

World Wide Web

Péter Jeszenszky

Faculty of Informatics, University of Debrecen

jeszenszky.peter@inf.unideb.hu

Last modified: September 6, 2025

The Birth of the Web (1)

- The World Wide Web was born in CERN.
- Both the idea and the implementation came from Tim Berners-Lee (TBL).
 - For more information about TBL, see:
<https://www.w3.org/People/Berners-Lee/>
- The idea:
 - Tim Berners-Lee. *Information Management: A Proposal*. March 1989.
<https://www.w3.org/History/1989/proposal.html>
 - He recommended a hypertext information system to CERN.
 - Tim Berners-Lee, Robert Cailliau. *WorldWideWeb: Proposal for a HyperText Project*. 12 November 1990.
<https://www.w3.org/Proposal.html>

The Birth of the Web (2)

- TBL is the creator of the following:
 - The first web server (CERN httpd) (December 24, 1990)
<https://www.w3.org/Daemon/>
 - The first web browser and HTML editor (WorldWideWeb) (December 25, 1990) <https://www.w3.org/People/Berners-Lee/WorldWideWeb.html>
 - HTML (HyperText Markup Language)
<https://info.cern.ch/hypertext/WWW/MarkUp/MarkUp.html>
 - HTTP (HyperText Transfer Protocol)
<https://www.w3.org/Protocols/HTTP/AsImplemented.html>
 - URI (Universal Resource Identifier), originally called as UDI (Universal Document Identifier)
<https://info.cern.ch/hypertext/WWW/Addressing/Addressing.html>

The Birth of the Web (3)

- The first public website: <http://info.cern.ch/>
(launched on: August 6, 1991)
 - See: *Restoring the first website*
<https://first-website.web.cern.ch/first-website/>

History

- See:
 - <https://thehistoryoftheweb.com/>

Idea

- Originally, the idea of the Web is based on the following cornerstones:
 - Identifying resources by global identifiers (URIs)
 - Client-server model
 - Hypertext markup language (HTML)

Web Architecture

- Architecture of the Web from a contemporary viewpoint:
 - *Architecture of the World Wide Web, Volume One* (W3C Recommendation, 15 December 2004)
<https://www.w3.org/TR/webarch/>
 - The client-server model is not mentioned at all in the text!

Web Architecture: Concepts (1)

- **World Wide Web:** an information space in which the items of interest (referred to as resources) are identified by URIs.
- **Resource:** anything that might be identified by a URI.
 - **Information resource:** a resource which has the property that all of its essential characteristics can be conveyed in a message.
- **Uniform Resource Identifier (URI):** a global identifier in the context of the Web.
- **Representation:** data that encodes information about resource state.

Web Architecture: Concepts (2)

- **Content negotiation:** offering multiple representations for a resource and selecting the one that is the most appropriate when a representation must be served.
- **Dereferencing a URI:** using a URI to access the referenced resource.
 - Access may take many forms, including retrieving, adding, or modifying a representation of the resource, and deleting some or all representations of the resource.

Web Architecture: Concepts (3)

- **Web agent:** a person or a piece of software acting on the Web on behalf of a person, entity, or process.
 - For example, a web crawler.
- **User agent:** one type of Web agent, a piece of software acting on behalf of a person.
 - For example, a web browser.

Architectural Bases of the Web

- **Identification:**

- Resources are identified by global identifiers called URIs.

- **Interaction:**

- Web agents communicate using standardized protocols that enable interaction through the exchange of messages.
 - Web protocols include, for example, HTTP, HTTPS, and WebDAV.
- A message may include data as well as metadata about a resource, the message data, and the message itself.

- **Data Formats:**

- The choice of interaction protocol places limits on the formats of representation data and metadata that can be transmitted.
- The Web itself does not constrain the data formats that can be used by content providers.
 - For a data format to be usefully interoperable between two parties, the parties must agree (to a reasonable extent) about its syntax and semantics.

