The State of Web Standards

Larry Masinter

Xerox Palo Alto Research Center

May 1996

Purpose of this talk

- Describe the standards process

- Survey current Web-related standards

- Introduce acronyms and buzzwords

- Describe relation to other activities

Organization of talk

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

- Part 2: Recent activities
    - What's the latest news?
    - What are the hard problems?

Vision for the "World Wide Web"…

- ## One network, everyone on it
  - Interoperability across the world

- ## Merged modes of communication
  - Retrieve, mail, broadcast, collaborate

- ## All media
  - Text, sound, video, animation

Three categories of web standards

- Content
  - what are the objects we're moving around?
- Protocols
  - how do they get moved?
- Naming
  - how to reference something not in hand?

Larry Masinter   The State of Web Standards

But first, some words about …
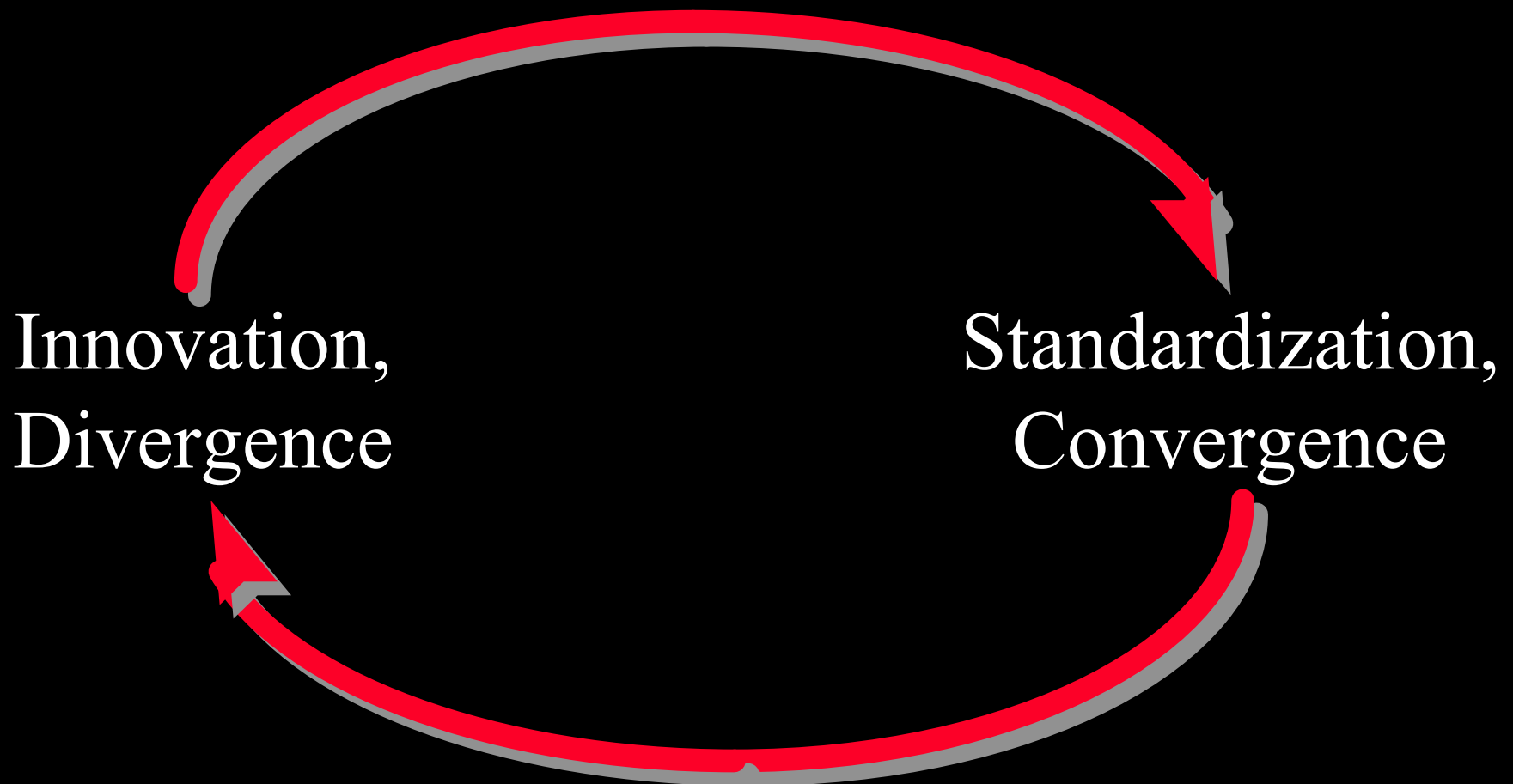
- Standards
- Organizations
- Politics

The nice thing about standards...

