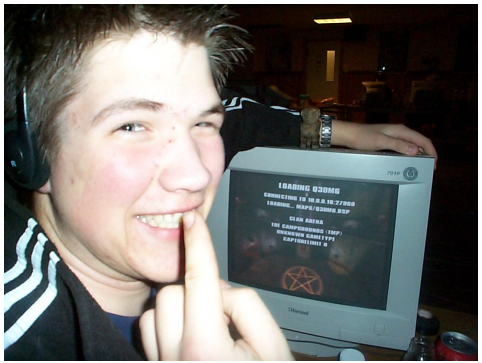
Improving interop with web-platform-tests

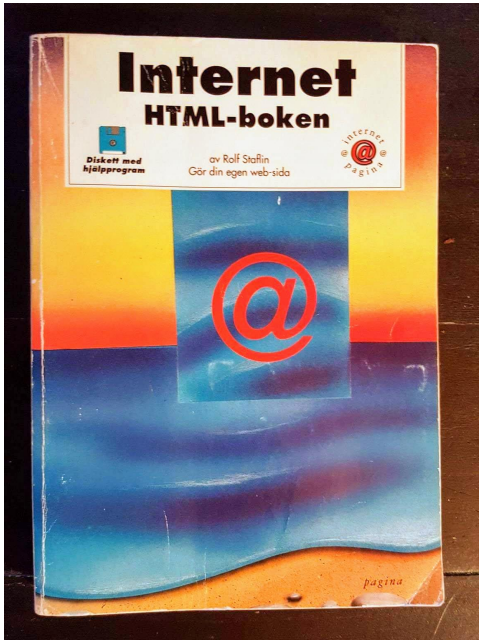
Philip Jägenstedt <foolip@chromium.org> at Web Engines Hackfest 2017

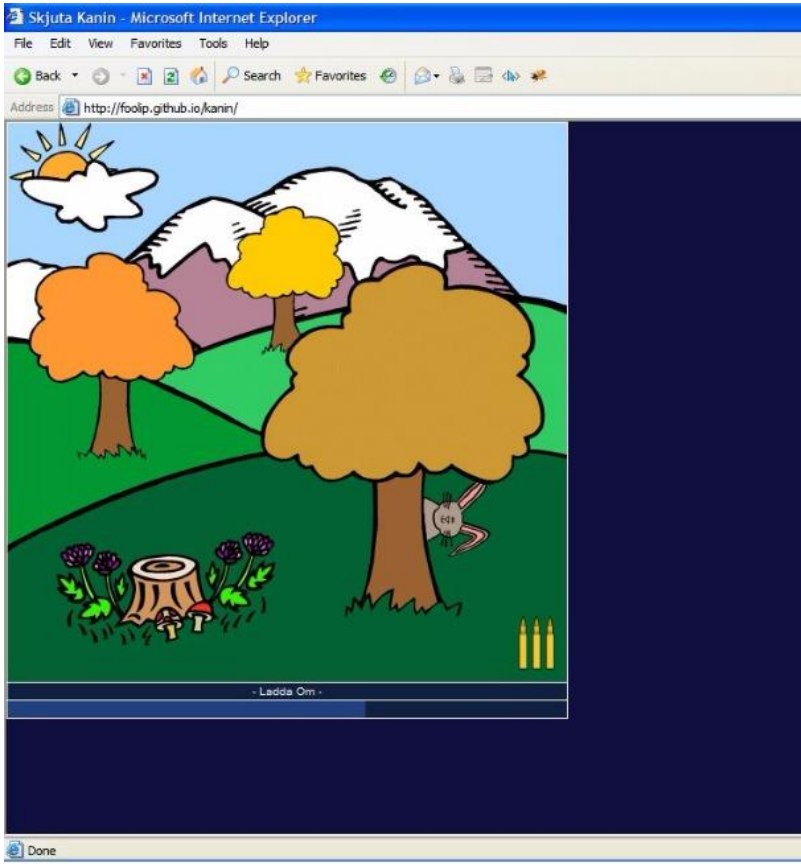
Slides: bit.ly/hackfest-wpt

Why care?

Joy




Internet
HTML-boken
Diskett med hjälpprogram
av Rolf Staflin
Gör din egen web-sida
pagina


Skjuta Kanin - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://foolip.github.io/kanin/
- Ladda Om -
Done

Mortality

Choice

Web Engines Hackfest
HACKFEST VENUE SPONSORS COC CONTACT
2-4 October | A Coruña, Galicia, Spain
WEB ENGINES HACKFEST 2017
THE HACKFEST

Web Engines Hackfest
2-4 October | A Coruña, Galicia, Spain
WEB ENGINES HACKFEST 2017

Web Engines Hackfest
HACKFEST VENUE SPONSORS COC CONTACT
2-4 October | A Coruña, Galicia, Spain
WEB ENGINES HACKFEST 2017
THE HACKFEST

Web Engines Hackfest
2-4 October | A Coruña, Galicia, Spain
WEB ENGINES HACKFEST 2017
THE HACKFEST

Web Engines Hackfest
2-4 October | A Coruña,
WEB ENGIN HACKFEST 2
THE HACKFEST
Who is attending?