Web Architecture: Example Scenario

URI

```
http://weather.example.com/debrecen
```

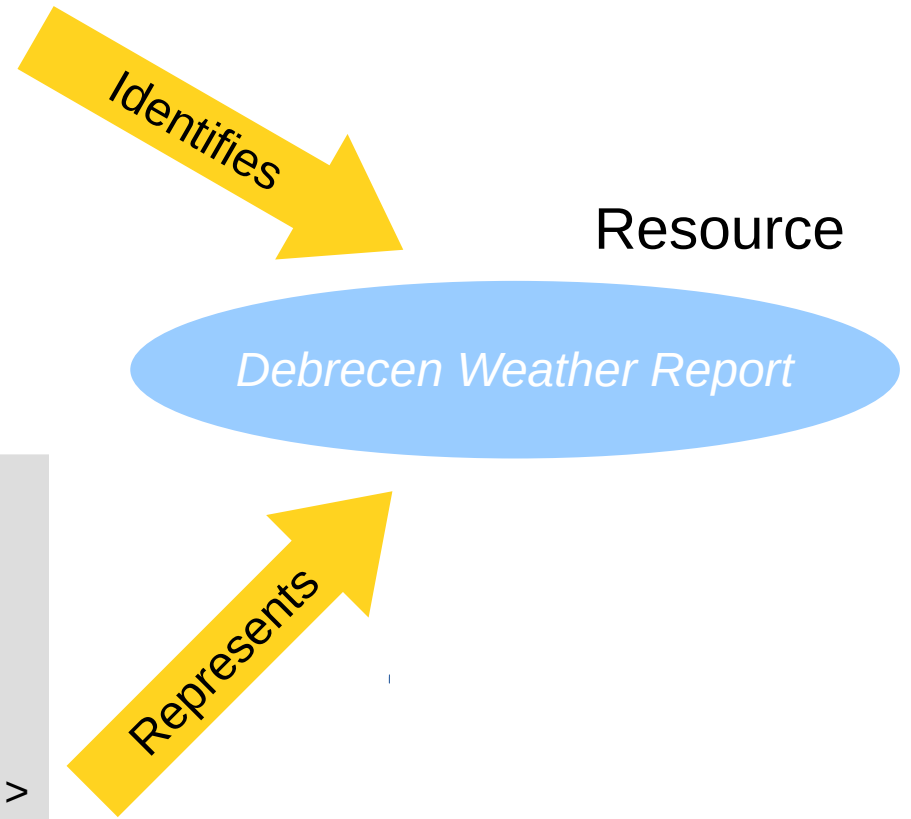
Representation

Metadata:

```
Content-Type:  
application/xhtml+xml; charset=utf-8
```

Data:

```
<!DOCTYPE html>  
<html xmlns="http://www.w3.org/1999/xhtml">  
  <head>  
    <title>10 Day Weather Forecast for  
      Debrecen</title>  
  ...
```



Standards (1)

- A standard is a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose.
 - See:
<https://web.archive.org/web/20200101101550/https://www.iso.org/standards.html>

Standards (2)

- By origin, there are three types of standards:
 - **De facto standards:** arise from common usage or market acceptance.
 - Examples: the QWERTY keyboard layout, TeX, PDF (before 2008).
 - **De jure standards:** are mandated by regulators at the local, state, federal, and/or international level.
 - Examples: International System of Units (SI), PDF (from 2008).
 - **Voluntary consensus standards:** are specified within a range of private institutions, including engineering societies, trade associations, accredited standards-setting organizations, and industry consortia.
 - Examples: the Internet protocol suite (commonly known as TCP/IP), HTML, CSS.
- See:
 - Andrew L. Russell. *Open Standards and the Digital Age*. Cambridge University Press, 2014. <https://arussell.org/open/>

Open Standard (1)

- There is no single, universally accepted definition:
 - *OpenStand: The Modern Paradigm for Standards* (IEEE, ISOC, IETF, IAB, W3C, ...)
<https://open-stand.org/>
 - *Open Standards Requirement for Software* (Open Source Initiative) <https://opensource.org/osr/>
 - ...
- Further information: *Open standard*
https://en.wikipedia.org/wiki/Open_standard

Open Standard (2)

- In general, an open standard is a standard that is freely available for use and adoption to anyone.
- Open standards are typically developed via a collaborative process.

Web Standards

- The following organizations are responsible for web standards:
 - *Ecma International* <https://www.ecma-international.org/>
 - *International Organization for Standardization (ISO)* <https://www.iso.org/>
 - *Internet Engineering Task Force (IETF)* <https://www.ietf.org/>
 - *Unicode Consortium* <https://unicode.org/consortium/consort.html>
 - *Web Hypertext Application Technology Working Group (WHATWG)* <https://whatwg.org/>
 - *World Wide Web Consortium (W3C)* <https://www.w3.org/>
 - ...

Internet Assigned Numbers Authority (IANA)

- Coordinates the allocation of codes and numbers that form the basis for the operation of the Internet.
<https://www.iana.org/>
 - Manages the DNS root zone, and the `.int` and `.arpa` domains.
 - Coordinates the allocation of IP addresses globally.
 - Maintains registries of codes and numbers used in a variety of Internet protocols.
 - See: *Protocol Registries* <https://www.iana.org/protocols>
- IANA is a function that is currently performed by the Internet Corporation for Assigned Names and Numbers (ICANN), a not-for-profit corporation.

