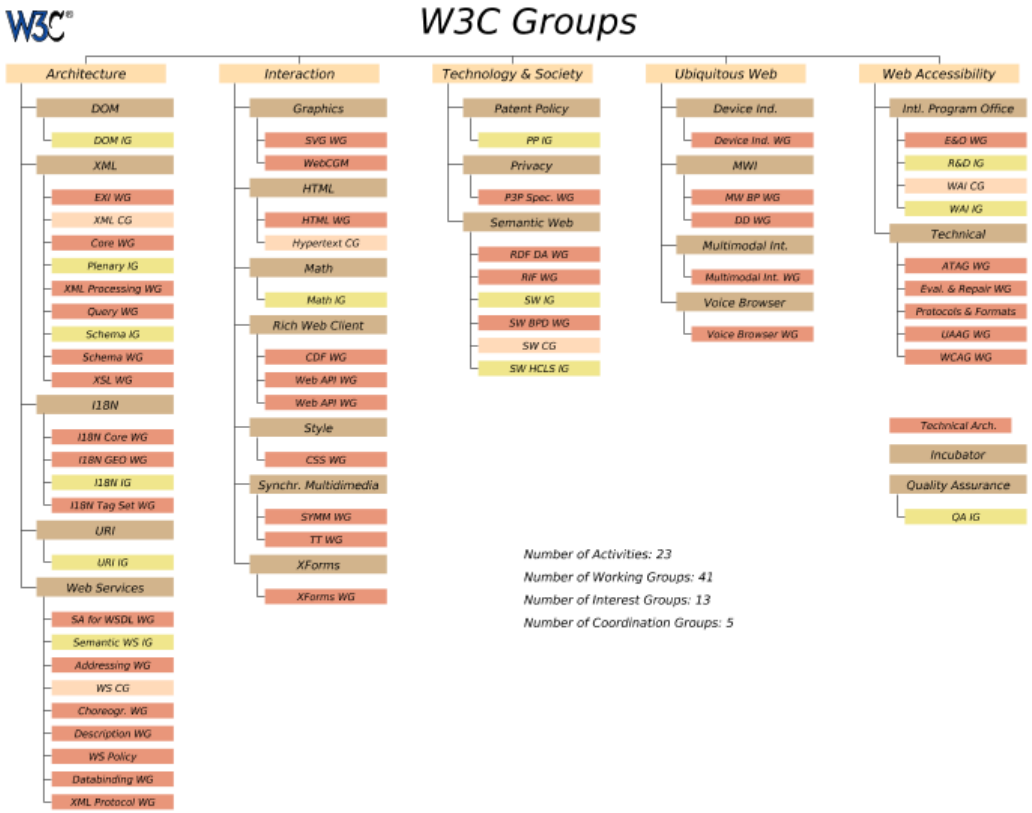
W3C staff

- More than 50 researcher and engineers
- Very international team (residence in 9 countries, around 12 nationalities…)
- Their role is:
  - *provide directions to W3C*
  - *coordinate the activities of W3C*
  - *facilitate active member participation*
  - *communicate the results of W3C*

Typical W3C work flow

1. A *W3C Workshop* is organized in an area of interest
   - *possible starting point for standardization*
   - *members can have members submissions that are taken into account, too*
2. A *Working group* (WG) is formed
   - *members have the possibility to review, and vote on the charter of the group (or to oppose its creation…)*
3. WG regularly publishes drafts to seek comments from the public
4. Implementations of the new technology are called for
5. Members review the final proposal
6. If final review is positive, W3C publishes the new Recommendation

W3C groups and activities



So, what do we *do* ?

Some highlights for this time

1. The “horizontals”
2. Mobile Web
3. Semantic Web

The “horizontals”…

The Web is for everybody!

- Regardless of language, culture, geographical location
- Regardless of user capabilities
- Regardless of device types and capabilities

Horizontal activities at W3C

- W3C has a number of activities to reinforce those principles
  - *"horizontal" review of all W3C technologies:*
    - internationalization, multimodality, accessibility, device independence, …
    - specification can be "sent back" to the drawing board if problems occur!
  - *separate education and outreach activities:*
    - tutorials, information for designers, quicktips, guidelines
      - some of those guidelines, like WCAG, are part of legislation in a number of countries!