Pain

Lambtaco
@Lambtacotweets

Follow

I noticed that i Microsoft Edge. Seems to work fine in Chrome and Firefox. Try a different browser?

> **AlextheTroller** @Alex_theTroller
> @Lambtacotweets Yo Alex here, decided to give 2.0 a try and the controls are iffy, enable hold to dash and try to jump left while dashing.

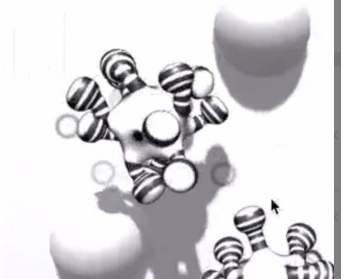
6:51 PM - 20 Sep 2017

Lars Berg
@laserberg

Follow

knobby rolls
larsberg.net/sketches/2017/ …

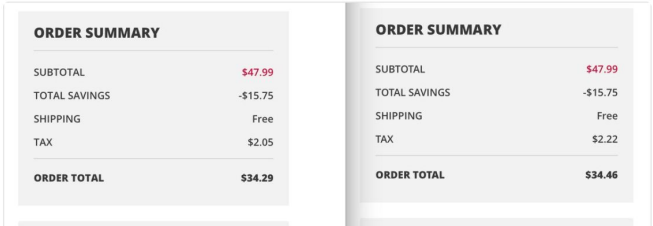
I made this in chrome & was bummed how different it sounds in FF & safari
#threejs #tonejs #cannonjs



Thom McGrath
@tekcor

Follow

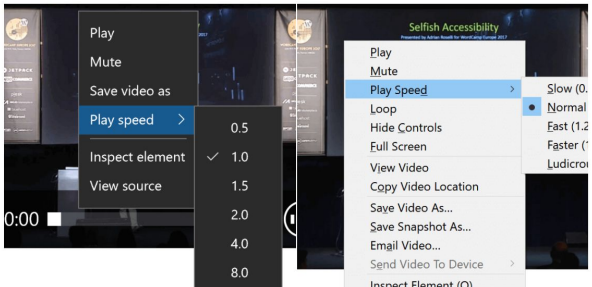
Left: Chrome. Right: Safari. Two different tax calculations for the same address.



Adrian Roselli
@aardrian

Follow

Playback speed for <video> content works only in Windows, only in Firefox and Edge.
rosel.li/092217
Safari test by @scottohara.



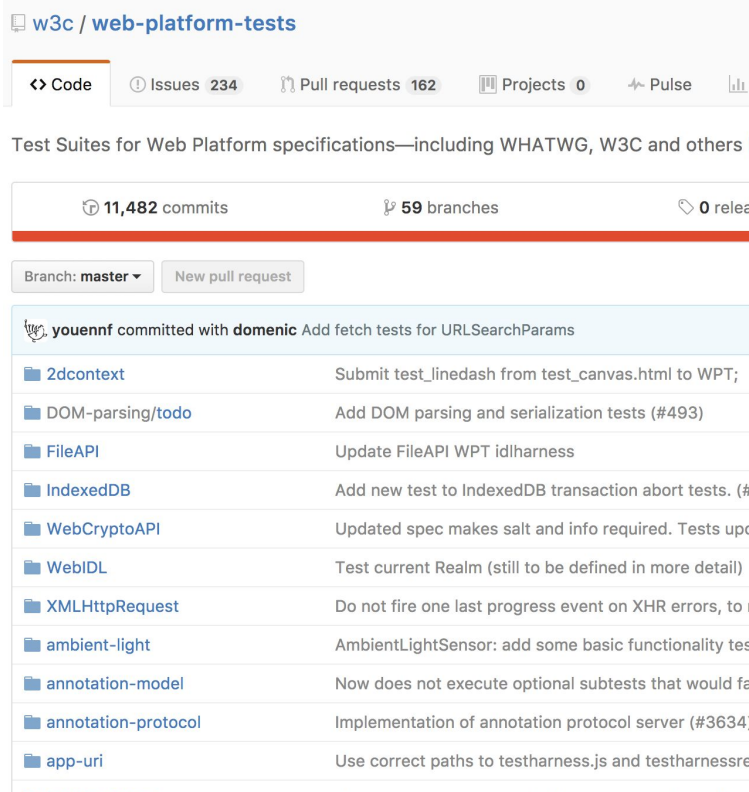
What is the problem?