Internet Engineering Task Force (IETF)

- An international standards organization developing Internet standards.
 - For example, IETF develops the Internet protocol suite (commonly known as TCP/IP).
 - The IETF has no formal membership, no membership fee, participation is open to anyone.
 - Mailing lists: <https://www.ietf.org/list/>
 - The technical work is done in working groups.
- Formation: 1986
 - See: *IETF Turns 25 on 16 January 2011*
<https://www.ietf.org/mail-archive/web/ietf-announce/current/msg08366.html>
- Publishes Internet standards-related specifications in the RFC series of documents.

Request for Comments (RFC) (1)

- The RFC series contains technical and organizational documents about the Internet.
- The RFC series of documents began in 1969 as part of ARPANET project.
 - The first RFC:
 - Steve Crocker. *Host Software*. RFC 1, 7 April 1969.
<https://www.rfc-editor.org/info/rfc1>

Request for Comments (RFC) (2)

- RFC Editor edits, publishes, and catalogs RFCs.
<https://www.rfc-editor.org/>
- By origin, the RFC series is split into four streams:
 - The Internet Engineering Task Force (IETF) Stream
 - The Internet Architecture Board (IAB) Stream
 - The Internet Research Task Force (IRTF) Stream
 - The Independent Submission Stream
- Further information about the RFC series:
 - Russ Housley (ed.), Leslie L. Daigle (ed.). *The RFC Series and RFC Editor*. RFC 8729, February 2020.
<https://www.rfc-editor.org/rfc/rfc8729>

Request for Comments (RFC) (3)

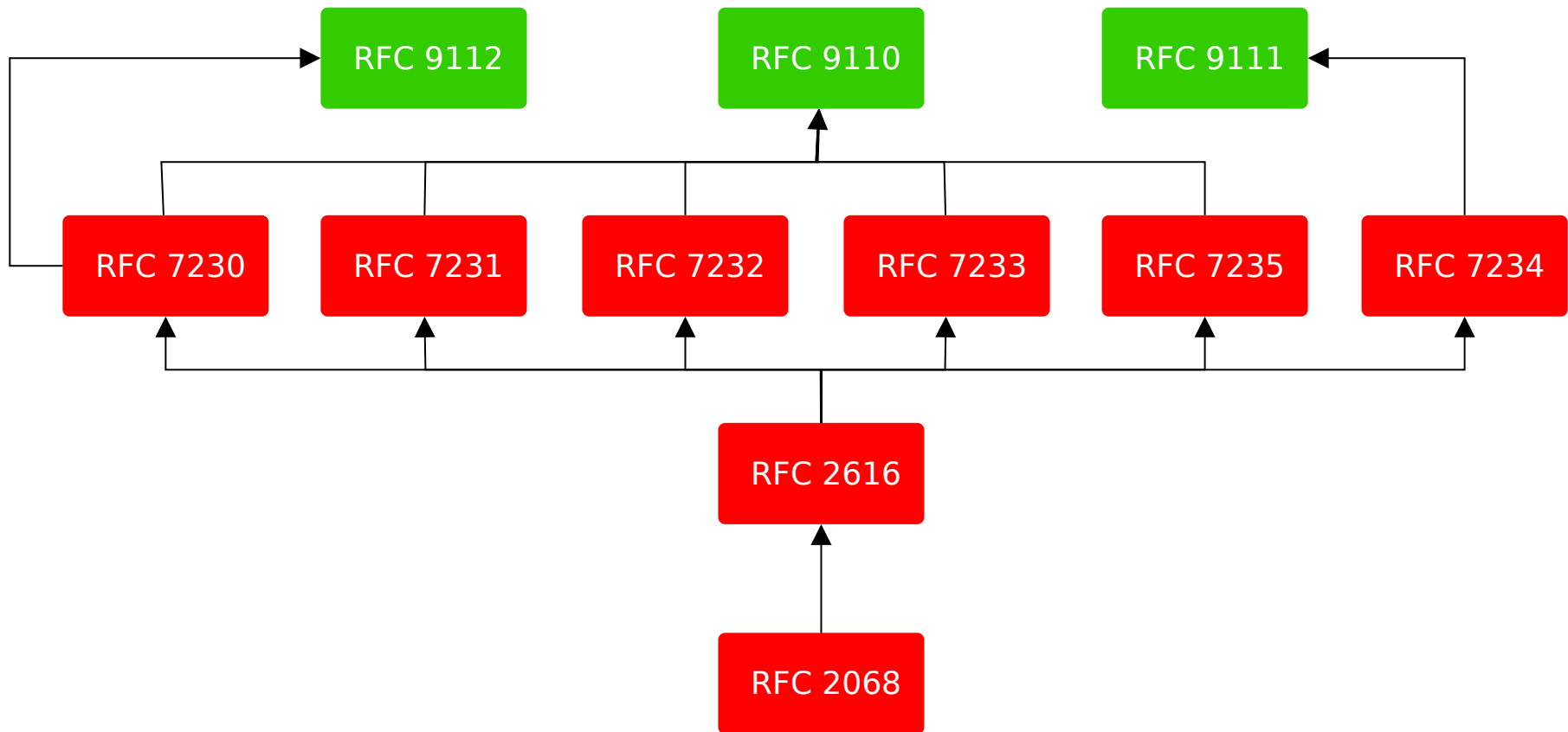
- Each RFC is identified by a number, such as RFC 9110.
- Each RFC is available in ASCII text, such as:
<https://www.rfc-editor.org/rfc/rfc9110.txt>
 - The same RFC in HTML:
<https://www.rfc-editor.org/rfc/rfc9110.html>
- The list of all RFCs:
<https://www.rfc-editor.org/rfc-index.html>

