



The future of applications: W3C TAG perspectives

Henry S. Thompson
School of Informatics
University of Edinburgh

W3C Technical Architecture Group
28 March 2011

The World Wide Web Consortium



Founded by Tim Berners-Lee to sustain the value of the Web for everyone and "lead the Web to its full potential"

Hosted by MIT in US, Keio University in Japan and ERCIM in Europe
~60 employees, about half working remotely

Over 300 members: organisations and companies

Recently awarded substantial support from ISOC

Standards work carried out by Working Groups made up of member representatives and invited experts

100s of existing Recommendations, dozens in progress from >50 WGs

Mostly 'horizontal' technologies

- HTML (itself, Forms, CSS)
- XML (itself, XSL T, Schema, Query, Processing Model, Linking)

- Multimedia (PNG, SVG, SMIL, VoiceXML)
- Mobile (Mobile Web Initiative ('one web'), Mobile CSS, Geospatial (!))
- Semantic Web (RDF, OWL, Rules)
- Accessibility
- Internationalisation

The Technical Architecture Group of the W3C



Originally the Director (Tim B-L) was responsible for maintaining consistency across activities

As the Consortium grew and the scope of its activities enlarged, this task became impossible

The TAG was established eight years ago to pick up and broaden this task

- Identify and promote core principles of Web Architecture
- Look backwards at what has made the Web succeed
- And forwards to what must be done to protect that success
- While embracing as much as possible that emerges

And, perhaps most importantly for *tonight's* panel

[H]elp coordinate cross-technology architecture
developments inside and outside W3C

(From the TAG's formal charter)

TAG membership



Tim Berners-Lee *ex officio*

Three appointed members

- Noah Mendelsohn (independent) (chair)
- Jonathan Rees (Creative Commons)
- Jeni Tennison (independent)

Five elected

- Dan Appelquist (Vodafone, London)
- Ashok Malhotra (Oracle, New York)
- Peter Linss (Hewlett-Packard, San Diego)
- Henry S. Thompson (University of Edinburgh)

- Larry Masinter (Adobe, California)

Plus Yves Lafon, staff contact

Weekly 'phone conferences, four face-to-face meetings yearly

A parenthesis on the meaning of the word 'architecture'



I suspect that the IETF community and the W3C community have different initial examples in mind when they hear the word *architecture*

"The *architecture* of the Internet"

- Primarily something concrete:
 - Physical connectivity
 - WANs, LANs, packet routing
 - Hard-edged requirements

"The *architecture* of the World Wide Web"

- Primarily something abstract:
 - Specification connectivity
 - Social contracts for responsible/effective use

- Best practices

Architecture of the World Wide Web



A grandiose concept

- In practice an ongoing *post-hoc* analysis project with respect to a vast, messy, poorly delimited real-world phenomenon

A document

- The first publication from the **Technical Architecture Group** of the W3C (World Wide Web Consortium)
- [Architecture of the World Wide Web, Volume One](#)
- The goal of the document is to "preserve [the properties underlying the success] of the information space as the technologies evolve"
- It contains a number of "Principles, Constraints and Good Practice Notes"

Grandmother observations about the Web



"Global naming leads to global network effects"

- A prerequisite for citation
- A challenge for many existing schemes

"To benefit from and increase the value of the World Wide Web, agents should provide [http:] URIs as identifiers for resources"

- I.e. not just *any* global name, but a particular kind of global name

More from Grandma Webarch



It's good for the ownership of a name (URI) to be manifest in the form of that URI

- We need to know who is responsible for the name and its use

"A URI owner should not associate arbitrarily different URIs with the same resource."

"Agents do not incur obligations by retrieving a representation."