Tragedy of the commons

- We usually think of our browser as the product
- Web developers want a cohesive platform
- Let’s think of the **platform as the product**
- Neglected, so improving is easy
- Apply basic software engineering

WPT one year ago:

- No one's primary test suite
- Harder to use than non-shared tests
- No continuous integration
- No dashboard/waterfall
- Out of sync with specs



Progress so far

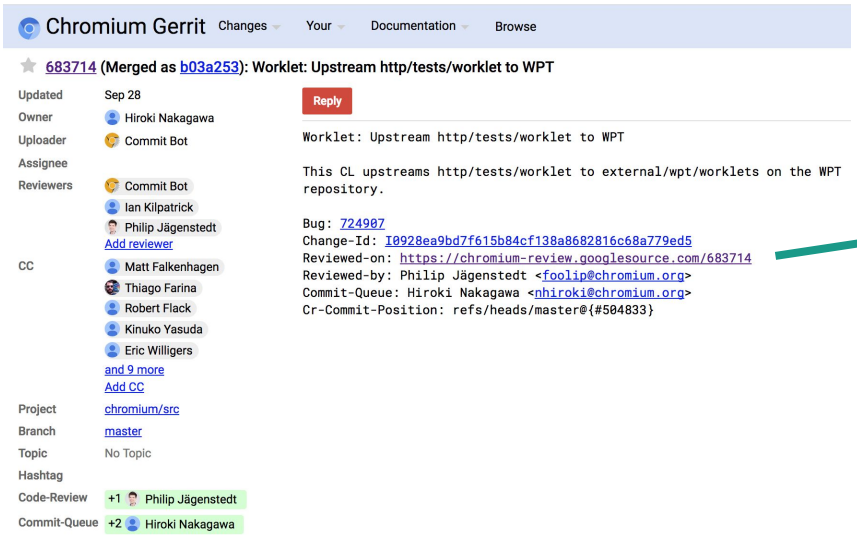
2-way sync

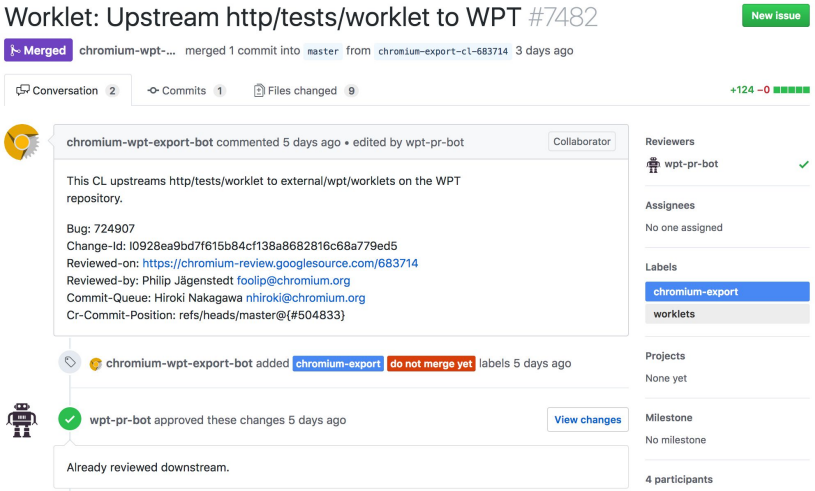
WPT changes imported into Chromium within 12 hours

Chromium changes exported to WPT within 1 hour

Humans not in the loop, but on rotation if needed

2-way sync (export)