Request for Comments (RFC) (4)

- Published RFCs never change.
- Various errors are fixed by errata.
- Amendments can be also made by writing and publishing a revised RFC.
 - An RFC can obsolete or update earlier RFCs.

Request for Comments (RFC) (5)

- Example: Hypertext Transfer Protocol – HTTP/1.1



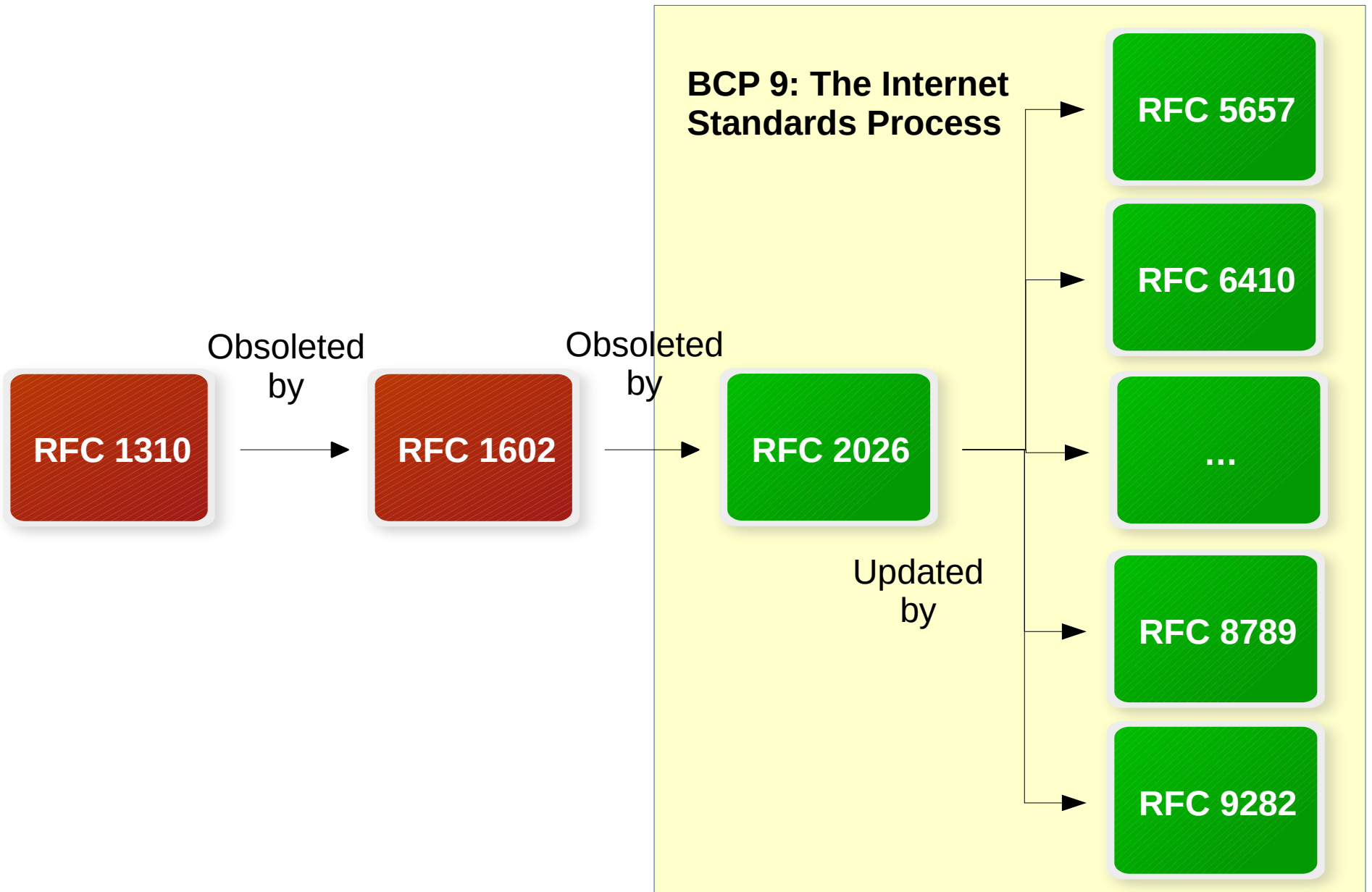
Request for Comments (RFC) (6)