Example: international text

Leading the Web to its Full Potential…
Duent la Web al seu ple potencial…
Het Web tot zijn volle potentieel ontwikkelen…
Amener le Web vers son plein potentiel…
Alle Möglichkeiten des Web erschließen…
Οδηγώντας τον παγκόμιο ιστό στο μέγιστο των δυνατοτήτων του…
Hogy kihasználhassuk a Web nyújtotta összes lehetőséget…
वेब की सम्पूर्ण क्षमता के उपयोग की दिशा में अग्रणी…
Sviluppare al massimo il potenziale del Web…
引发网络的全部潜能…
웹의 모든 잠재력을 이끌어 내기 위하여…
Levando a Web em direcção ao seu potencial màximo…
Раскрывая весь потенциал Сети…
Guiando la web hacia su máximo potencial…
Se till att Webben når sin fulla potential…
Ohjaamassa Webin kehittymistä täyteen mittaansa…
Webの可能性を最大限に導き出すために:
لإيصال الشبكة المعلوماتية إلىأقصى إمكانياتها…
להוביל את הרשת למיצוי הפוטנציאל שלה…

Example: international text (cont)

- One would think that this is only an issue of character set (e.g., Unicode)
- That is *not* the case:

The mobile web…

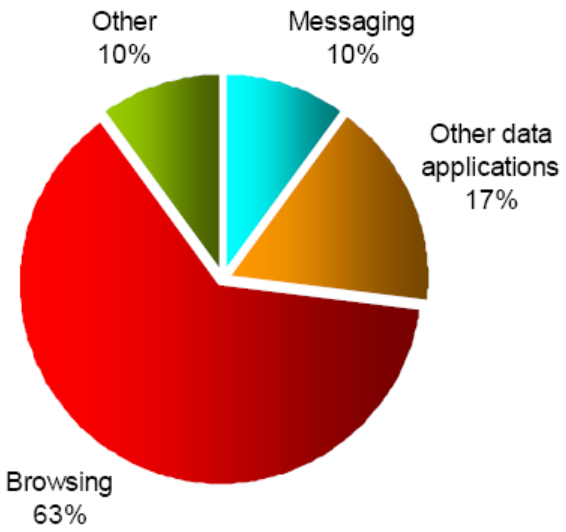
What is “mobile”?

- *Currently* W3C concentrates on mobile/cell phones and network aware PDA-s
  - *but a workshop on Ubiquitous Web took place earlier this year!*
- Question: what does W3C contribute to this environment?

Characteristics of mobile

- Extremely dynamic market
- Big business in Europe and Asia, with US catching up fast
  - *extremely dynamic market: ≈800M units sold in 2005, 63% of installed phones are Web capable (est.)*
- Potentially huge number of users
  - *40 Million new users per year in China alone!*
  - *future: one PC per family, but one (or more!) mobile per person…*
  - *Potentially huge number of users in developing countries (where, for many people, mobile is the only gateway to the Web/Internet!)*

Mobile web usage is growing (1)



Source: Nokia study, 2005 — Smartphones — Singapore, Germany, UK

Mobile web usage is growing (2)

- T–Mobile Web'n'Walk (a Web portal)
  - *330 page views per month per user*
  - *489% increase in data volume per user*
  - *199% increase in data access (excl. SMS)*
  - *source: Opera, April 2006*
- BBC
  - *number of requests to mobile content doubled in 2005*
  - *approaching 250 million/day*
  - *28% of mobile user only access BBC content from mobile phones, not from PC*
  - *source: BBC, November 2005*

It is multipolar World

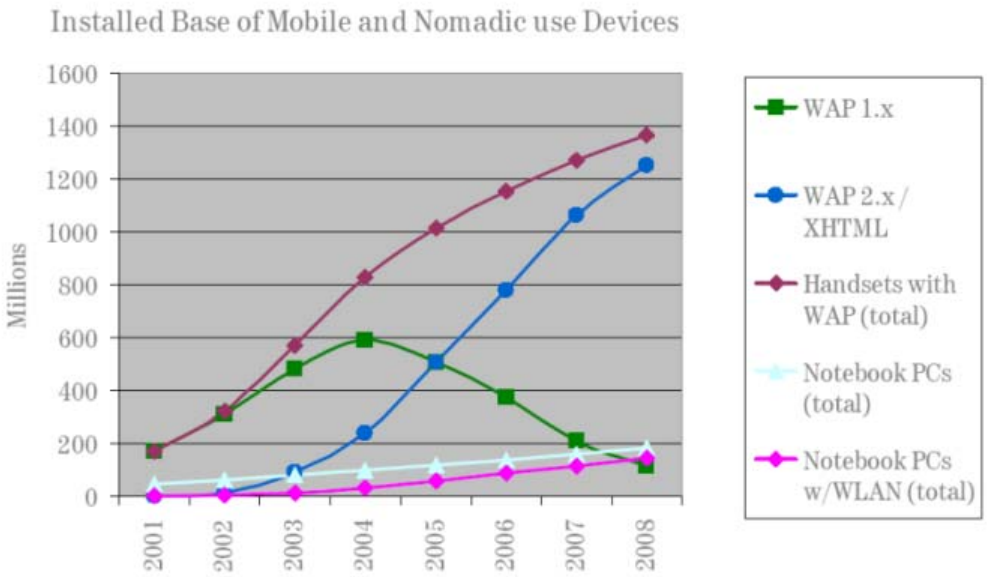
- Variety of hardware architectures
  - *Nokia, HP, Samsung, Palm, Motorola, DoCoMo, Sharp, SonyEricsson, KDDI, Sony, Dell, Sagem, Fujitsu, …*
  - *they represent different architectures, processors, displays, user interface styles, …*
- Operating systems evenly spread the field
  - *proprietary, Symbian, PalmOS, Windows Mobile/CE, Linux, …*
  - *none of them dominates!*
- Thriving software industry for all variants