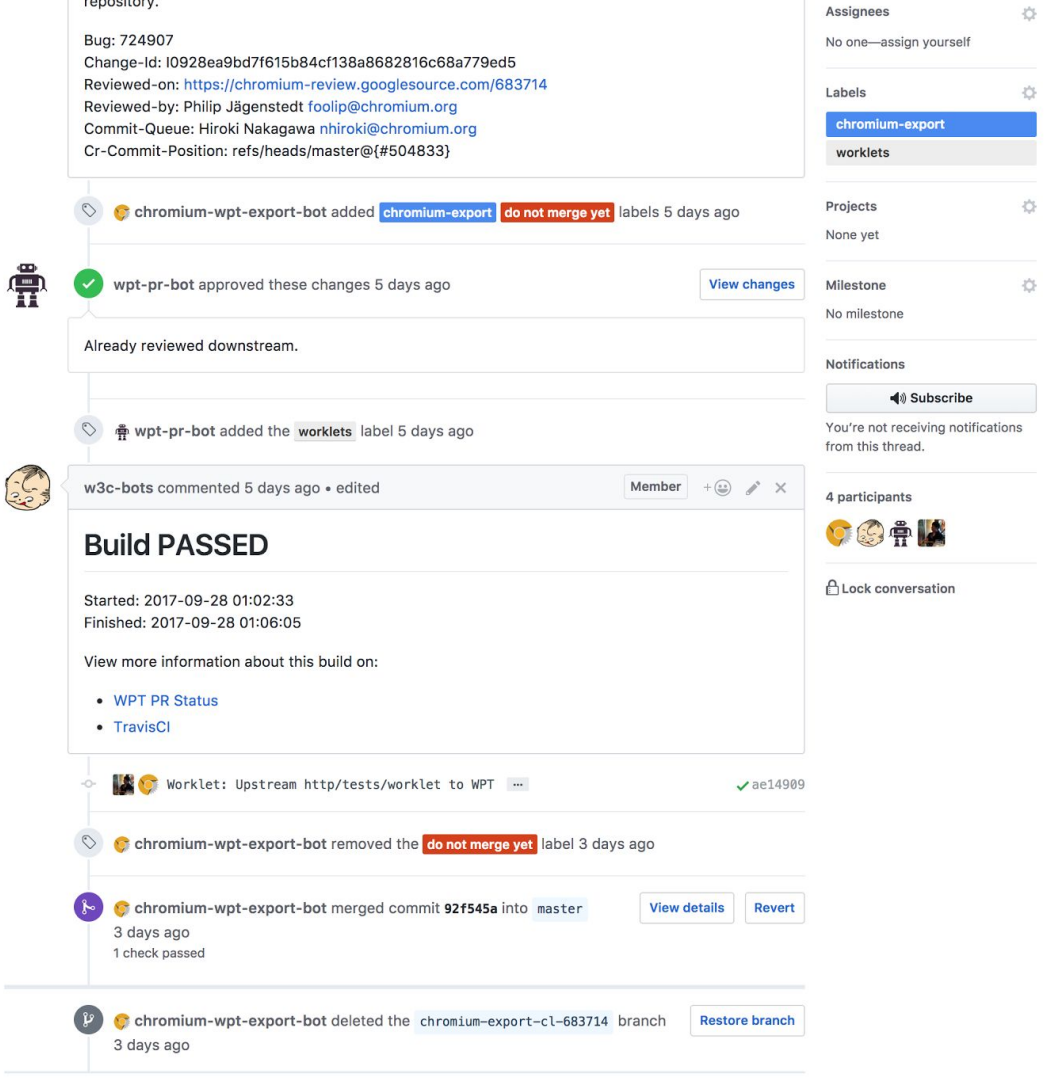




Per-PR testing

Catches unstable tests

Does not require tests to pass



wpt.fyi

web-platform-tests dashboard

Root About

Search test files, like cors/allow-headers.htm

WPT /

Data below are intended for web platform implementers and do not contain useful metrics for evaluation or comparison of web platform features. Also note that tested Edge and Safari are not pre-release versions (#109, #110).

Spec	chrome 62.0 linux 3.16 @de6ce4a47f Sep 30 2017	edge 15 windows 10 @d4ccae09e1 Sep 29 2017	firefox 57.0 linux 3.16 @de6ce4a47f Sep 30 2017	safari 10 macos 10.12 @de6ce4a47f Sep 30 2017
/2dcontext	1818 / 1912	1469 / 1742	1843 / 1912	1760 / 1908
/FileAPI	491 / 499	374 / 493	494 / 499	445 / 497
/IndexedDB	1578 / 1579	763 / 1312	1530 / 1579	1270 / 1579
/WebCryptoAPI	22858 / 35622	959 / 16961	17305 / 37253	119 / 1250
/WebIDL	436 / 445	249 / 445	440 / 445	191 / 297
/XMLHttpRequest	1483 / 1563	734 / 1096	1518 / 1564	814 / 1373
/accelerometer	3 / 71	3 / 71	3 / 71	3 / 71
/ambient-light	3 / 33	3 / 33	3 / 33	3 / 33
/apng	0 / 1	0 / 1	1 / 1	1 / 1
/assumptions	12 / 17	15 / 17	16 / 17	14 / 17
/audio-output	12 / 13	3 / 14	2 / 13	2 / 13
/background-fetch	23 / 179	14 / 72	23 / 179	7 / 16
/battery-status	34 / 36	4 / 10	4 / 11	4 / 11
/beacon	51 / 101	55 / 67	95 / 101	21 / 100
/bluetooth	1 / 2	0 / 1	1 / 2	1 / 2
/clear-site-data	21 / 21	12 / 14	13 / 19	7 / 17
/clipboard-apis	12 / 31	3 / 31	3 / 31	2 / 12
/compat	5 / 19	16 / 19	19 / 19	15 / 19

WPT improvements in Chromium

dom/	79% → 99%
editing/	0% → 90%
html/	74% → 93%
selection/	16% → 99%

Standards

The WHATWG Blog

Please leave your sense of logic at the door, thanks!