- The series of IETF RFCs contains the following two important sub-series:
 - ***Best Current Practice (BCP)***:
 - BCPs document guidelines, processes, or the operation of the IETF itself.
 - *BCP Index*: <https://www.rfc-editor.org/rfc/bcp/>
 - ***Internet Standard (STD)***:
 - *STD Index*: <https://www.rfc-editor.org/rfc/std/>

Request for Comments (RFC) (7)

- BCPs and STDs are assigned a number in their subseries while retaining their RFC number.
 - Example:
 - Scott O. Bradner. *The Internet Standards Process – Revision 3*. **BCP 9**, RFC 2026, October 1996. <https://www.rfc-editor.org/rfc/rfc2026>
 - Tim Berners-Lee, Roy T. Fielding, Larry Masinter. *Uniform Resource Identifier (URI): Generic Syntax*. **STD 66**, RFC 3986, January 2005. <https://www.rfc-editor.org/rfc/rfc3986>
- Several RFCs may share the same BCP or STD number.
 - For example, an STD number identifies a standard not a document.

Request for Comments (RFC) (8)



Request for Comments (RFC) (9)

- **Standards Track**: the set of maturity levels of RFCs that are intended to become Internet Standards.
 - Originally, three maturity levels were used:
 - Proposed Standard
 - Draft Standard
 - Internet Standard
 - Currently, the Proposed Standard and Internet Standard maturity levels are used.
- See:
 - Scott O. Bradner. *The Internet Standards Process – Revision 3*. BCP 9, RFC 2026, October 1996. <https://www.rfc-editor.org/rfc/rfc2026>
 - Russell Housley, Dave Crocker, Eric W. Burger. *Reducing the Standards Track to Two Maturity Levels*. BCP 9, RFC 6410, October 2011. <https://www.rfc-editor.org/rfc/rfc6410>

Request for Comments (RFC) (10)

- **Internet-Draft:** a draft version of a specification made available for informal review and comment during the development.
 - May or may not eventually be published as an RFC.
 - Is subject to change or removal at any time.
 - Is valid for a maximum of six months.
 - Should not be cited or quoted in any formal document, except as “work in progress”.
 - Example:
 - Austin Wright (ed.), Henry Andrews (ed.), Ben Hutton (ed.), Greg Dennis. *JSON Schema: A Media Type for Describing JSON Documents*. 10 June 2022. <https://datatracker.ietf.org/doc/id/draft-bhutton-json-schema-01.html>

Request for Comments (RFC) (11)

- On nearly every April 1 since 1989, one or more funny RFCs has been published.
 - Example:
 - Jogi Hofmueller (ed.), Aaron Bachmann (ed.), Johannes Zmoelnig (ed.). *The Transmission of IP Datagrams over the Semaphore Flag Signaling System (SFSS)*. RFC 4824, April 1 2007. <https://www.rfc-editor.org/info/rfc4824>
- See:
 - *April Fools' Day Request for Comments*
https://en.wikipedia.org/wiki/April_Fools%27_Day_Request_for_Comments

World Wide Web Consortium (W3C)

- The W3C is an international community where member organizations, a full-time staff, and the public work together to develop open web standards.
 - See: <https://www.w3.org/about/>
- W3C publishes documents called Recommendations that define Web technologies and are considered Web standards.
 - See: <https://www.w3.org/standards/>

W3C Design Principles

- **Web for All:** the Web must be available to all people, whatever their hardware, software, network infrastructure, native language, culture, geographical location, or physical or mental ability
 - Related concepts: web accessibility, internationalization
- **Web on Everything:** the Web must be accessible from a wide variety of devices.
 - E.g., mobile phones, smart phones, interactive television systems, domestic appliances, ...
- See:
 - *Our mission – Our design principles*
<https://www.w3.org/mission/#principles>
 - *Vision for W3C* <https://www.w3.org/TR/w3c-vision/>

History of the W3C

- Was founded at MIT in October 1994.
- The director is Tim Berners-Lee, the inventor and creator of the Web.
- Has published more than 300 recommendations since 1996.
 - See:
<https://www.w3.org/TR/?status%5B0%5D=standard>

