

# yt-dlp and gallery-dl

(or: *“If you liked it, you should’ve made a copy”*)



**Tom Ryder**

[tom@sanctum.geek.nz](mailto:tom@sanctum.geek.nz)

<https://sanctum.geek.nz/>

# Cloud?

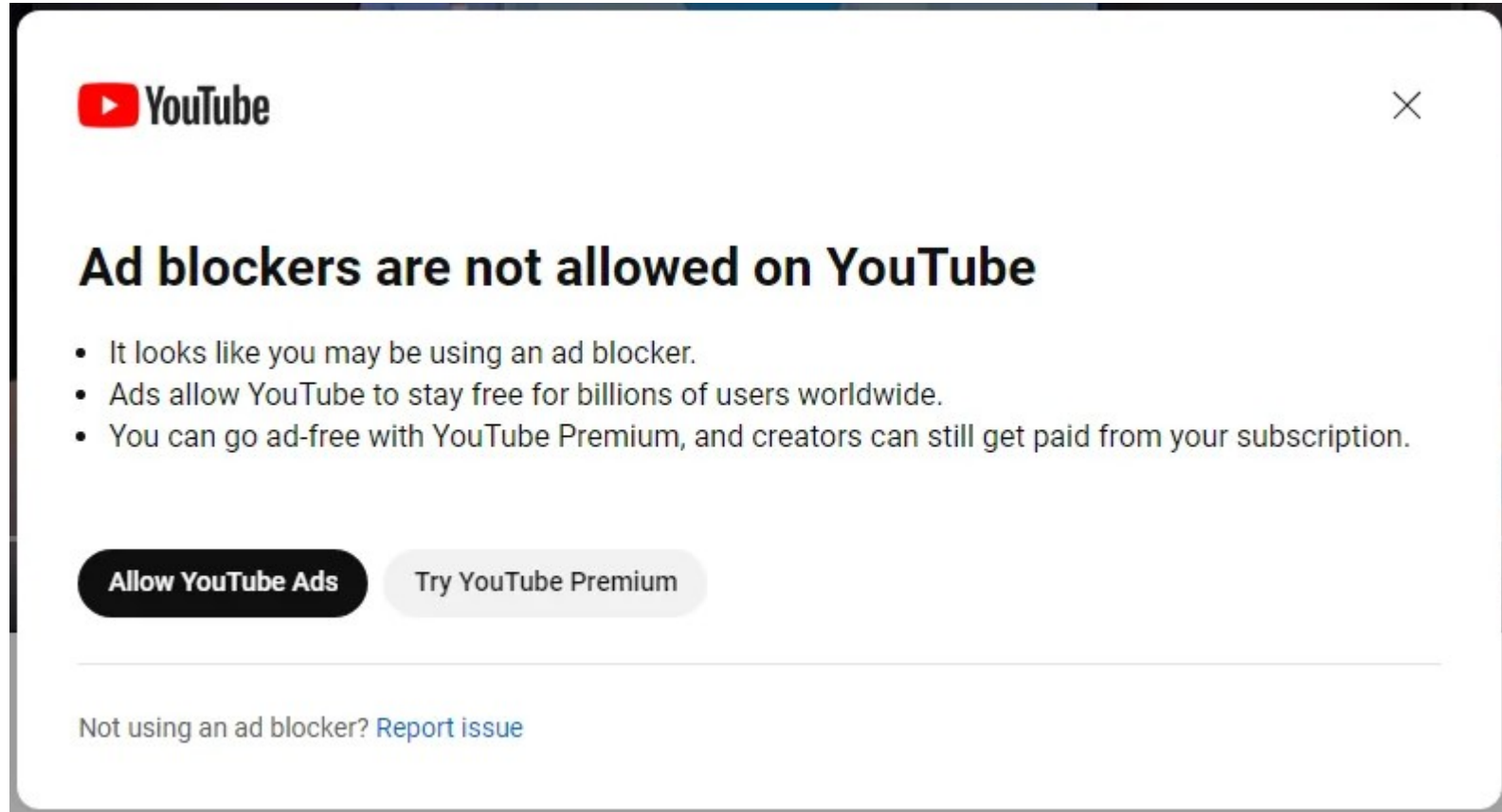


# Not a fan



My Computer  
(not yours)

# The indignity of it!



# Ethics

- I'm not going to tell you not to pay for YouTube or other media services.
- Go ahead and do that, especially if you can support creators directly.
- But it's not just about the ads, it's also about availability, and actually *having* your data.
- Streaming is not good enough.

