Internet Standards for the Web: Part I

Larry Masinter
April 1998
Outline of tutorial

- Part 1: Current State
  - Standards organizations & process
  - Overview of web-related standards

- Part 2: Recent activities
  - What’s happening with web standards?
  - What are the hard problems
Purpose of Part I

- What’s a “standard”?
- How are standards made?
- What are the standards for the web?
- Introduce terms and set in context
What’s the World Wide Web?

- One network, everyone on it
- Multiple media
- Multiple modes of communication
What’s **important?**

*One Network, Everyone On It*
All kinds of “media”

- documents
- video
- music
- interactions

Documents, videos, worlds, music
Many modes of communication

- publish
- broadcast
- interact
“The nice thing about standards…”

- … there are so many of them to choose from
- … by the time things become standards, they're obsolete.
- … real standards are set by the market, not committees

Standards the only way that everyone can play

tragedy of the common

“where do you want to go today?”
Standards follow Innovation

Innovation, Divergence

Standardization, Convergence
Standards vs. Design

- **Design**
  - choose between alternatives (A, B, or C)
  - optimize function, performance, reliability

- **Standard:**
  - choose one, some, all, “undefined”, “implementation dependent”, “discoverable”
  - optimize flexibility, interoperability, politics, extensibility, enforced cooperation
Who writes web standards?

- Standards organizations
- Consortia
- Companies
- Individuals
Internet Engineering Task Force

- Defines standards for the Internet
- Different rules, structure than most other standards organizations
Internet Society

- Non-governmental organization created to coordinate Internet activities
- Umbrella organization for IETF
IETF structure

Internet Architecture Board (IAB)

Internet Engineering Steering Group (IESG)

Applications
- HTTP
- HTML
- URC
- ...12 WG

Security
- WTS
- PKIX
- ...6 WG

Transport
- ...6 WG

User Services
- ...7 WG

...9 areas
IETF Working Groups

- Open organizations
  - no formal membership, all volunteer

- Most work happens via email
  - may meet at IETF meetings (3 a year)

- Small focused efforts
  - published goals and milestones

- No formal voting
  - “Rough consensus and running code”
IETF Documents

- Internet-Drafts
  - works in progress, no formal status
  - deleted after 6 months

- RFCs (Request For Comments)
  - Archived series of documents
  - RFC 1796: “Not all RFCs are Standards”
IETF RFC Categories and Process

Standards Track

**Proposed Standard**
- complete, credible specification
- demonstrated utility

6 months - 2 years

**Draft Standard**
- multiple independent interoperable implementations

4 months - 2 years

**Standard**
- operational stability

Other Categories

**Experimental**
- not ready for standards track

**Informational**
- Important but not standards track

**Historic**
- superseded or otherwise unused
World Wide Web Consortium

- Members are vendors and user organizations
- Paid (and volunteer) staff
- *Develops* web-related standards
- Hosts workshops, working groups
W3C and IETF relationship

- W3C focus on Web; IETF general Internet
- W3C *researches and develops* protocols
- IETF *reviews and resolves* standards
- W3C staff participate actively in IETF
W3C Process

- Advisory Committee sets priorities
- New work requires member approval
- Exploratory workshops spawn working groups
- Working groups are closed
  - outside experts invited, though
Other groups setting standards

- Consortia, vendor groups, trade groups
  - European Computer Manufacturer’s Association (ECMA)
  - CommerceNet
  - Internet Messaging Consortium

- ANSI, ISO, ITU
Categories for Web Standards

- Content (e.g., HTML)
  - kinds of objects we’re moving around?
- References (e.g., URLs)
  - how to talk about something not in hand?
- Protocols (e.g., HTTP)
  - how do things move around the net?
Standards for Content

- Packaging
- Documents
- Images
- Media: audio, video, music
- Interactive content
- Metadata
Why standards for content?

- Preservation
  - Can you read Word 2.4 files?
- Interoperability
  - Multiple implementations
- Global communication
  - Standards designed for consistency over features
Content Packaging: labeling data

- **MIME:**
  - Multipurpose Internet Mail Exchange
    - Originally designed for mail

- Allows
  - Multiple media
  - Multiple character sets
  - Multiple languages

*MIME media types*
Internet Media Types (‘‘MIME types’’)

- Standard way of naming data formats
- Hierarchical structure with parameters
- Applications use MIME to decide how to interpret data (instead of file extension)
MIME Major Types

- **text**: sequences of characters
- **image**: bitmaps in various forms, e.g., gif, jpeg, tiff, png
- **audio**: sounds in various forms
- **video**: animations
- **message, multipart**: special purpose
- **application**: catch-all
MIME subtype

- Standard registry: “image/tiff”, “application/postscript”
- Registry rules: security, both standard & private (vnd)
- “application/vnd.ms-word”
Standards for Web Document formats

- HTML, SGML and XML
- Page layout: PDF
- Proprietary application formats
  (word, wordperfect, etc.)
SGML and XML

- Standard Generalized Markup Language
- A way of writing
  (ways of writing documents)
- DTD (Document Type Definition)
  defines elements and rules about them
- XML (from W3C) is simplification
Markup: saying things about parts