A World of varieties

Where are we?

- But… we are still at the beginning
  - *systems and application software not always mature yet*
  - *infrastructure under constant development (eg, network)*
  - *more simplicity is needed for average user*
- Standardization is (even more) important!

Standardization is (even more) important!



Source: T-Mobile

The players

- Lots of hardware and software vendors (of course)
- Two main industry consortia outside W3C:
  - *Open Mobile Alliance (OMA):*
    - integrated some older consortia (WAP Forum, SyncML Initiative, …)
    - specifies interoperable technical specification for Mobile devices
  - *3rd Generation Partnership Project (3GPP)*
    - specifies technical specification for 3rd Generation GSM networks
  - *roughly: 3GPP is the radio, OMA is the application level*
    - but there are overlaps; they try to cooperate and synchronize

Position of W3C

- OMA and 3GPP often *integrate* existing technologies (when available and possible)
  - *only if the technology does not exist, do they define it themselves*
- W3C's expertise lies in the development of the basic Web architecture
  - *W3C provides already a number of "building blocks"; these are integrated in 3GPP/OMA specifications*
- Bottomline: there is good cooperation among W3C, OMA, and 3GPP

Example: XHTML/CSS

- XHTML Basic: a “minimized” profile of XHTML
  - *no frames, scripting; only simple tables (no colgroup, tbody/thead/tfoot, justification in cells)*
  - *had an early adoption for WAP 2*
- CSS Mobile: under development
- Important for simple devices
- For higher end devices, XHTML Basic may not be that relevant any more…
  - *there are browsers that can manage XHTML 1.1+CSS*

Example: SVG

- SVG has two “Mobile profiles”: Tiny and Basic
- Newer phones come with SVG built in (122 different types end of June 2006)
  - *some Web Browsers have SVG Tiny built in (Opera, NetFront, …)*
- W3C is working on SVG 1.2 Tiny (in strong synchrony with 3GPP)
- SVG Mobile becomes *the* vector graphics tool for Mobile!

Example: XForms

- XForms aims at an enhancement of traditional HTML forms
- XForms (full) is a W3C Recommendation
- XForms Basic should become a Rec later this year
  - *e.g., very restricted requirements on XML Schemas*

(courtesy of solidapp.com)

W3C's Mobile Web Initiative (MWI)

"Making Web access from a mobile device as simple, easy and convenient as Web access from a desktop device"

- Complements the work of OMA and 3GPP and the work done elsewhere at W3C
- Launched in May 2005 with a separate set of directed sponsorship
- The general approach:
  - *solve interoperability and usability issues for end users and content providers*
  - *not geared at new technology*
  - *explain how to use existing technology and improve implementations*

Mobile Web Best Practices working group

- Audience: Web content providers/Web developers
- Issue: how to make Web content work on mobile devices?
  - *rules to follow*
  - *things to look out for*

Best practices

- Studied existing “tips and tricks” (W3C Accessibility, iMode, Opera, Openwave, Nokia,…)
- 60 “Best Practices”; examples:
  - *thematic consistency/“One Web”*
  - *no table for layout, no spacers-GIFs, no frames*
  - *screen estate constraints: small top navigation, avoid large graphics*
    - has an overview of the typical current set of devices
  - *keep URI-s for sites short*
  - *scrolling should be in one direction*
  - *…*
- Close-to-final release issued last week!