7   There are so many of them to choose from.

7   By the time things become standards, they're obsolete.

7   Real standards are set by the market, not committees.

but...

4   Standards promote interoperability.

Standards follow rather than lead innovation in the cycle

Innovation, Divergence

Standardization, Convergence

Who makes standards?

- Standards organizations
- Consortia
- Companies
- Individuals

Some Standards Organizations

- Internet Engineering Task Force (IETF)
- International Organization for Standardization (ISO)
- *And many others: ANSI, AFNOR, IEEE, etc.*

Welcome to ISO Online
ISO International Organization for Standardization

IEEE STANDARDS

ANSI American National Standards Institute
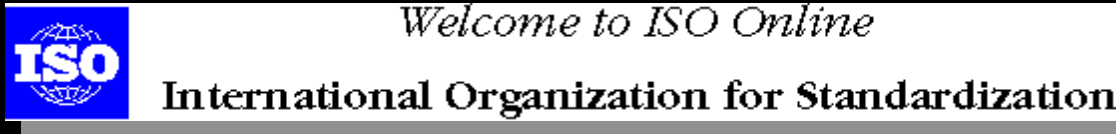
Internet Engineering Task Force

- Defines standards for the Internet
- Different rules, structure than most other standards organizations
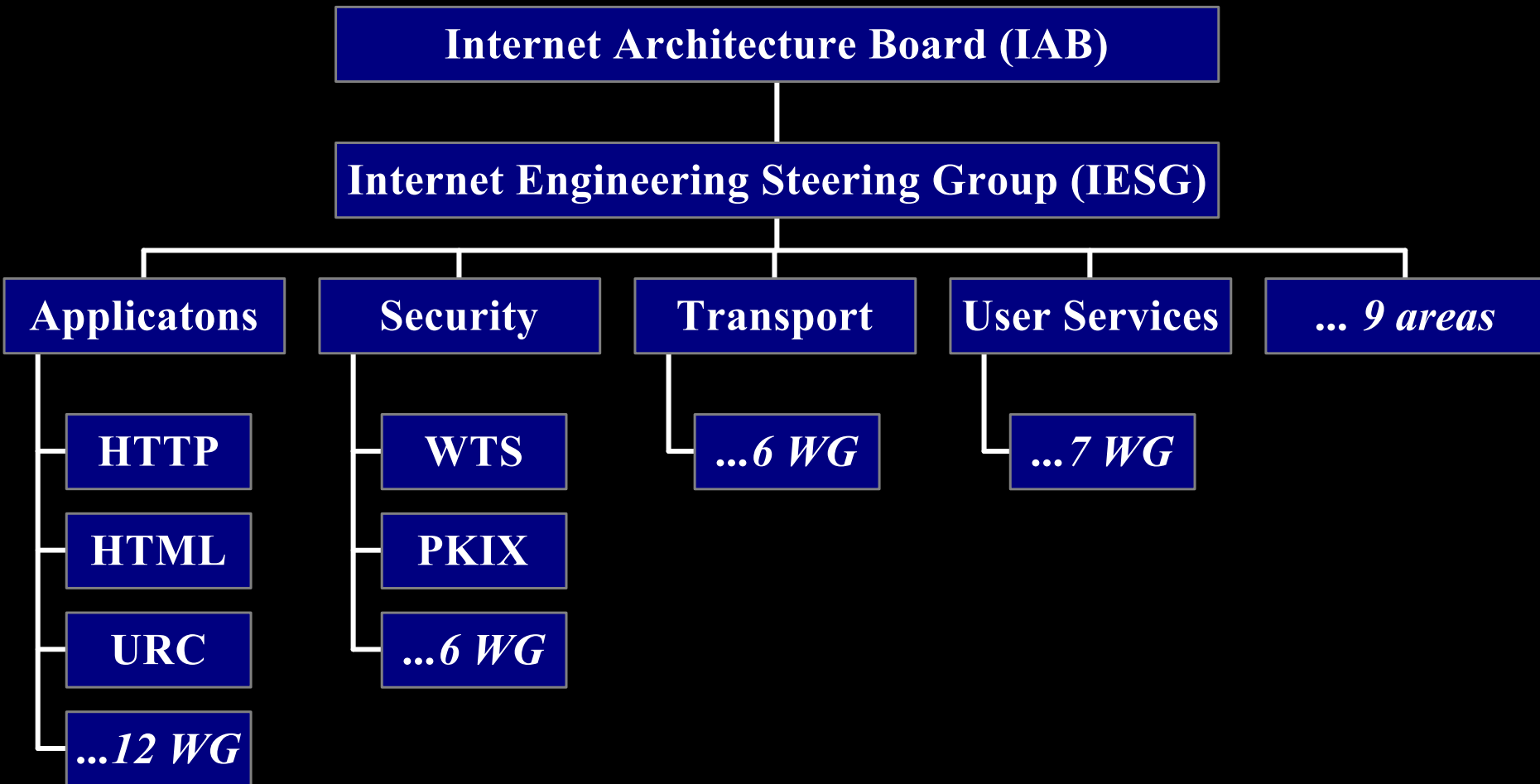- Formal relationship with ISO
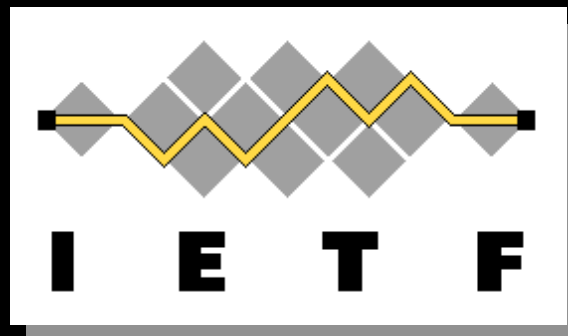
Internet Society

