

1. Figure out where links to Unlisted pony videos exist

2. Accumulate links

3. Take links/video IDs and identify which are pony, unlisted, and pre-2017

4. Reformat into list of URLs to save

All sections are gone over in greater detail in the following pages. Note that nothing's set in stone, just what I've been able to pull from the thread and other stuff.

Ideal Guidelines

- Ideally want to **utilize/repurpose** existing resources as much as possible (scripts, archive sites (TPA), YouTube API, etc.) If we don't have to build from scratch, let's try not to.
- Need a **resource list(s)** of available scripts to use, what websites/forums/wikis to look at (and priorities of each), channels to look at, etc.
- Maybe eventually create processes/resources that **any anon can** use to **help** (ie. downloadable programs, browser extensions)

Priorities for scripts to be built (top = highest priority, bottom = lowest priority)

Priorities for places to be searched/scraped (")

1. Figure out/List where links to Unlisted pony videos exist

- a. Ancient websites and archives of said websites (desuarchive)
- b. Utilize Internet Archive/Wayback Machine for channels (and videos)
- c. Old forums (prioritizing oldest threads first, ex. [>>37165365](#))
- d. Annotations (HUGE untapped potential [>>37162070](#), [1](#), [2](#))
- e. Pyxia's site [>>37169367](#)
- f. Tumblr blogs (many are archived on Rome)
- g. >Reddit
- h. EqD (last priority, see [>>37168145](#))

2. Accumulate links

- a. **Build scrapers** for websites if they can be made
 - i. Utilize/Repurpose existing resources as much as possible so we don't make something that's already been made before
 - ii. Ex. Desuarchive scraper already built [37165580](https://github.com/37165580/Desuarchive)
 - iii.
- b. **People without any coding experience** go and manually search Wayback Machine, annotations, etc. anything that scrapers can't be built for.
 - i. Keep track of which channels/videos have already been scoured

3. Take links/video IDs and identify which are **pony**, **unlisted**, and **pre-2017**

Need **3 lists**:

- a. Pony & Unlisted & Pre-2017 (most dire obviously)
- b. Pony & Unlisted but NOT Pre-2017
- c. Pony but NOT Unlisted

- d. First **standardize** links, **dedupe**, and **identify** what has
 - i. already been saved
 - ii. already been found and evaluated
- e. For what's already been saved on TPA, there is [recursivedir.txt](#) which can be checked against using video IDs. Would be awesome if something could be built where a list of links can be thrown in and they are all checked (would be ideal to have the page downloaded locally and checked, rather than needing to load it every time, cuz it's really big)
 - i. Only URLs that begin with /eqd_yt/, /youtube/, and /quickchive/ would be needed for the recursivedir.txt file, the rest can all be thrown out as they're not videos or pertinent to this.
 - ii. Note: recursivedir.txt does not update automatically, but can probably ask for it to be updated. (Updated as of 2021-05-16)
- f. For what's already been found and evaluated, a masterlist of what's been found could be checked against (ideally something that can continually be updated like another spreadsheet or Ponepaste's RAW version or something rather than a new file each time)
- g. Then, **evaluate** which videos are **pony**. This will most likely have to be manually done.
- h. YouTube's API can export video links, their titles, creator, unlisted status, and upload date into a .csv file. These can then be uploaded into Sheets and then easily sorted, filtered, and checked.
 - i. If it's taking too long, may need to cut this part out, but there are many more videos unlisted than just pony videos.

4. Reformat into clean list of just URLs to save

- a. Give to Spikebot for TPA (just listed like <https://www.youtube.com/watch?v=UmPDdAr-gkc> or the full channel link). Can download videos on multiple threads at once, so can have multiple playlists/channels being downloaded at the same time.
- b. Download for this eventually [>>37169957](https://www.youtube.com/watch?v=UmPDdAr-gkc)
- c. Want to have redundancy and not have a single point of failure, but may want to prioritize TPA right now because it already has an existing framework, site, downloader bot, and anyone can access the data.
- d. For channels, many multiple channels have already been downloaded on TPA, but I don't think focusing on them should be a priority as of right now. If they have unlisted videos, they won't show up when you try and download the channel anyway. The unlisted videos only be in playlists the channel may have created (or otherwise been found).

Also, reason why I drew this up is because at least with Pastebin, everything was mostly contained to the board. With Mediafire, the only sites being searched on were pony and thus the stuff could be assumed to be pony (mostly). This doesn't have either of those conveniences and also has a lot more qualifications (only certain videos should be prioritized due to both time and space available). And links are *everywhere*.

I'd like to create a video on this so any anon can understand what's going on and how they can help if they'd like to. It'd mostly go over the framework in general, how to happen upon these websites, channels, forums, blogs, etc. to search, and how people without coding experience can help out.

stuff

[>>37161678](https://www.youtube.com/watch?v=UmPDdAr-gkc) important stuff

[>>37161982](https://www.youtube.com/watch?v=UmPDdAr-gkc), [>>37163016](https://www.youtube.com/watch?v=UmPDdAr-gkc), and [>>37165897](https://www.youtube.com/watch?v=UmPDdAr-gkc) video downloader for playlists I believe

[>>37165268](https://www.youtube.com/watch?v=UmPDdAr-gkc) video downloader (says is old and may be buggy)

[>>37163067](https://www.youtube.com/watch?v=UmPDdAr-gkc) and [>>37165307](https://www.youtube.com/watch?v=UmPDdAr-gkc) throw in a playlist and identify which videos are unlisted (at-risk)

[>>37165580](https://www.youtube.com/watch?v=UmPDdAr-gkc) guy who did the pastebin/mediafire stuff from before (thank you by the way, you're awesome)

>>[37165701](#) scraper for <https://unlistedvideos.com/>

>>[37165580](#) stuff for wayback machine scraping potentially

>>[37162136](#) annotations stuff

>>[37161906](#) comment downloader (no exact timestamps)

Submission form:

<https://docs.google.com/forms/d/17U6Eanhq-hynjVIZ0kl93q-UHWGnQrUXkRx-7sAw-jY/edit>