Device Description working group

- Issue: how do I reliably find out the technical characteristics of a device?
  - *currently: all providers make their own testing*
- Device description needed for content adaptation
- Ongoing Work
  - *“landscape” document: survey of existing technology*
  - *“ecosystem” document: understand who does what and why*
- Probable future work: shared, open device description database

Potential future works at MWI

- “MobileOK” validator
- Device Description Database
- Test suites
- Training
- …

Semantic Web…

Problems leading to the Semantic Web…

- Tasks often require to *combine* data on the Web:
  - *hotel and travel infos may come from different sites*
  - *searches in different digital libraries*
  - *various databases within an organization (eg, after company mergers)*
  - *etc.*
- Humans combine these information easily, even if different terminologies, terms, languages, etc, are used…
- Machines have real problems with that!

Example: automatic airline reservation

- Your automatic airline reservation
  - *knows about your preferences*
  - *builds up knowledge base using your past*
  - *can combine the local knowledge with remote services:*
    - airline preferences
    - dietary requirements
    - calendaring
    - etc
- It communicates with *remote* information (i.e., on the Web!)
- (M. Dertouzos: The Unfinished Revolution)

Example: data(base) integration

- Databases are very different in structure, in content
- Lots of applications require managing *several* databases
  - *after company mergers*
  - *combination of administrative data for e-Government*
  - *biochemical, genetic, pharmaceutical research*
  - *etc.*
- Most of these data are now on the Web (though not necessarily public yet)

What is needed?

- Data should be available on the Web for further processing by other machines and programs
- Data should be possibly merged, connected, combined on a Web scale
- Sometimes, data may describe other data (e.g, using metadata)…
- … but sometimes the data is to be exchanged by itself, like a calendar or travel preferences
- Machines may also need to *reason* about that data
- *The "Semantic Web" is an infrastructure extending the current Web for the interchange and the integration of data on the Web,*

What is needed (technically)?

- To make data machine processable, we need:
  - *unambiguous names for resources (that may also bind data to real world objects): URI-s*
  - *a common data model to access, connect, describe the resources: RDF*
  - *access to that data: SPARQL*
  - *define common vocabularies, ontologies: RDFS, OWL, SKOS*
  - *…*

RDF triples

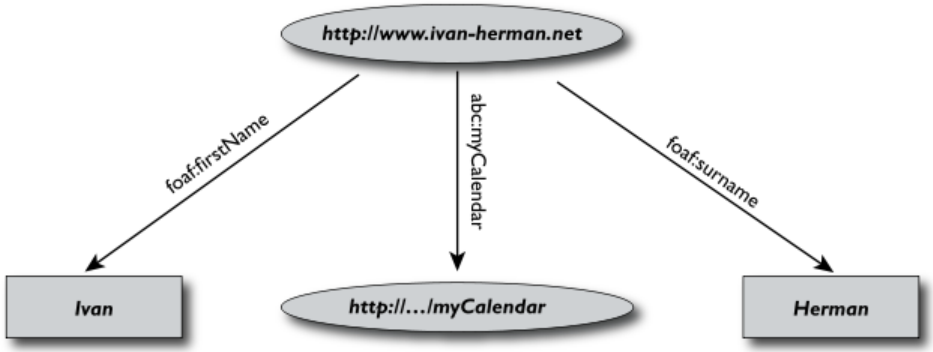
- We said “interchange” and “connection” of data… ie, resources have to be *connected*
- But a simple connection is not enough… it should be named somehow
  - *a connection from me to my calendar is not the same as the connection from me to my CV (even if all of these are on the Web)*
  - *the first connection should somehow say “myCalendar”", the second “myCV”*
- Hence the RDF Triples: *a labelled connection between two resources*

RDF triples (cont.)

(http://www.ivan-herman.net, http://…/myCalendar, http://…/calendar)

- This triple connects my home site with my calendar, using a `myCalendar` "predicate"
  - *note that URIs are also used to name the connection itself*
- *RDF* is a general model for such triples
  - *… with machine readable formats (RDF/XML, Turtle, n3, RXR, …), where RDF/XML is the "official" format*

A simple RDF example




```
<rdf:Description rdf:about="http://www.ivan-herman.net">
    <foaf:name>Ivan</foaf:name>
    <abc:myCalendar rdf:resource="http://…/myCalendar"/>
    <foaf:surname>Herman</foaf:surname>
</rdf:Description>
```