W3C: A Few Milestones (1)

- **October 1996:** *PNG (Portable Network Graphics) Specification Version 1.0*
<https://www.w3.org/TR/REC-png-961001>
- **December 1996:** *Cascading Style Sheets, level 1*
<https://www.w3.org/TR/REC-CSS1-961217>
- **February 1998:** *Extensible Markup Language (XML) 1.0*
<https://www.w3.org/TR/1998/REC-xml-19980210>
- **April 1998:** *Mathematical Markup Language (MathML) 1.0 Specification*
<https://www.w3.org/TR/1998/REC-MathML-19980407/>

W3C: A Few Milestones (2)

- **October 1998:** *Document Object Model (DOM) Level 1 Specification* <https://www.w3.org/TR/REC-DOM-Level-1/>
- **November 1999:** *XSL Transformations (XSLT) Version 1.0* <https://www.w3.org/TR/1999/REC-xslt-19991116>
- **December 1999:** *HTML 4.01 Specification* <https://www.w3.org/TR/html401/>
- **January 2000:** *XHTML 1.0: The Extensible HyperText Markup Language* <https://www.w3.org/TR/2000/REC-xhtml1-20000126/>
- **May 2001:** *XHTML 1.1 – Module-based XHTML* <https://www.w3.org/TR/2001/REC-xhtml11-20010531/>

W3C: A Few Milestones (3)

- **October 2004:** *XML Schema* <https://www.w3.org/TR/xmlschema-0/>
<https://www.w3.org/TR/xmlschema-1/>
<https://www.w3.org/TR/xmlschema-2/>
- **June 2011:** *Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification* <https://www.w3.org/TR/CSS2/>
- **September 2011:** *Selectors Level 3*
<https://www.w3.org/TR/2011/REC-css3-selectors-20110929/>
- **June 2012:** *Media Queries*
<https://www.w3.org/TR/2012/REC-css3-mediaqueries-20120619/>
- **October 2014:** *HTML5 – A vocabulary and associated APIs for HTML and XHTML*
<https://www.w3.org/TR/2014/REC-html5-20141028/>

W3C: A Few Milestones (4)

- **December 2019:** *WebAssembly Core Specification*
<https://www.w3.org/TR/wasm-core-1/>
- **April 2020:** *Web of Things (WoT) Architecture*
<https://www.w3.org/TR/wot-architecture10/>
- **January 2021:** *WebRTC 1.0: Real-Time Communication Between Browsers*
<https://www.w3.org/TR/2021/REC-webrtc-20210126/>
- **April 2022:** *Media Queries Level 3*
<https://www.w3.org/TR/mediaqueries-3/>
- **May 2023:** *EPUB 3.3* <https://www.w3.org/TR/epub-33/>
- **August 2024:** *Geolocation* <https://www.w3.org/TR/geolocation/>

W3C Operation (1)

- Currently, W3C has 347 Members from around the world (September 5, 2025).
 - The list of members: <https://www.w3.org/membership/list/>
 - Adobe, Amazon, Apple, CERN, Google, IBM, Intel, Meta, Microsoft, SZTAKI, ...
- Geographic or interest-based communities of individuals interested in W3C's activities: W3C Chapters <https://chapters.w3.org/>
 - Hungary Chapter <https://chapters.w3.org/hungary/>

W3C Operation (2)

- Development is carried out by working groups.
- Deliverables produced by working groups include technical reports, test suites, and open source software.
- Working groups are composed of experts in the area in question, each of whom can be any of the following:
 - a member of the W3C Team (e.g., a W3C employee),
 - an individual representing a member organization (typically, an employee of a member organization)
 - an individual participating as an invited expert.

W3C Operation (3)

- Currently, W3C has 45 working groups (September 5, 2025).
<https://www.w3.org/groups/wg/>
 - Cascading Style Sheets (CSS) Working Group
<https://www.w3.org/Style/CSS/members>
 - HTML Working Group <https://www.w3.org/groups/wg/htmlwg/>
 - Internationalization Working Group
<https://www.w3.org/International/i18n-activity/i18n-wg/>
 - Web Applications Working Group
<https://www.w3.org/groups/wg/webapps/>
 - Web Machine Learning Working Group
<https://www.w3.org/groups/wg/webmachinelearning/>
 - ...