- Semantic markup
  `<part-no>N1025B</part-no>`

- Structural markup
  `<H1>N1025B</H1>`

- Presentation markup
  `<font face=aslan>N1025B</font>`
HyperText Markup Language (HTML)

- An application of SGML (more or less)
- A way of writing text
  that includes links
  and (mainly) structural markup
  with some other things (like images) embedded.
HTML design goals

- **lingua franca** for the web
- Hypertext views of existing documents
- Simple, scaleable
- Platform independent
- Support for visually impaired
- Interoperability with common editors
HTML standards

- 1994: 2.0 (baseline) RFC 1866
- 1996: 3.2 (tables, forms, presentation)
- 1998: 4.0 (style sheets, lots more) W3C Recommendation
HTML/4.0

- More complete tables
- File Upload
- Internationalization
- Embedded objects
- Extensions
- Style sheets
Beyond HTML: XML

- simplification of SGML
- Allows multiple kinds of documents, separate semantics from presentation

Why XML?

- Think beyond this year
  - Can you read Word 3.2 documents?
- Think beyond the PC
  - Different devices
  - Different uses (searching, indexing, translation)
Character sets: beyond ASCII

- European languages: ISO-8859-1 (Latin 1)
- The rest of the world: variety of systems
- Identifying the charset used: a registry
- A single charset? Unicode (UTF-8)
Other content on the web

- Images
- Page layout
- Video
- Audio
Images on the Web

- **gif**: Graphics Interchange Format
  - 8-bit color, transparent areas; patent cloud
- **jpeg**: Joint Photographic Expert Group
  - Lossy compression for photos, not line art
- **tiff**: Tagged Image File Format
  - Issues over tag standardization
- **png**: Portable Network Graphics
  - Calibration, hypertext links
Page layout on the Web

- **Postscript**
  - Designed for printer control
  - `application/postscript`

- **Portable Document Format (PDF)**
  - Useful for screen presentation and printing with exact layout
  - `application/pdf`
Video formats on the Web

- MPEG
- QuickTime
- AVI
Audio and Music

- audio/basic
- Audio hasn’t taken off
- MIDI and music unevenly deployed
More web content-types

- Desktop applications
  - Word, Excel, etc.
- 3-D renderings
  - VRML, etc
- Active content
  - Java
  - JavaScript, Document Object Model
Standards for MetaData and the Web

- Cataloging (Dublin Core)
- Ratings (PICs)
- Digital Signatures (proving authenticity)
- Copyright (who owns this material?)
Identifiers in the Web: URIs

- **URL**: locations
  - *New York Public Library, second floor, third aisle, second shelf, third book from left*

- **URN**: location-independent names

- **URC**: descriptions
  - *genre: book, title: The Ecology of Vision; author: J.N.Lythgoe; Date: 1979; Publisher: Clarendon Press, Oxford*
URL Requirements

An object that describes the location of a resource

- Global scope
- parsable
- transportable in many contexts
- extensible
- not loaded with other information
Some URL schemes

- http://host.dom/path
- ftp://host.dom/path
- gopher://host.dom/selector
- news:group.name
- news:article-id
- mailto:email-name@host.dom
- file:///C:/dos/path
- telnet://host.dom
Relative URLs

- “base” + “relative URL”
  => “absolute URL”
- Defines what “base” is for various contexts
- Not defined in terms of scheme
Uniform Resource Names (URN)

- name independent of location; allows for replication, migration
- separate problems of naming authority and name assignment resolution mechanism: finding information about the thing named
  - location(s), metadata
Network Protocols for the Web

Major activities:
- send and receive (email)
- publish and retrieve (web)
- broadcast and subscribe (news, push)

Of course, there’s more:
real time interaction, pay for things, share secrets, query databases, etc.
Standards for Internet protocols

- Sending (SMTP, POP, IMAP, fax)
- Publish, retrieve (HTTP)
- Broadcast communication (NNTP), push
- and more..
  - directory access (LDAP)
  - interactive sessions (TELNET)
HyperText Transfer Protocol (HTTP)

- Started as a simple protocol, designed for the 1990 vision of the World Wide Web
  - Open connection to widget.com
  - send “GET /product.html”
  - read headers
  - read body
  - close connection
HTTP/1.0 added features

- Multiple content-types
  - Accept, language, charset, content-type
- More information
  - User-Agent, From, error codes
- Simple caching
  - last-modified, if-modified-since
- Basic Authorization
HTTP/1.1 Improvements

- Performance
  - pipelining
  - persistent connections
  - caching (Etags)

- Reliability
  - clear semantics for many headers

- New features
Putting the pieces together

- The web is just part of the Internet
- Distributed communication is built out of lots of pieces
- Integration of
  - web, mail, push, security, media,
  - worlds, libraries, identifiers, copyright
Future of Web Standards

- Innovation still leads, standards will follow
  - *This will not end*
- Organizations adapt too
  - *IETF, W3C change*
- Interoperability trumps features
  - *if you’re careful, you can have both*
- Avoiding the tragedy of the commons
  - *local greed over global optimization*
Internet Standards for the Web
End of Part I

Larry Masinter
April 1998