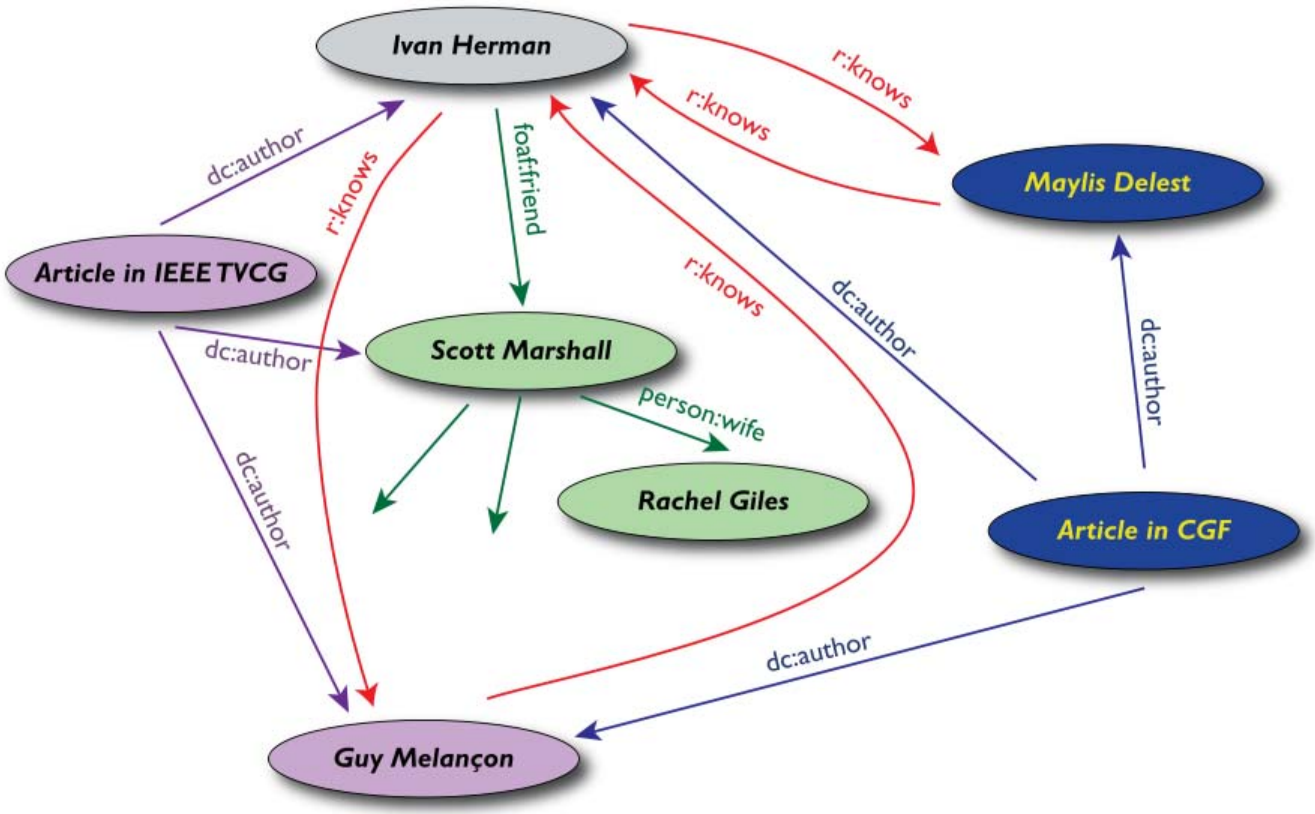
URI-s play a fundamental role

- *Anybody* can create (meta)data on *any* resource on the Web
  - *e.g., the same SVG file could be annotated through other terms*
  - *semantics is added to existing Web resources via URI-s*
  - *URI-s make it possible to link (via properties) data with one another*
- *URI-s ground RDF into the Web*
  - *information can be retrieved using existing tools*
  - *this makes the "Semantic Web", well… "Semantic Web"*

URI-s: merging

- It becomes easy to *merge* data
  - *e.g., applications may merge the SVG annotations*
- Merge can be done because statements refer to the *same* URI-s
  - *nodes with identical URI-s are considered identical*
- Merging is a *very* powerful feature of RDF
  - *metadata may be defined by several (independent) parties…*
  - *…and combined by an application*
  - *one of the areas where RDF is much handier than pure XML in many applications*

What merge can do...