Improving interoperability

January 30th, 2017 by Philip Jägenstedt in WHATWG

The goal of the WHATWG's Living Standards is to achieve interoperable implementations. With an ever-evolving web platform, we want changes to our standards to reach all implementations quickly and reliably, but from time to time there have been mishaps:

- Two table-related interfaces were ignored by implementers for a decade, but when finally removed from the standard it turned out that WebKit had just added them. Sorry!
- In 2007, the serialization algorithm for <pre> was changed to add an extra newline, so that it would roundtrip correctly. It was implemented in Presto (Opera) in 2012 (and the standard was tweaked), then implemented in Gecko in 2016, but they found it was not web compatible, so it was reverted. Oh well.
- We changed how non-empty URL attributes are reflected with support from two implementers, but nobody followed through with the change and it was later reverted. Ahem.

Three months ago, we changed the process for the HTML Standard to encourage writing tests and filing browser bugs for normative changes. (Normative means that implementations are affected.) This was the first step on a path towards improving interoperability and shortening the feedback cycle, and it has thus far exceeded our own expectations:

Pages

Home

FAQ

IRC

Code of Conduct

Submit an Article

Copyright

Archives

July 2017

June 2017

April 2017

January 2017

November 2016

September 2016

August 2016

Standards

80+ specs now have a policy to ask for tests with normative changes

WHATWG (16), Service Workers, Web Performance WG (12), IndexedDB, Pointer Lock, Payment Request, Web Animations, WebVTT, FX Task Force (6), CSS WG (34), SVG WG (8), Content Security Policy