- HTTP 'GET' is side-effect free (c.f. cookies)

"A URI owner should provide representations of the resource it identifies"

In a nutshell



The Web works because you can

- GET any URI from anywhere
- View Source
- Follow your nose
- Write URIs on the side of a bus
- Use generic tools
- Redirect, cache and proxy
- Crawl and index everything

Web architecture and applications



All those nice pithy observations

- Were written about a web of *documents*
- Every one of them needs to be re-assessed as we move towards a web of documents, data *and* applications

The Web is evolving

- Web architecture has to evolve with it

The rise and rise of port 80



HTTP, HTML and the browser have come to increasingly dominate the Internet

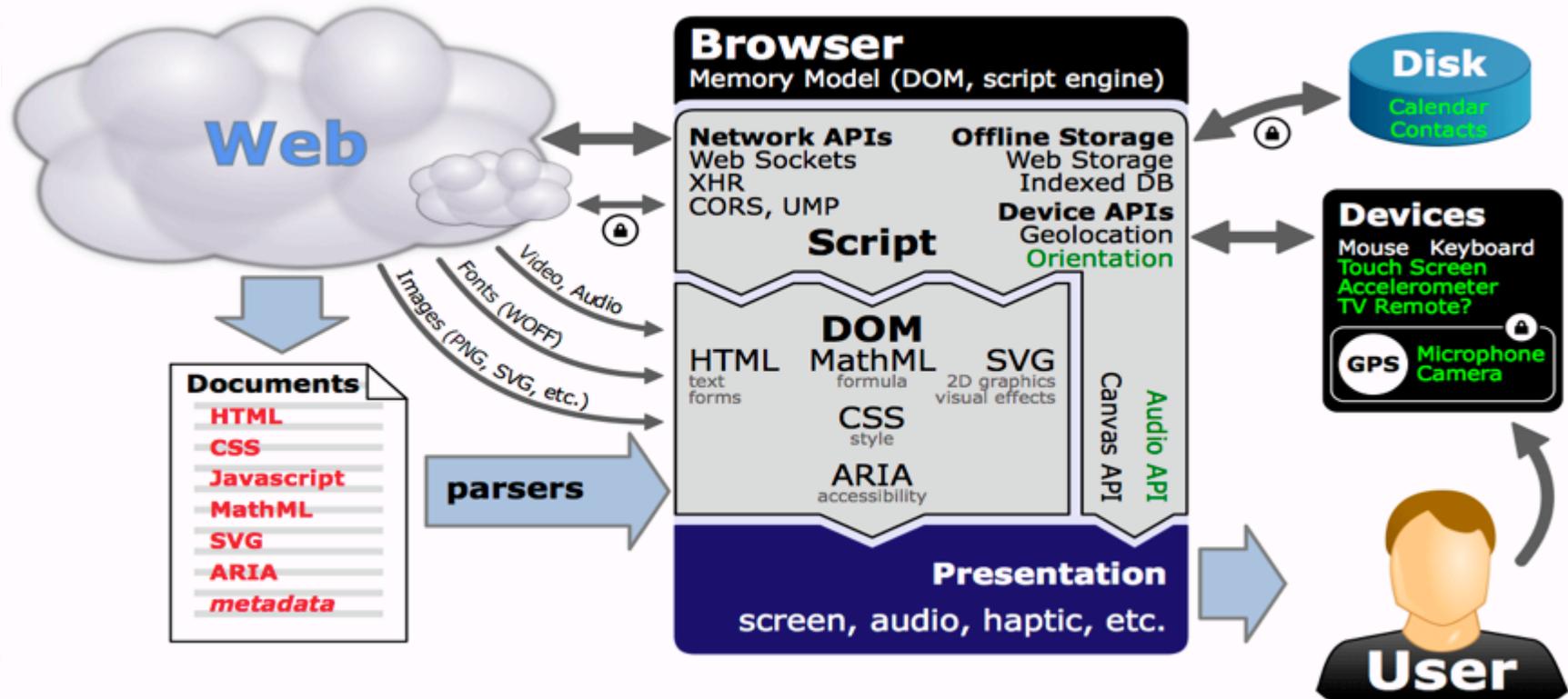
- FTP, NNTP and even SMTP have been supplanted
- The browser has become the information appliance of choice for a rapidly growing portion of users

