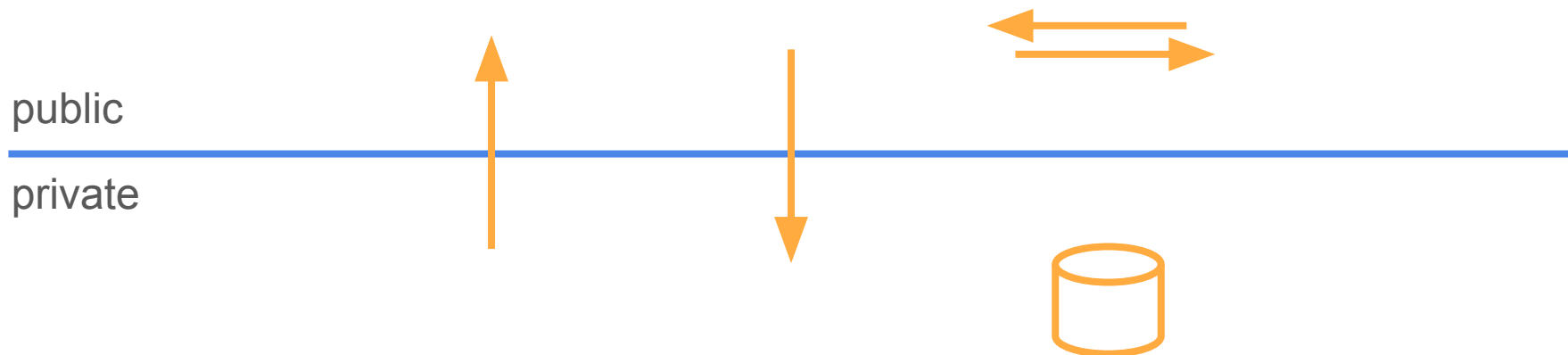# I CONTROL MY OWN DATA!

...so, what exactly can I do with it?

Nathan Willis

The free-software community talks about
"data sovereignty," "autonomy," and so forth
as *essential*

I've used takeout services *and* aftermarket tooling in the past: Google Reader + `ReaderIsDead`

Aside...

"How to rid yourself of Big Tech" guides ... not really helpful

public

private

I made a list & started requestin': Google, Yahoo, Facebook, Twitter, GitHub, Amazon, Instagram, Reddit, Identi.ca, Mastodon, GitLab, Discourse, Wikipedia, LinkedIn, Gpodder, Telegram, Firefox Sync, Signal, Slack, Matrix, etc, etc, etc....

Knowing how well this works matters

Can you actually take your data then live on your own?

If not, "sever all the connections you have with your friends and coworkers, destroy your data, lose everything you've posted" isn't much of a sales pitch

The plan:

- Look at what you can actually export

- Look at FOSS tooling & usability

- Case study on a social-network account (plus& offshoots)

- Some ideas about how the community can move the needle

What can you get?

- Several of the biggest service providers supply a static, local HTML browsing interface

- Most provide JSON … mostly without a schema

- Directly usable file formats are the exception (but exist!)

What can you get: Google Takeout

What can you get: Google Takeout

What can you get: Twitter archive

What can you get: Twitter archive

"Together, we're a thing of beauty, like a swan."

"On my own, I'll be like half a swan. All I'll do is make a big mess and die."

1. 'It's a Party and I'll Lie If I Want To' (2010), *Better Off Ted*, season 2, episode 12. Garfield Grove Productions. American Broadcasting Company, 17 August 2010.

What can you get: other highlights

Facebook: HTML index, posts, comments, media, etc seem complete

GitHub: Entire repos, message/issue archives, attachments....

LinkedIn: A bunch of CSV files....

Amazon: massive hierarchy of files, but with documentation

What can you get: a bunch of others

Instagram: Local HTML browser, original media, comments, saves/likes, followers/follows

Telegram: Desktop-only, decent HTML browser

Reddit: raw CSV

Element/Matrix: per-room HTML page; bad CSS; missing data

What can you get: GitLab

Not good: NONE.