W3C Participation

- Participation is open to the public, you can:
 - Review specifications and provide feedback
<https://www.w3.org/standards/review/>
 - Join mailing lists <https://www.w3.org/email/>
 - Join community and business groups <https://www.w3.org/community/>
 - Contribute to W3C open source software <https://www.w3.org/Status>
 - Translate specifications and other resources
<https://www.w3.org/Consortium/Translation/>
 - Attend events (e.g., conferences, workshops) organized by W3C
<https://www.w3.org/events/>
 - ...
- See: <https://www.w3.org/get-involved/>

W3C Technical Reports

- See the following about the various technical reports published by W3C:
 - *W3C Process Document* (18 August 2025) <https://www.w3.org/Consortium/Process/>
 - *Types of documents W3C publishes* <https://www.w3.org/standards/types>
- All technical reports: <https://www.w3.org/TR/>
- W3C documents are licensed under the W3C Document License.
 - See: *Document License* <https://www.w3.org/copyright/document-license/>
 - Further information: <https://www.w3.org/copyright/>

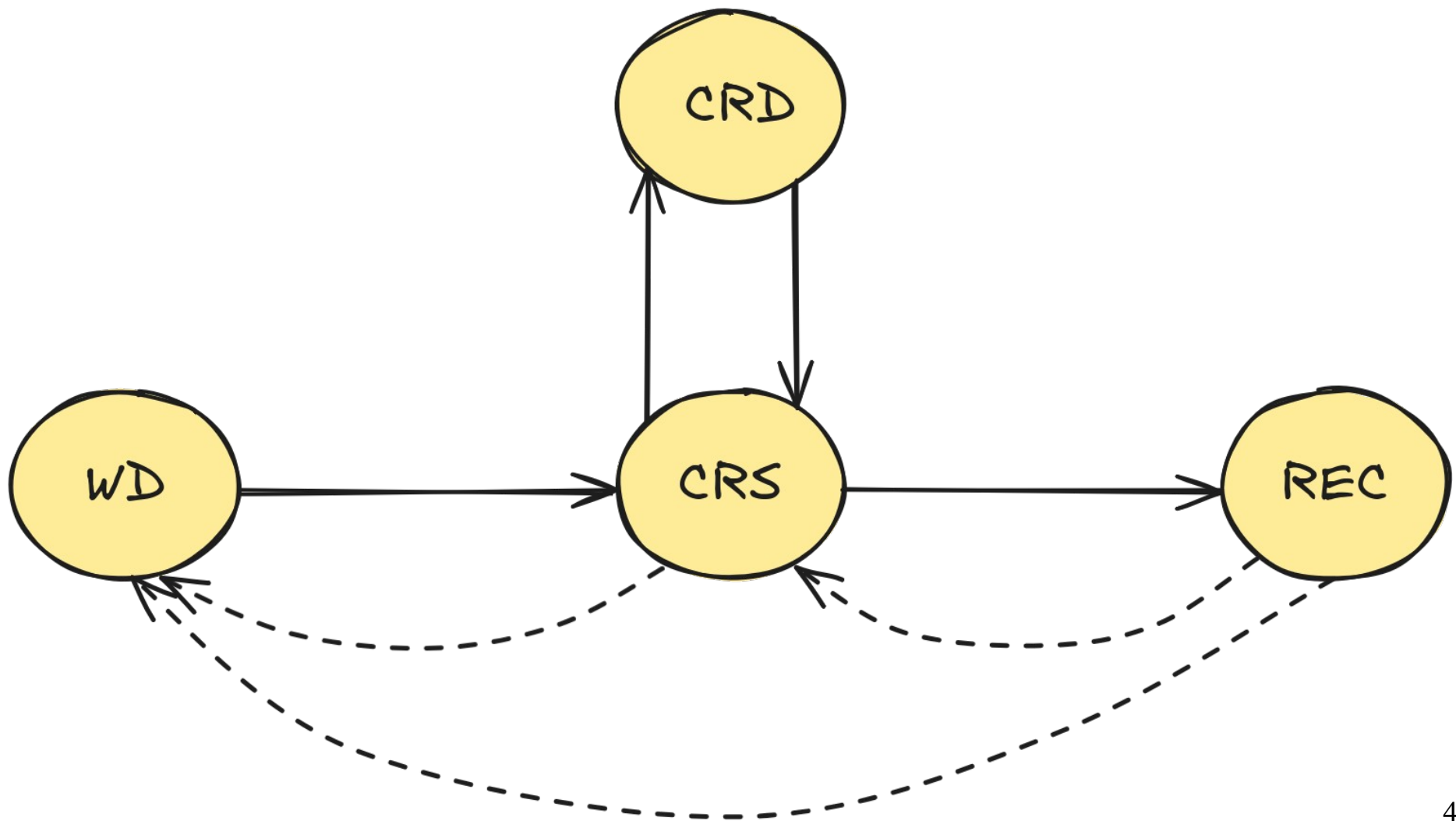
Maturity Levels of W3C Technical Reports (1)

- **Working draft (WD)**: a document that is published for review by the community (including W3C members), the public, and other technical organizations.
 - Some, but not all, Working Drafts are meant to advance to Recommendation.
- **Candidate recommendation (CR)**: a document that has already received wide review and is made available to gather implementation experience. It has two forms:
 - **Candidate recommendation snapshot (CRS)**:
 - **Candidate recommendation draft (CRD)**:
- **Recommendation (REC)**: a Web standard suitable for wide adoption.
- **Group Note (NOTE)**: a document that is not intended to be a formal standard.
 - Are published to document information other than technical specifications, such as use cases motivating a specification and best practices for its use.