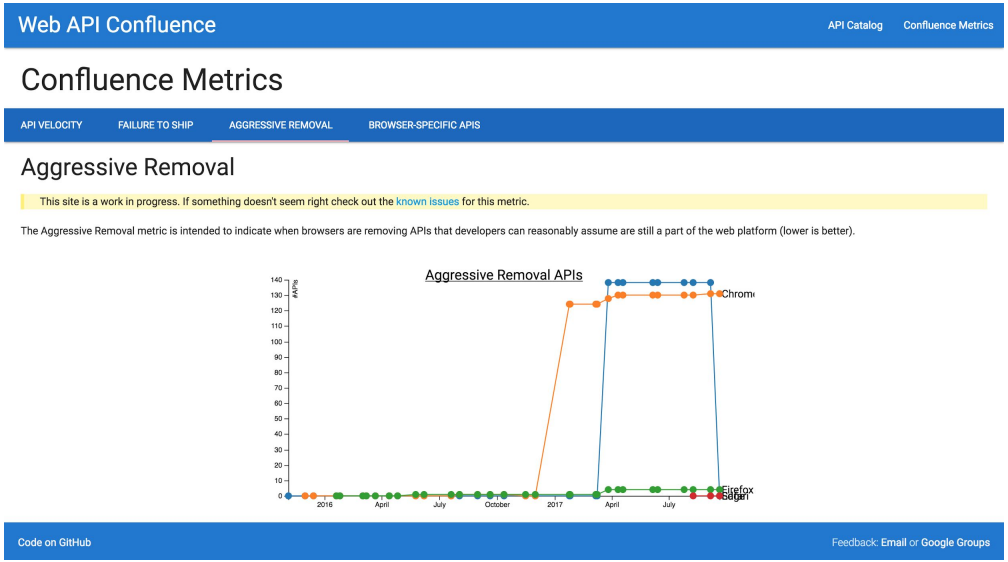
100+ remain

Non-WPT bonus

Visualizing API support as seen by traversing JS global object

Time-based metrics

web-confluence.appspot.com



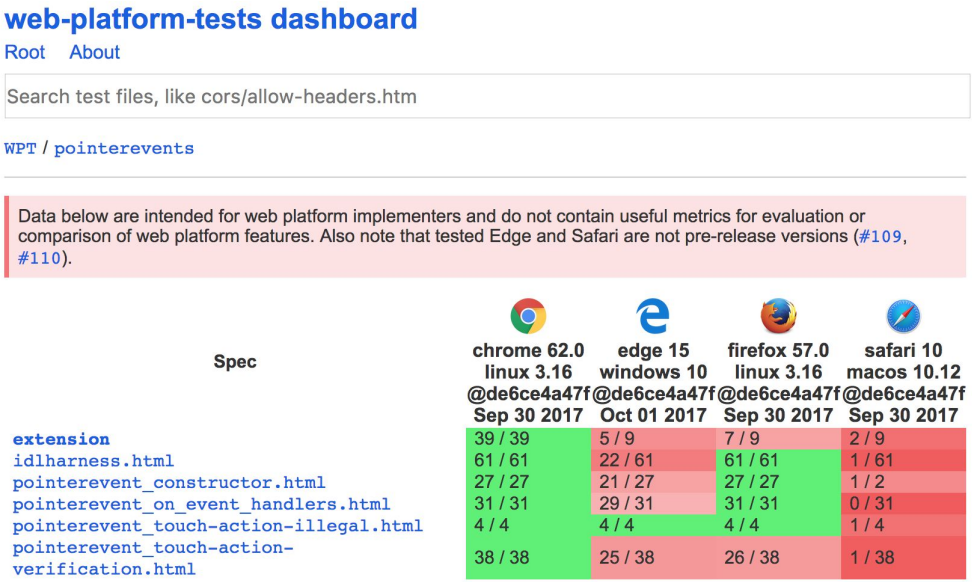
What’s next?

Automation

pointerevents/ has 71 manual tests

Automate input with WebDriver

Also: Permissions, WebRTC



web-platform-tests dashboard

Root About

Search test files, like cors/allow-headers.htm

WPT / pointerevents

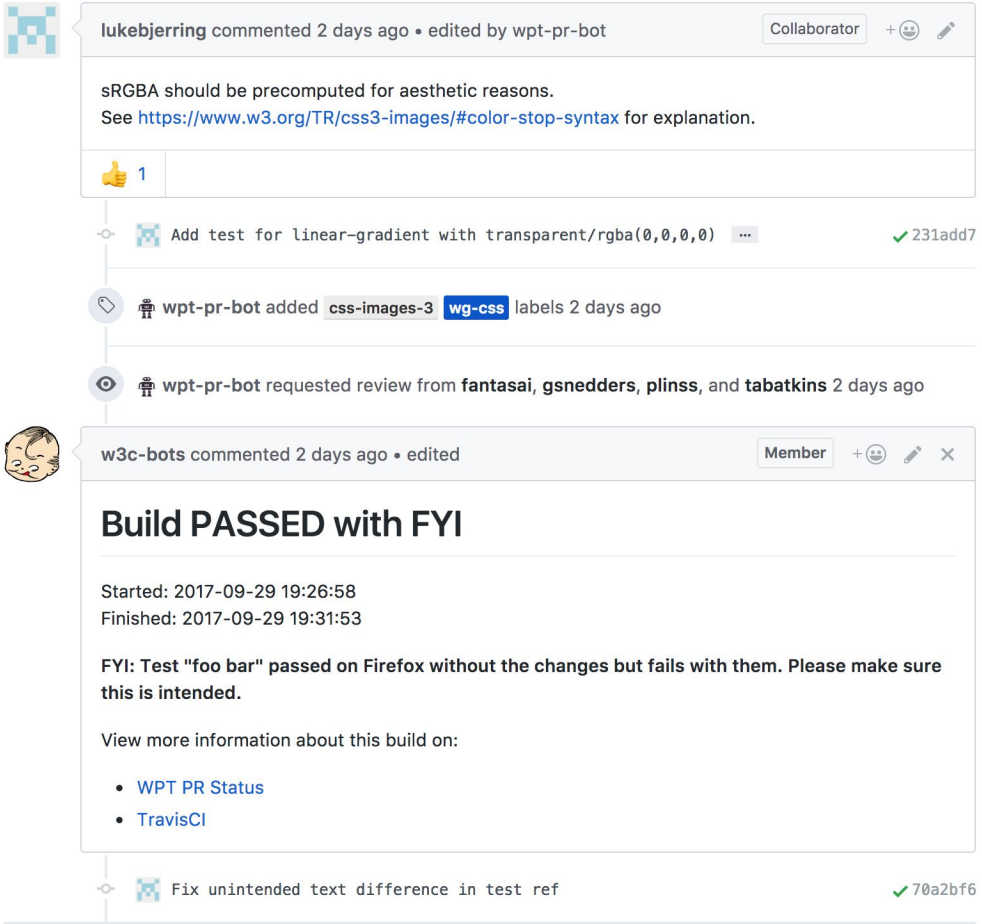
Data below are intended for web platform implementers and do not contain useful metrics for evaluation or comparison of web platform features. Also note that tested Edge and Safari are not pre-release versions (#109, #110).