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

IETF structure

Internet Architecture Board (IAB)

Internet Engineering Steering Group (IESG)

Applicatons | **Security** | **Transport** | **User Services** | **... *9 areas***

HTTP — **WTS** — **...*6 WG*** — **...*7 WG***

HTML — **PKIX**

URC — **...*6 WG***

...*12 WG*

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"
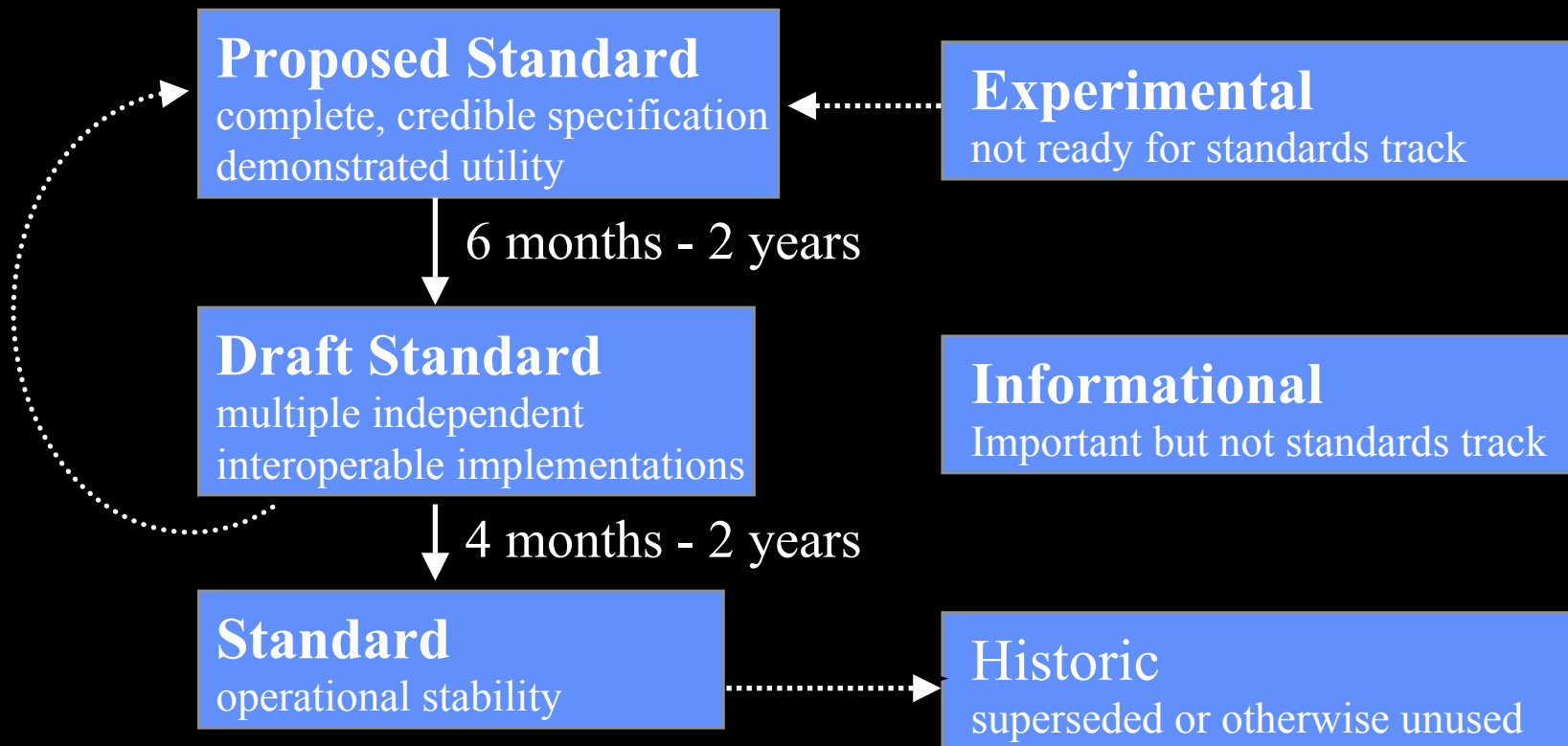
IETF Documents

- 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

IETF Scope

- Internet Standards:
    - Protocols
    - Data formats used in protocols
- Not appropriate:
    - Technology not directly related to protocols
    - Application Program Interfaces (API)

World Wide Web Consortium

- Members are vendors and users

- Paid staff

- Develops web protocols

- Hosts conferences

W3C and IETF relationship

- W3C develops new proposals
- IETF reviews proposals, resolves disagreements
- Not much overlap
- Cooperation when there is overlap
  - W3C staff participate actively in IETF

CommerceNet

- Consortium with focus on use of Internet for electronic commerce
    - Develop mechanisms
        - security, catalogs, EDI, connectivity
    - Education and training
    - Public policy issues

σ Standards & Organizations

- Lots of players

- a common goal:
  Interoperability

- a frequent goal:
  Market Domination

- Avoid the "tragedy of the commons"

〈 Standards for Web Content

- HTML

- MIME and Internet Media Types

- Survey of other web content

Short diversion: What's SGML?

- Standard Generalized Markup Language

- An ISO standard (ISO8879:1986)

- A way of writing
  (ways of writing documents)

- DTD (Document Type Definition)
  defines elements and rules about them

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

Why HTML isn't *just* an application of SGML

It's defined by an SGML DTD...