Tensions and problems have arisen because HTTP, HTML and the browser were not *designed* to be a distributed application delivery platform

- But it's manifestly too late to stop and "get it right" from first principles
- So we have to manage their ongoing re-purposing as best we can

Hence the TAG's work on Web Application Architecture

The W3C vision of the Open Web Platform





Standardising the Open Web delivery platform

Expanding and clarifying HTML

- More Multimedia: `<video>`, `<audio>`, `<canvas>`, `<svg>`, etc.
- More Expressivity: `<article>`, `<section>`, `<figcaption>`, `<footer>`, `<math>`, forms, etc.
- The Goal: High Interoperability
- Targetting Q2 2011 for Last Call, 2014 for W3C Recommendation.
 - Many features implemented already
- Key for interoperability is *testing*
 - Lots of tests are needed
 - No organized testing of HTML 4 was ever done

There are many demos available: here's [a collection of linked \(click on 'next'\) HTML5 demo pages](#) from Philippe le Hégarret of W3C

Behind the scenes: Javascript APIs



Three layers of specification:

1. Javascript the language, aka ECMAScript

- ECMA TC 39 is in charge

2. Web IDL, for API definition

- W3C's Web Applications WG is in charge

3. Individual APIs

- Web DOM Core API, Drag Drop API, Text Selection API, Undo History API
- 2D Context API, Web Storage API, WebSockets API, Web Workers API
- Web Messaging API, Geolocation API, Indexed Database API, Microdata API

- RDFa API, Element Traversal API, XMLHttpRequest API, Web Notification API
- DOM Level 3 Events API, Navigation Timing API, Multi-touch Events API
- CSSOM View Module API, Selectors API, File API, (WebGL API)
- Resource Timing API, User Timing API, Messaging API, Device API

Maybe be client-side only

- Or involve server or peer interaction
 - Via HTTP
 - Or custom protocol

May be intended for implementation

- Natively, by the browser
- Via plugin
- Via Javascript

The TAG and the IAB



So far I've talked about how we saw ourselves

- From within our own W3C space

WebArch (the document) was (mostly) squarely within W3C territory

- But many of our more recent concerns have drawn us towards IAB territory
- So it's clearly time to give some thought to demarcation

Some overlap is unavoidable

- Indeed even beneficial

But it needs to be recognised and managed

Example 1: Scalability



Right on the edge between W3C and IETF

Several years ago, W3C discovered that the HTML DTD and other similar static resources are being accessed up to 500M times/day, often from the same client

Architectural issue:

- There is no normative specification that prohibits or even discourages repeated access to the same, cacheable, Web resource
- RFC 2616 says only

provide an explicit expiration time [...] indicating that a response MAY be used to satisfy subsequent requests

Practical issues:

- Large organizations can't find the software that's doing it, may not be in a position to rewrite/reconfigure code
- Blocking miscreants: can disable access from entire organizations
- Tell them to proxy: large corporations report expense of proxying all Web access can be >> savings

The W3C's current response:

- “Tarpitting”
- when repeated access to a given resource is seen from given source, delay before responding – W3C server team reports this is the best compromise so far

Is this a *Best Practice*?

- If so, it should be documented
- If not, what *should* we do?

- Is there, for instance, a connection with the ongoing discussion of registries?
- And who's "we" in this case?

The W3C recently published its [priorities](#) going forward:

1. Powerful Web Apps
2. Data and Service Integration
3. Web of Trust
4. Television, Mobile and the Web of Devices
5. One Web for All

The TAG has organised its current activities under four broad headings, which are broadly in line with those institutional goals:

1. The future of HTML
2. Privacy

3. Web Application Architecture

4. Core Mechanisms of the Web

New areas for Web/Internet Architecture



All of the following (relatively) new webapps-related TAG issues are, it seems to us, IAB issues as well:

- Use of URIs for identifying (parts of) application state
- Rethinking privacy
- Web Security as it becomes Application Security
- Device APIs: the difficulty of distinguishing between protocol and API