# The internet is forever...



Hi [REDACTED]

We have reviewed your content and found severe or repeated violations of our [Community Guidelines](#). Because of this, we have removed your channel from YouTube.

We know this is probably very upsetting news, but it's our job to make sure that YouTube is a safe place for all. If we think a channel severely violates our policies, we take it down to protect other users on the platform - but if

© 2015 YouTube, LLC. All rights reserved. Privacy Policy Terms of Service

# ...until it isn't



This video is no longer available because the YouTube account associated with this video has been closed.

# Dust in the wind

- Internet content is *ephemeral*.
- The more unique your interest in something is, the more important it is that you save it.
- People are *really bad* at keeping stuff online.
- If individuals are bad, companies are even worse.
  - I hate [link rot](#), but it's just a fact of life.

# Right click, “Save as...”

- Ah, if only it were so simple!
- Saving the media *easily* without logging in undermines these sites’ business models.
- So they make it *difficult*, and it requires dedicated software tools.
- Enter: **yt-dlp** and **gallery-dl**.

# yt-dlp

- Download **videos**
- Supports many different sites:
  - YouTube, DailyMotion, BBC, Reddit...
- Written in Python, [available on PyPI](#)
- Actively maintained
- Skips a download if you've already got it

# yt-dlp—You spelled that wrong!

- **youtube-dl** was the original program.
  - youtube-dl was unmaintained for a long time; although it works, it's easily detected and gets rate-limited on YouTube.
- Use **yt-dlp** instead.
  - It's a drop-in replacement for youtube-dl.
  - It's actively maintained.

# yt-dlp—In action

```
~$ yt-dlp 'https://www.youtube.com/watch?v=ygI-2F8ApUM'
[youtube] Extracting URL: https://www.youtube.com/watch?v=ygI-2F8ApUM
[youtube] ygI-2F8ApUM: Downloading webpage
[youtube] ygI-2F8ApUM: Downloading ios player API JSON
[youtube] ygI-2F8ApUM: Downloading android player API JSON
[youtube] ygI-2F8ApUM: Downloading m3u8 information
[info] ygI-2F8ApUM: Downloading 1 format(s): 247+251
[download] Destination: BRODYQUEST [ygI-2F8ApUM].f247.webm
[download] 100% of 24.32MiB in 00:00:01 at 16.48MiB/s
[download] Destination: BRODYQUEST [ygI-2F8ApUM].f251.webm
[download] 100% of 4.04MiB in 00:00:00 at 26.97MiB/s
[Merger] Merging formats into "BRODYQUEST [ygI-2F8ApUM].webm"
Deleting original file BRODYQUEST [ygI-2F8ApUM].f251.webm (pass -k to keep)
Deleting original file BRODYQUEST [ygI-2F8ApUM].f247.webm (pass -k to keep)
~$ █
```

# yt-dlp—Download a video

```
$ yt-dlp 'https://www.youtube.com/watch?v=sgKK0YRQyMQ'
```

- Automatically picks the best format available!
  - This might be more than you need.