… plus a description of what the tags mean

… plus some rules about how to display things

…plus some rules about interaction with forms and URLs

… plus some rules about what to do if you see a tag you don't know

HTML 2.0

- RFC 1866: IETF Proposed Standard
- Lots of HTML (as of 1994)...
  - structure, headings, paragraphs, forms, menu, lists, hyperlinks, embedded images
- … but not all.
  - no tables, fonts, colored backgrounds, or Java

HTML 2.0 elements

- Document attributes in header
  - *title, base, links*
- Structure
  - *headings (H1 ... H6), paragraph, address, block*
- Lists, Forms
  - *bullet, numbered, definition, menu*
- Hyperlinks
- Embedded images
  - *simple, image map, image in form*

… more HTML 2.0 elements

- Phrase markup
  - *emphasized, strong*
  - *citation, variable, sample, keyboard*
- Limited typographical elements
  - *bold, italic, monospace*
- Forms
  - *small and large text input, select one-of-many, "radio buttons"*
  - *submit, reset, clear, with URL for action*

σ Summary: HTML 2.0

- HTML 2.0 Proposed Standard
  has many features

- It only has a subset of the HTML that is
  now in common use

- Standardization has been difficult

current activities & future in Part II

Other data on the Internet: MIME

- Multi-Purpose Internet Mail Exchange
- RFC 1521, 1522 and follow-ons
- headers in messages to describe body
- media types for registering formats
- encodings for transfer
- character sets

Internet Media Types ("MIME types")

- Standard way of naming data formats

- Hierarchical structure with parameters

- web, email, netnews applications
  use MIME to decide how to interpret data

- use instead of file extension (`logo.gif`)

- `text, image, audio, video, multipart, application`

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

Other content on the web

- Full SGML
  - catalogs, encapsulation
- Page layout
  - Postscript, Portable Document Format (PDF)
- Video
  - MPEG, QuickTime, AVI
- Audio
  - Basic, RealAudio

Other content on the web

- Desktop applications
  - Word, Excel, etc.
- 3D graphics
  - VRML and follow-ons
- Interactive applications
  - Java and others

σ Content on the web

- Lots of innovation
- Much of it outside of standardization
- For now, that's OK
- Ultimately, it isn't

σ Content needs standards

- Benefits from open standards:
    - Interoperability, more platforms & tools
    - Preservation
    - Cost
- Vendors prefer lock-in
    - sell more tools, software libraries, training, etc.
- Demand open formats

〈 Network Protocols for the Web

- There are mainly three things people do on the net
  - send (email)
  - get (web)
  - broadcast (news)

Of course, there's more:

  real time interaction, pay for things, share secrets, query databases, etc.

HyperText Transfer Protocol (HTTP)

- Started as a simple protocol, designed for the 1990 vision of the World Wide Web

- **`http://widget.com/product.html`**

  – 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/1.0 specification

- IETF Informational RFC

  (Approved March 28, but RFC not assigned as of April 27)

- HTTP as it was practiced in 1995

- Many features "listed but not described"

  Current implementations differed in interpretation too much

σ HTTP standard

- HTTP/1.1: Proposed Standard soon
  – Clarify ambiguities in HTTP/1.0
  – Improve performance and load on Internet
- HTTP/1.2
  – things that didn't make 1.1
- HTTP-NG
  – redesign rather than incremental
  – distributed object systems ?

… more in part II

Other related protocol work

- Secure HTTP (S-HTTP)
  - proposed standard soon
- Secure Sockets Layer (SSL)
  - and variations
- Internet Payment
  - no standards yet
- Voluntary Access Control
  - charter, but no proposal forwarded yet

⟨ Identifiers in the Web

- ## URL: locations

  – *New York Public Library, second floor, third aisle, second shelf, third book from left*

- ## URN: location-independent names

  – `QP:475.L95;   ISBN:0-19-854529-0`

- ## URC: descriptions

  – *genre:* book*, title:* The Ecology of Vision*; author:* J.N.Lythgoe*; Date:* 1979*; Publisher:* Clarendon Press, Oxford

Uniform Resource Locators

- RFC 1630: *Uniform Resource Identifiers in the World Wide Web*

- RFC 1736: *Functional Recommendations for Internet Resource Locators*

- RFC 1738: *Uniform Resource Locators*
  - Proposed Standard

URL Requirements

An object that describes the location of a resource

- Global scope

- parsable

- transportable in many contexts

- extensible

- not loaded with other information

URL Proposed Standard

- limited repertoire of characters

    - not all of ASCII

    - encoding for bytes that can't be directly represented as one of those characters

        `phrase%20with%20spaces`

- *scheme:scheme-specific-part*

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`**

URLs in plain text

- Recommendations
  - `<URL:http://host.dom/path/part>`
  - no hyphens when line breaks
- Does a name need a name?
  - is "tel:" part of your telephone number?

