

# URLs



**The Current State of URLs is Disastrous**

# History of URLs

1994: RFC 1738

1998: curl

2005: RFC 3986

2010: IETF

2011: IETF specification

All downhill from here

**Writing a URL today, should the syntax be presumed to still work in 30 years?**

# URL problems in 2018

“I entered this URL in my shell”

“This website sends me this redirect URL”

“I copied this URL from my browser”

# I entered this URL in my shell

Non-ascii byte codes sent as-is; browsers encode them  
“reserved” characters used; curl uses them for globbing

IDN 2003 vs 2008 vs RFC 3987 (CVE-2016-8625)

“curl dont work with russian charaters in Windows” (#2454)

# This website sends me this redirect URL

Wrong amount of slashes

White space

Non-ascii byte codes presumed to get re-encoded

Backslashes\*

\* = currently not supported by curl

# I copied this URL from my browser

The “input field” is not a URL

Userinfo interop issues (@, :, # part of user/password)

Weirdo numerical IP addresses

Non-ascii components

“curl doesn't like `http://✂.ws`” (#2429)



# Where do we go?





# Will this be fixed?

Not anytime soon

No IETF action – hot potato (IETFers are WHATWG skeptics)

WHATWG is very hard to affect (by me)

WHATWG sees no problems (WHATWGs are IETF skeptics)

Some tools/libraries buy into WHATWG URL

Many IETF people pretend WHATWG URL doesn't exist

# Why not just go WHATWG URL for curl?

Importance decreases



Backwards compatibility to curl users

Catch-22 with browser agenda

The specification *moves*

Craziness (“unlimited amounts” etc)

Hard to parse spec language

# Embracing the hybrid URL

RFC 3986, except when it's not

WHATWG URL, except when it's not

IDN 2008, mostly

Wait on and work for... the grand unified URL spec

<https://github.com/bagder/docs/blob/master/URL-interop.md>