Some of these, e.g. privacy, are areas where active cooperation is already under way;

In some, e.g. IPv6, the TAG is happy to look to the IAB for guidance;

Others, e.g. the diversification of fragment identifier usage, are areas where the TAG has, at least so far, been driving

Example 2: Privacy



Some small concrete steps underway in this huge and poorly-delimited area

Minimization principle

- Send only what is requested (and no more); request only what is necessary (and no more)
- Mimimizes surface area of attack and potential for subsequent misuse
- Daniel Appelquist [API Minimization](#)(work in progress)

Do Not Track

- US FTC has been driving interest in DoNotTrack, to provide an opt out from Web usage tracking

to identify a user unnecessarily

- Browsers starting to implement
- [Workshop on Web Tracking and User Privacy](#) 28, 29 April, Princeton, USA, sponsored by W3C
 - Following on, in part, from joint IAB/W3C/ISOC [workshop on privacy](#) last year
- *Overview of Universal Opt-Out Mechanisms for Web Tracking* [draft-cooper-web-tracking-opt-outs-00](#)
- *Do Not Track: A Universal Third-Party Web Tracking Opt Out* [draft-mayer-do-not-track-00](#)
- Overview of state of play: [Web Tracking and User Privacy: The Next Steps](#), Thomas Roessler, 2011-03-10

The view for 10,000m: "[Information Accountability](#)" Abelson, Berners-Lee et. al., *CACM* v. **51**, No. **6**, 2008

- We've been thinking about privacy the wrong way
- Users cannot competently decide their privacy settings
 - They cannot determine the future consequences of their choices
- Instead of user-selected settings
 - A legal framework to protect users against misuse of private data

Example 3: Fragment identifiers and client-side state



Increasingly, what is displayed on the browser page is constructed by JavaScript on the client using bits retrieved from the Web in the background

- And the view is updated in response to user actions

To allow bookmarking, sharing, **back** functionality etc. requires the page URI to be updated to record such changes in browser state

- Many applications have used the URI's fragment identifier field to record state
 - One reason for doing this is because it avoids a page reload

But this in turn renders the distinction thus captured invisible to search engines

- Hence the recent exploration of the so-called hash-bang (#!)

convention

The situation is

- moving fast, mostly outside standards bodies
- incompatible with the existing media type registrations

Ashok Malhotra [Repurposing the Hash Sign for the New Web](#) (work in progress)

Thinking again about Web Architecture



Or, when does a quantitative difference become qualitative?

Browsers have always had to *compute* what people see on the screen

- Based on the bits that come over the wire

That hasn't changed

None-the-less the old naive picture

- Where a server responded to a GET with a bucket of bits it found in its file system
- And the client rendered those bits according to the specified media type
 - And `text/html` wasn't really very different in that regard from `image/jpeg`

is a best a pretty misleading way to think about things today

- And looks pretty certain to be even *less* helpful tomorrow

With so much of the action now taking place client-side

- Is it really true that the Web works because you can . . .
 - Crawl and index everything?

With so many new APIs/protocols

- Expressed as conventions for XML/JSON interchange
 - Dependent on a significant subset of the Open Web Platform stack
 - Thus significantly expanding the demands on client capabilities
- What happens to the "One Web for All" goal?

Conclusions



The IAB and the TAG have interests in common going forward

- Certainly in the area of Web Apps
- And this area is a (if not *the*) high-growth area right now
- Existing architecture analyses and documents give us a starting point
 - But need to be re-examined, possibly modified, almost certainly augmented, to encompass *web application* architecture

A monthly W3C-IETF liaison call, plus one or two by-chance overlaps in personnel, are probably insufficient to manage the cooperation which is necessary to address these common interests effectively.

- The W3C in general, and the TAG in particular, are looking for new and better ways to co-operate

At the end of the day, there is only *one* reason for a standards body to exist

- Interoperability

The IETF and the W3C should and, I hope *will*, do better at *institutional* interop :-)