RDF may not be enough…

- Creating data and using it from a program works, provided the program *knows* what terms to use!
- We used terms like:
  - `foaf:name`, `abc:myCalendar`, `foaf:surname`, …
  - *etc*
- Are they all known? Are they all correct? (it is a bit like defining record types for a database)

Possible issues to handle

- What are the possible terms?
  - *"is the set of data terms known to the program?"*
- Are the properties used correctly?
  - *"do they make sense for the resources?"*
- Can a program *reason* about some terms? Eg:
  - *"if «A» is left of «B» and «B» is left of «C», is «A» left of «C»?"*
  - *obviously true for humans, not obvious for a program …*
  - *… programs should be able to deduce such statements*
- If somebody else defines a set of terms: are they the same?
  - *clearly an issue in an international context*

Ontologies

- The Semantic Web needs a support of *ontologies*:

"defines the concepts and relationships used to describe and represent an area of knowledge"

- We need a *Web Ontologies Language* to define:
  - *the terminology used in a specific context*
  - *possible constraints on properties*
  - *the logical characteristics of properties*
  - *the equivalence of terms across ontologies*
  - *etc*
- This is done by RDFS (RDF Schemas) and OWL (Web Ontology Language)

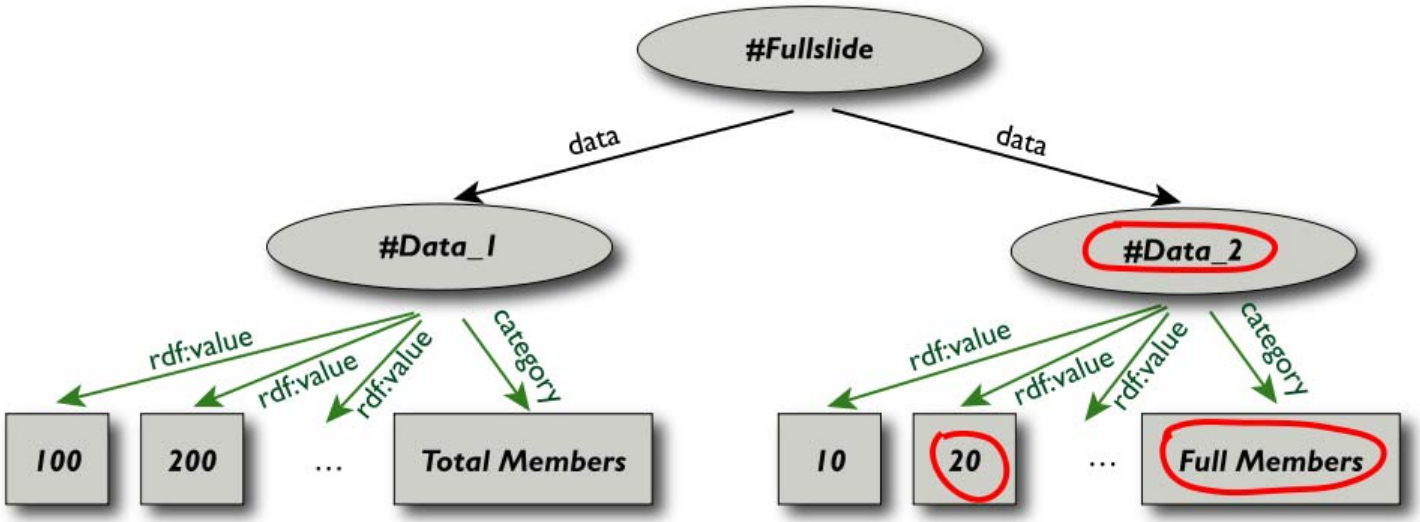
The newest element in the puzzle: SPARQL

- A query language for RDF
- RDF is a graph… SPARQL is based on *graph patterns* (i.e.: small graphs with unbound variables)

Simple SPARQL example

```
SELECT ?cat ?val # note: not ?x!
WHERE { ?x rdf:value ?val. ?x category ?cat }
```