Relative URLs

- RFC 1808: *Relative Uniform Resource Locators*

  ```
  ../image.gif
  ./dir1/dir2/sample
  ```

- "base" + "relative URL"
  => "absolute URL"

- Defines what "base" is for various contexts

- Not defined in terms of scheme

Uniform Resource Names

- RFC 1737: Functional Requirements for Uniform Resource Names

- location-independent designators

- Requirements
  - global scope, persistent, scaleable

…more in Part II

URC: Uniform Resource Characteristics

- Syntax for carrying metadata
  - Title
- A standard set of tags useful for describing Internet resources
- Standards work:
  - URC working group forming

σ References on the Web

- URLs are used widely
  - some minor issues with new URL schemes
- URN and URC work has been slow
  - innovation before standardization

σ Summary, Part I

- Many organizations and people are involved in producing standards

- Standards are progressing for
  - data: HTML
  - protocols: HTTP
  - references: URL

Part II will cover more about current activities and difficult problems

The State of Web Standards Part II

Larry Masinter

Xerox Palo Alto Research Center

May 1996

Overview of Part II

Recent events and current activities

- Content
  - beyond HTML 2.0
- Protocols
  - HTTP and follow-ons
- References
  - URLs, URNs, URCs

HTML Working Group activity

- Tables
- File Upload
- Internationalization
- Embedded objects
- Extensions

… but

HTML-WG to finish current work and close, W3C will continue activities

HTML Tables

- February 1, 1996 draft
  `draft-ietf-html-tables-06.txt`

- Recent changes include:
    - more formatting control
    - incremental display
    - compatibility with popular browsers
    - compatibility with CALS

File Upload

- RFC 1867 (Experimental)

- Add a way that a form can ask a user for a file as well as data to be typed in
  `<INPUT TYPE=FILE>`

- A better encoding for data returned from filling out forms (`multipart/form-data`)

HTML Internationalization

- Extended character sets
  - SGML numeric character references `Ӓ` always refer to ISO 10646
  - MIME charset specifies encoding
- **LANG** attribute for noting language of sections in multi-lingual text
- Form submission
- Minor extra enhancements

Internationalization problems:

- Non-ASCII characters in URLs

- Non-ASCII simple query forms

- Interaction with `<FONT>` and style sheets

HTML Style and Style Sheets

- Presentation descriptions
  - In a separate resource
  - In the HTML head
  - Inline on each element
- How are styles described?
  - Cascading Style Sheets (CSS)
  - Other proposals?

The debate over inline style (**`<FONT>`** or equivalent)

4   People want it

8   They'll misuse it

4   Inline style displays faster incrementally

8   Precomputed styles

4   It's easier to enter inline markup

8   Automated tools make styles just as easy

4   *"Give them rope"*

8   *"They'll hang themselves"*

Compound Documents in HTML

- Many tags with similar purpose
  - **`EMBED, FIG, IMG, OBJECT, APPLET`**

- Can these be merged?
  - several proposals made
  - convergence is elusive

HTML Link model

- Beyond **`<A HREF="…">`**

- Showing relationships internally
  `REL=MADE, REL=PREVIOUS`

- Redefining button-bar elements
  `<LINK REL=xxx HREF="…">`

HTML Feature identification

- Some mechanism of registering HTML extensions

- Some mechanism of delivering HTML with conditional features
  "if you do 12-dimensional tables, use this; if not, use this instead"

- Possibly some mechanism of client/server negotiation for conditional features

HTML and IETF

- IETF usually does protocols, not data formats

- HTML/2.0 was important enough to be taken up by IETF

- HTML-WG was behind schedule and not making good progress

- Industry was going different directions

σ HTML Standards status

- Standardization has been hard
- Probably won't be a HTML 3.0 standard
- IETF HTML-WG to close
  - finish current activities; extensions registration
- W3C and others to develop features
- Standardization to lag innovation

Other media standards

- MIME revision in progress
  New hierarchical name space for vendor-defined data types
  `application/vnd.ms-excel`

- New (patent-free) compression mechanisms

- Much activity in multimedia, outside standard organizations

σ Content: Registration vs. Standardization

- Meta-standard: a standard way of saying which non-standard thing you did

- A way to solve impasse when standardization is not possible

- Register your types!