Maturity Levels of W3C Technical Reports (2)

- A recommendation may become superseded, obsolete, or rescinded:
 - **Superseded Recommendation:** a specification that has been replaced by a newer version.
 - Example:
 - *XHTML 1.0 The Extensible HyperText Markup Language (Second Edition)*
<https://www.w3.org/TR/xhtml1/>
 - **Obsolete Recommendation:** a specification that the W3C has determined lacks sufficient market relevance to continue recommending it for implementation.
 - Example:
 - *The 'view-mode' Media Feature* <https://www.w3.org/TR/view-mode/>
 - **Rescinded Recommendation:** a specification that W3C no longer endorses.

Maturity Levels of W3C Technical Reports (3)



WHATWG (1)

- *Web Hypertext Application Technology Working Group (WHATWG)* <https://whatwg.org/>
 - A community committed to the evolution of the Web that develops standards implementable in web browsers.
 - Pronunciation: *what-wee-gee, what-wig, what-double-you-gee*
 - Lásd: *How do you spell and pronounce WHATWG?*
<https://whatwg.org/faq#spell-and-pronounce>
 - Further information: *WHATWG – FAQ*
<https://whatwg.org/faq>

WHATWG (2)

- Standards:
 - DOM <https://dom.spec.whatwg.org/>
 - Fullscreen API <https://fullscreen.spec.whatwg.org/>
 - HTML <https://html.spec.whatwg.org/>
 - URL <https://url.spec.whatwg.org/>
 - WebSockets <https://websockets.spec.whatwg.org/>
 - XMLHttpRequest <https://xhr.spec.whatwg.org/>
 - ...
- See: *WHATWG – Standards* <https://spec.whatwg.org/>

WHATWG (3)

- History:
 - It was founded by programmers of Apple, the Mozilla Foundation, and Opera Software in 2004 who were concerned about the W3C's activity related to the development of HTML.
- Operation:
 - Its operation is coordinated by the Steering Group whose current members are Apple, Google, Microsoft, and Mozilla.
- Participation:
 - Participation is open to the public.
 - See: *WHATWG – Participation* <https://participate.whatwg.org/>

WHATWG (4)

- Development model:
 - The WHATWG develops specifications called “Living Standards” that are continuously updated.
- Living standards are licensed under the CC BY 4.0 license.
<https://creativecommons.org/licenses/by/4.0/>
 - See: *WHATWG – Intellectual Property Rights Policy*
<https://whatwg.org/ipr-policy>

Size of the Web

- The total number of indexed web pages: ≈ 4 billion.
 - See: <https://www.worldwidewebsite.com/>
- The total number of web sites: > 1 billion.
 - See: *Netcraft – August 2025 Web Server Survey*
<https://www.netcraft.com/blog/august-2025-web-server-survey>

Wayback Machine (1)

- A service that allows people to visit archived versions of Web sites.
 - Website: <https://archive.org/web/>
 - It makes the phrase “The internet doesn't forget” true.
 - A sub-project of the Internet Archive project launched in 1996.
 - Contains over 2 petabytes of data compressed, or 150+ billion web captures, including content from every top-level domain, 200+ million web sites, and over 40 languages.
- Further information:
 - <https://help.archive.org/help/wayback-machine-general-information/>
 - <https://help.archive.org/help/using-the-wayback-machine/>
 - <https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/>

Wayback Machine (2)

- Collects web pages that are publicly available.
- When a dynamic page contains forms, JavaScript, or other elements that require interaction with the originating host, the archive will not contain the original site's functionality.

Wayback Machine (3)

- Example:
 - Snapshots saved for <https://www.w3.org/>:
 - https://web.archive.org/web/*/https://www.w3.org/
 - Snapshot saved on September 3, 2003 at 14:02:22:
 - <https://web.archive.org/web/20030903140222/https://www.w3.org/>

Wayback Machine (4)

- A useful feature: **Save Page Now**
<https://web.archive.org/save/>
 - Capture a web page as it appears now for use as a trusted citation in the future.
 - Saves a specific web page one time only.