Spec	chrome 62.0 linux 3.16 @de6ce4a47f Sep 30 2017	edge 15 windows 10 @de6ce4a47f Oct 01 2017	firefox 57.0 linux 3.16 @de6ce4a47f Sep 30 2017	safari 10 macos 10.12 @de6ce4a47f Sep 30 2017
extension	39 / 39	5 / 9	7 / 9	2 / 9
idlharness.html	61 / 61	22 / 61	61 / 61	1 / 61
pointerevent_constructor.html	27 / 27	21 / 27	27 / 27	1 / 2
pointerevent_on_event_handlers.html	31 / 31	29 / 31	31 / 31	0 / 31
pointerevent_touch-action-illegal.html	4 / 4	4 / 4	4 / 4	1 / 4
pointerevent_touch-action-verification.html	38 / 38	25 / 38	26 / 38	1 / 38

More per-PR testing

Test regressions should be intentional.

mockup →



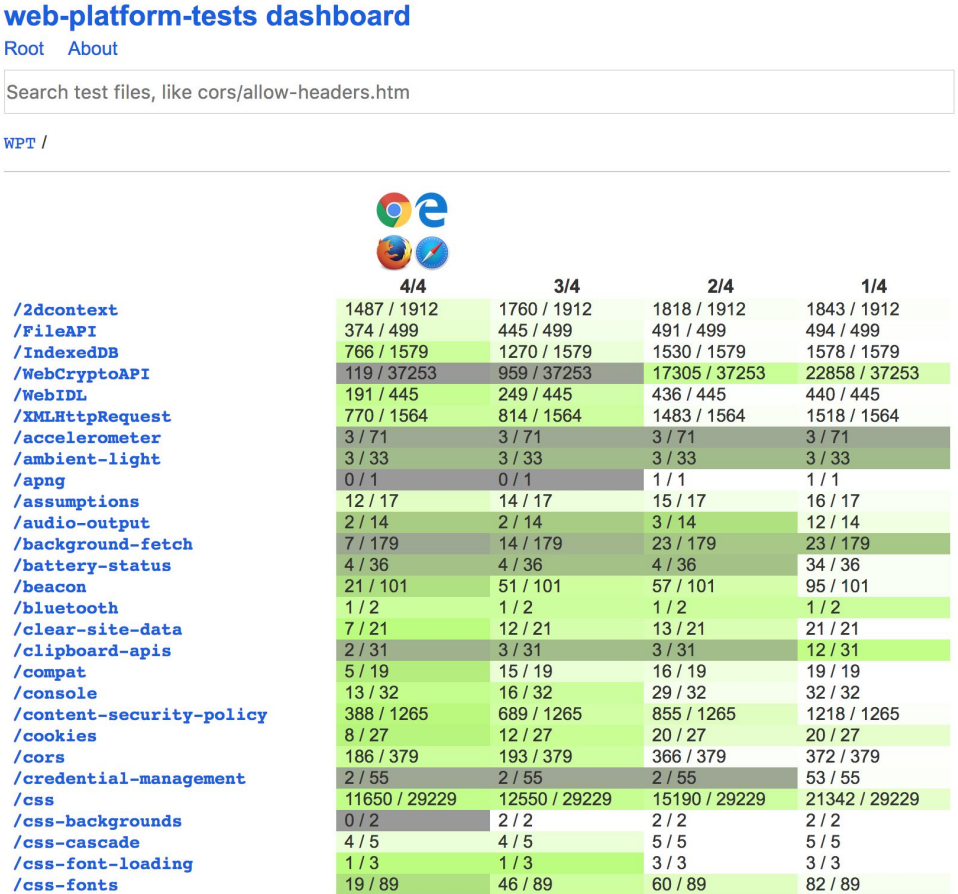
More wpt.fyi

An “interop health metric” where the incentives always point towards interop.

Per-browser reports without summary numbers, to be used for prioritization.

Also investigating time-based metrics.