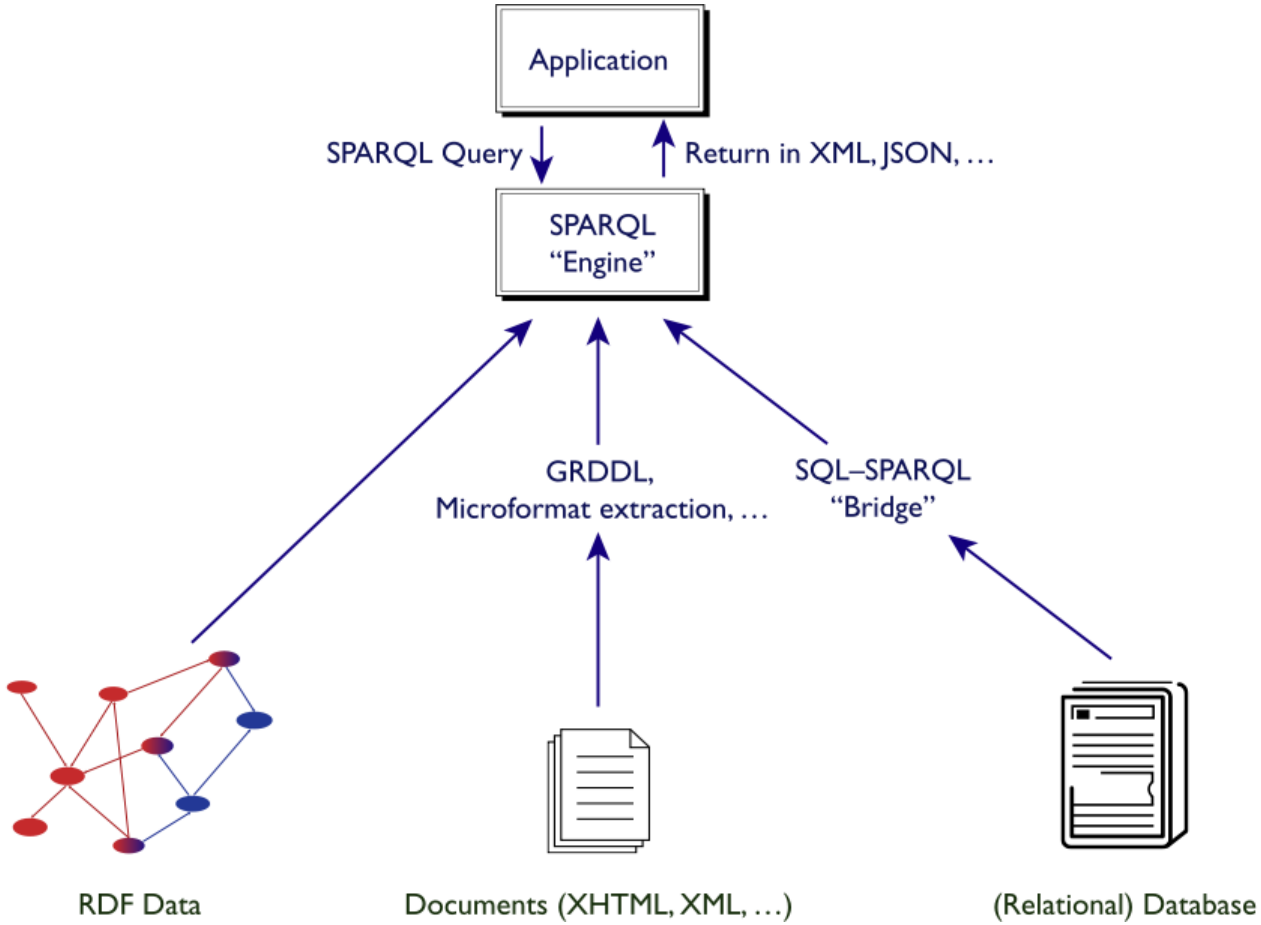
- Returns: `[["Total Members",100],["Total Members",200],…,["Full Members",10],…]`



SPARQL usage in practice

- *Locally*, i.e., bound to a programming environments
- *Remotely*, e.g., over the network or into a database
  - *separate documents define the protocol and the result format*
    - SPARQL Protocol for RDF with HTTP and SOAP bindings
    - SPARQL Results XML Format
    - there is also a JSON binding (though this is not a W3C document…)
- There are already a number of applications, demos, etc.,

SPARQL usage in practice (cont.)



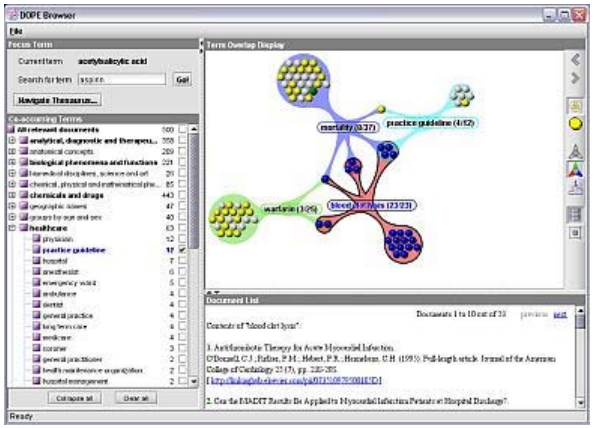
Some SW application examples

Example: portals

- Vodafone's Live Mobile Portal
  - *search application (e.g. ringtone, game, picture) using RDF*
    - page views per download decreased 50%
    - ringtone up 20% in 2 months
- Sun's SwordFish: public queries for support, handbooks, etc, go through an internal RDF engine for White Paper Collections and System Handbook collections
- Nokia has a somewhat similar support portal