# yt-dlp—Channels and playlists

- Pass yt-dlp a channel or playlist URL instead of a single video...
- Does what you want, gets videos in multiple files:

```
$ yt-dlp 'https://www.youtube.com/@Computerphile'  
$ yt-dlp 'https://www.youtube.com/watch?  
list=PL5QgoWHRW35Jf-Vxao7ceYwvXGc1Mk0yu'
```

# yt-dlp—Limit resolution/filesize

- This video's 4K version is 5.6 GiB in size...
- Maybe the 720p version is good enough for you? It's only 225 MiB.

```
$ yt-dlp --format='best[height<=720]'  
'https://www.youtube.com/watch?v=czBL7JQl248'
```

# yt-dlp—Audio only

```
$ yt-dlp --extract-audio 'https://www.youtube.com/watch?v=4a8CAS6I2DI'
```

- Get audio in .mp4 (AAC) or .opus format
  - Use the --format option to choose
- Add the --embed-metadata option to get tags for your media player

# yt-dlp—mpv support

- Watch videos directly with **mpv**:

```
$ mpv 'https://www.youtube.com/watch?v=sgKK0YRQyMQ'
```

- Might need configuring in `~/.config/mpv.conf`:

```
script-opts-append="ytdl_hook-ytdl_path=yt-dlp"
```



# yt-dlp—Mimic a browser

- yt-dlp can use **PhantomJS** to run JavaScript, emulating a browser.
- This might help avoid scraping counter-measures on some sites.
- You have to install it separately.



# gallery-dl

- Download **images** (usually in sets)
- Supports many different sites:
  - Imgur, Instagram, Patreon, Reddit...
- Written in Python, [available on PyPI](#)
- Actively maintained
- Skips a download if you've already got it

# gallery-dl—In action

```
"$ gallery-dl 'https://imgur.com/gallery/E0Yzn46'  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_001_g68dNsZ.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_002_dM9gw6P.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_003_QaJpQZN.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_004_LJ4K2Fr.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_005_HcizRtM.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_006_mNk1X0j.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_007_Y29JXpv.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_008_2DviXz8.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_009_7xbacaz.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_010_bMwFh3.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_011_qv4FiIo.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_012_yKlWHL.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_013_wRimjx.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_014_f0s1XAAb.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_015_oRJ0Yt5.jpg  
./gallery-dl/imgur/E0Yzn46 - Top 100 me.020 pt. 1/imgur_E0Yzn46_016_kcMlrGG.jpg
```

# gallery-dl—Use your cookies

- Instagram often requires login to download, so:
  - Log in to Instagram in browser
  - Export cookies to file ([Firefox](#), [Chrome](#))
  - Have gallery-dl use your cookies file for the download:

```
$ gallery-dl --cookies=cookies.txt  
'https://www.instagram.com/reel/CskI8aPu4ZR'
```

# gallery-dl—Rate-limit

- Request images only at a certain bitrate:

```
$ gallery-dl --rate-limit=500k  
'https://imgur.com/gallery/E0Yzn46'
```

- Might avoid tripping scraper blocks.

# Out of touch, out of time—1/2

- Both yt-dlp and gallery-dl are very much moving targets.
  - *Especially* yt-dlp.
- Site APIs and paths change, and they try to circumvent scraping.
  - *Especially* when the site has paid options.
- Your OS' versions might be out of date.
  - *Especially* if you prefer OS stability (Debian).

# Out of touch, out of time—2/2

- I suggest: keep them both in a **Python venv**, and update them regularly:

```
$ python3 -m venv dltool  
$ . dltool/bin/activate  
(dltool) $ pip install --upgrade gallery-dl yt-dlp
```



# Teaser for budding archivists

- So, you're getting the hang of this archiving thing, and you wonder to yourself:
  - “Why just videos and image galleries? Why not snapshot and archive whole web pages?”

# Teaser for budding archivists

- So, you're getting the hang of this archiving thing, and you wonder to yourself:
  - “Why just videos and image galleries? Why not snapshot and archive whole web pages?”
  - “Why not”, indeed?



# Questions? (Lightning-fast!)

- yt-dlp source code
- yt-dlp on PyPI
- gallery-dl source code
- gallery-dl on PyPI

**Email:** [tom@sanctum.geek.nz](mailto:tom@sanctum.geek.nz)

**Website:** <https://sanctum.geek.nz/>

**Fediverse:** [@tejr@mastodon.sdf.org](https://mstdn.social/@tejr)