mockup →



web-platform-tests dashboard

Root About

Search test files, like cors/allow-headers.htm

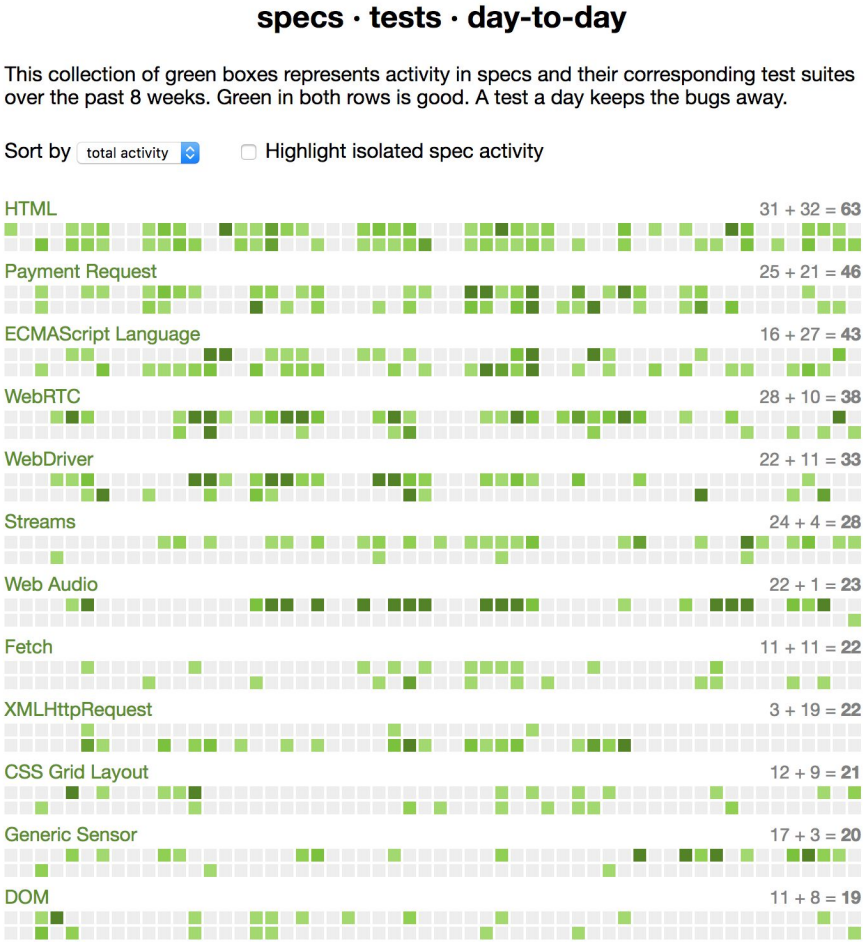
WPT /

	4/4	3/4	2/4	1/4
/2dcontext	1487 / 1912	1760 / 1912	1818 / 1912	1843 / 1912
/FileAPI	374 / 499	445 / 499	491 / 499	494 / 499
/IndexedDB	766 / 1579	1270 / 1579	1530 / 1579	1578 / 1579
/WebCryptoAPI	119 / 37253	959 / 37253	17305 / 37253	22858 / 37253
/WebIDL	191 / 445	249 / 445	436 / 445	440 / 445
/XMLHttpRequest	770 / 1564	814 / 1564	1483 / 1564	1518 / 1564
/accelerometer	3 / 71	3 / 71	3 / 71	3 / 71
/ambient-light	3 / 33	3 / 33	3 / 33	3 / 33
/apng	0 / 1	0 / 1	1 / 1	1 / 1
/assumptions	12 / 17	14 / 17	15 / 17	16 / 17
/audio-output	2 / 14	2 / 14	3 / 14	12 / 14
/background-fetch	7 / 179	14 / 179	23 / 179	23 / 179
/battery-status	4 / 36	4 / 36	4 / 36	34 / 36
/beacon	21 / 101	51 / 101	57 / 101	95 / 101
/bluetooth	1 / 2	1 / 2	1 / 2	1 / 2
/clear-site-data	7 / 21	12 / 21	13 / 21	21 / 21
/clipboard-apis	2 / 31	3 / 31	3 / 31	12 / 31
/compat	5 / 19	15 / 19	16 / 19	19 / 19
/console	13 / 32	16 / 32	29 / 32	32 / 32
/content-security-policy	388 / 1265	689 / 1265	855 / 1265	1218 / 1265
/cookies	8 / 27	12 / 27	20 / 27	20 / 27
/cors	186 / 379	193 / 379	366 / 379	372 / 379
/credential-management	2 / 55	2 / 55	2 / 55	53 / 55
/css	11650 / 29229	12550 / 29229	15190 / 29229	21342 / 29229
/css-backgrounds	0 / 2	2 / 2	2 / 2	2 / 2
/css-cascade	4 / 5	4 / 5	5 / 5	5 / 5
/css-font-loading	1 / 3	1 / 3	3 / 3	3 / 3
/css-fonts	19 / 89	46 / 89	60 / 89	82 / 89

More testing in standards

Visualizing the
day-to-day work
in specs and tests

foolip.github.io/day-to-day/



Make WPT the default in Chromium

Triage new test failures

Upstream as much of LayoutTests as possible

- ~70k files, compared to ~60k in WPT now
- Challenge: duplication and pre-fork WebKit tests

Can we relax?

No!
(trick question)

You can help!

- Write and use wpt for your work
- Tell us about what's not working
- Make 2-way sync work for everyone
  - Automatic 2-way sync in Gecko is coming!
  - Edge?
  - WebKit?

We can do it! \o/