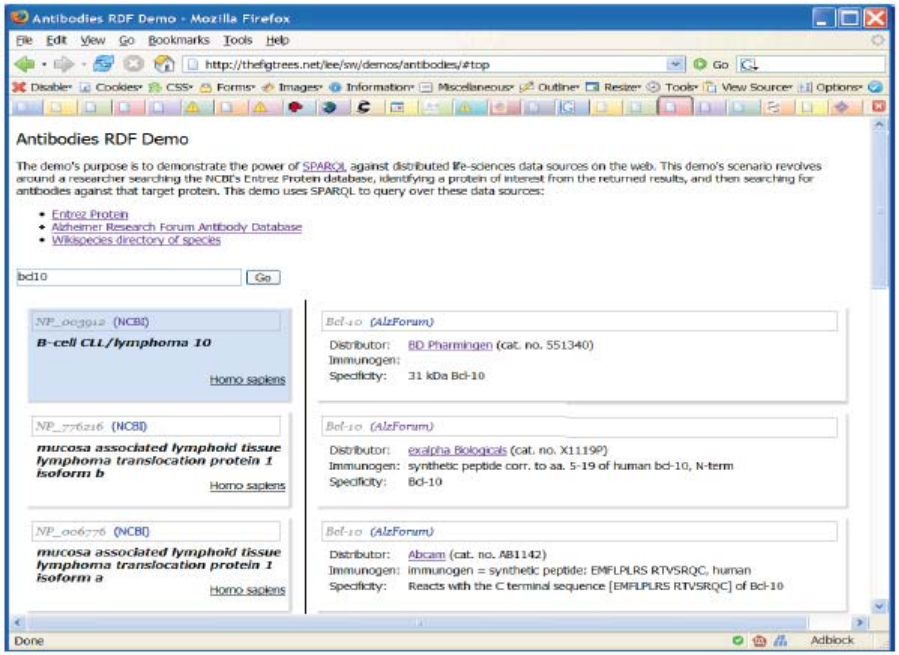
Example: data integration

- Semantic integration of different data sources
- RDF/RDFS (possibly with OWL and/or SKOS) based vocabularies as an “interlingua” among system components
- Many different projects and R&D on this: Boeing, MITRE Corp., Elsevier, EU Projects like Sculpteur and Artiste, national projects like MuseoSuomi, …



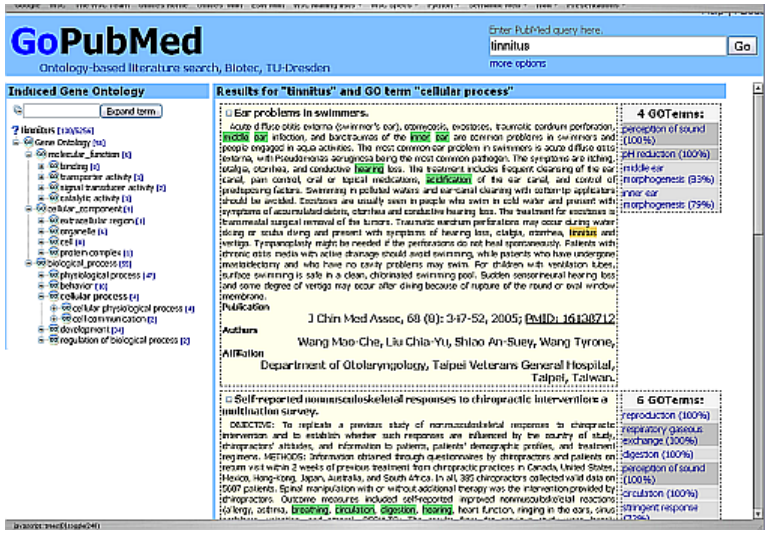
Example: Antibodies Demo

- Scenario: find the known antibodies for a protein in a specific species
- Combine four different data sources
- Use SPARQL as an integration tool



Example: improved search via ontology: GoPubMed

- Improved search on top of pubmed.org
- Search results are ranked using the specialized ontologies
- Extra search terms are generated and terms are highlighted
- Importance of *domain specific ontologies* for search improvement



Further information

More information about W3C:

http://www.w3.org/Consortium/

Contact the Australia Office:

http://www.w3c.org.au/

Mail me:

ivan@w3.org

Thanks you for your attention!