- Not a substitute for convergence

〈 The problems with HTTP

- HTTP traffic clogs Internet
- TCP/IP designed for "congestion control"
- Some trans-ocean links are always congested
- Internet routing caches not useful:
  too many short connections

Things are more complex now

- Multiple objects per click
- Many more users: HTTP dominates traffic
- Multiple connections: self-congestion
- Spiders and search engines
- Proxies, caches, shopping baskets

Prospective growth

- To meet projected demand, web capacity needs to increase 10,000-fold.

- Improvements in infrastructure will result in at most 100 times more capacity.

- Protocols and use of network need to be 100 times more efficient.

Toward better web performance

- Persistent connections

- Multiplexed connections

- Protocol improvements to allow caching reliably

- Deployment of caches by national networks, Internet Service Providers

HTTP/1.1 Highlights

- **`HOST`** header

- caching

- content negotiation

- byte ranges

- state and sessions

- persistent connections

Other HTTP work

- extensions, demographics
- feature negotiation
  - Media type parameters
  - Display size, color
- beyond access
  - version management
  - search

HTTP-NG

- "Next Generation" design
- Not required to be compatible
- Design goals:
  - simple
  - performance
  - asynchronous operation
  - mandatory display

HTTP and distributed objects

- Specify protocol with formal specification language
- Tune transport for situation
- Allow multiple transports
- ILU: Inter-Language Unification
  - Distributed object technology
  - Freely available from Xerox
  - CORBA compatible

Web Security

- WTS working group
  - S-HTTP to be Proposed Standard
- Connection-based security
  - SSL
- Digest Authentication
- Payment on the Internet
  - IPAY-WG?

Access control and ratings

- Rating of entertainment content for adult themes

- How to deal with cultural differences

- Multiple rating services

- Voluntary Access Control working group didn't start

σ Web network protocols

- Save the Internet from the Web!
  - Local decisions can have global impact
- Many features still needed
- The "tragedy of the commons" is still a threat

〈 References in the Internet

- New URL schemes

- URNs in development

- URC syntax developments

- Unsolved problems

New URL schemes

- **`nttp://host/article-id`**
- z39.50 URL schemes
- **`ldap:`** for Light-Weight Directory Access Protocol
- **`data:image/gif,,bbacd01xyz`**
- non-standard URLs
  - **`about:mozilla, aol:word, palace://host.dom`**

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
  - content

URN naming mechanisms

- A common syntax

  `urn:hdl:cnri.dlib/august95`

  `urn:lifn:some.domain:anything-goes-here`

  `urn:path:/A/B/C/doc.html`

  `urn:inet:library.bigstate.edu:aj17-mcc`

- Several different experimental resolution mechanisms

Still experimental

Uniform Resource Characteristics (URCs)

- describe attributes (title, author, data)

- useful for making a citation

- URC working group developing charter
  - structure of resource descriptions
  - at least two external syntax representations

Many previous standards to choose from

Some unsolved problems

- stuff goes away
  - Material behind URLs disappears
- **pimples.com**
  - vanity domains for billboard use
- Apple Computer and Apple Music
  - conflicts over short names
- **urn:hdl:MTV/I_quit**
  - how does authority migrate?
- **http://www.métro.paris.fr/métro**
  - Non-ASCII names

σ Current Web Standards

- Lots of activity

- Lots of innovation

- Lots of bad ideas as well as good ones

- Shake-out will take a long time

σ The Future

- Innovation leads, standards follow
  - This will not end
- Organizations adapt too
  - IETF, ISO are changing, albeit slowly
- Convergence is not inevitable
  - Things could worse instead of better
  - *You* can help

The State of Web Standards

Larry Masinter

Xerox Palo Alto Research Center

May 1996

THE DOCUMENT COMPANY

XEROX