What can you get: GitLab

Not good: NONE.

https://gitlab.com/gitlab-org/gitlab/-/issues/35790

What can you get: Firefox Accounts

Not good: NONE.

What can you get: Wikipedia

Not good: NONE.

What can you get: Identi.ca / pump.io

Not good: NONE.

What can you get: Identi.ca / pump.io

Not good: NONE.

API scraping is possible: http://repo.or.cz/pumpscripts.git

What can you get: Mastodon

Complicated: raw `ActivityPub` objects in JSON.

... plus *some* other things separately

What can you get: Mastodon
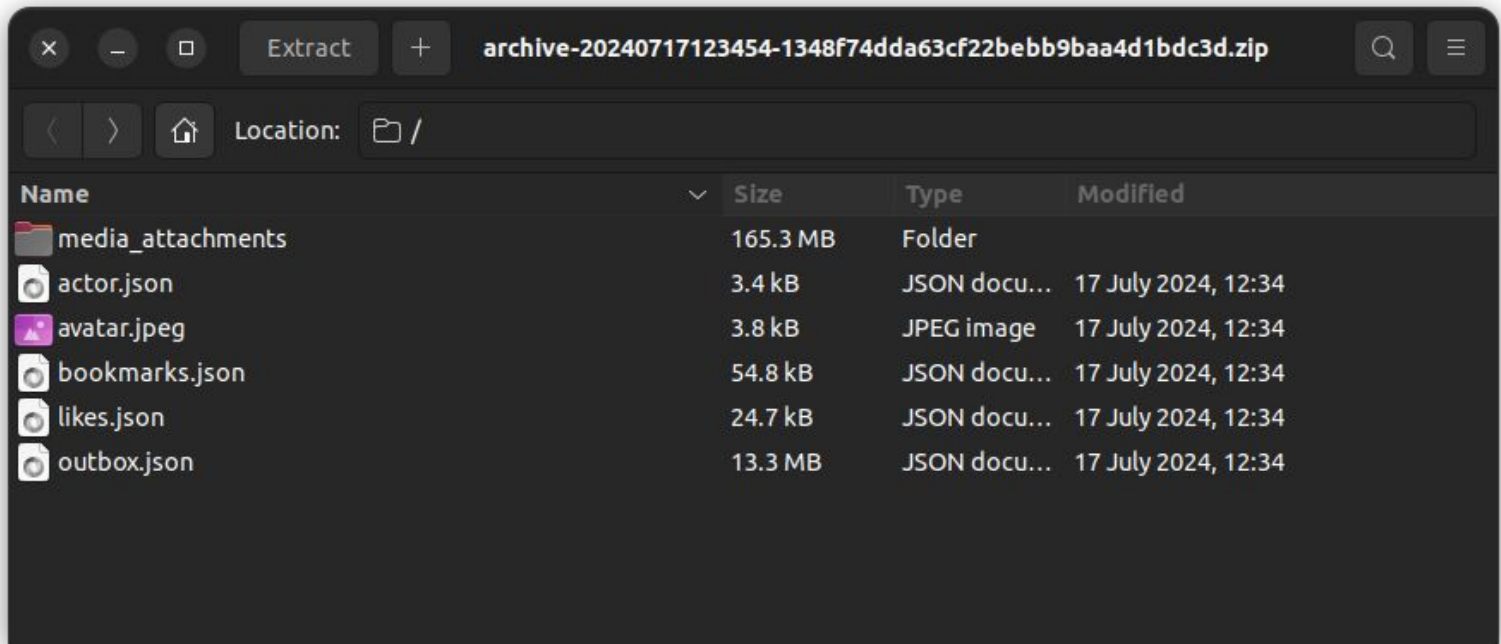
What can you get: Mastodon
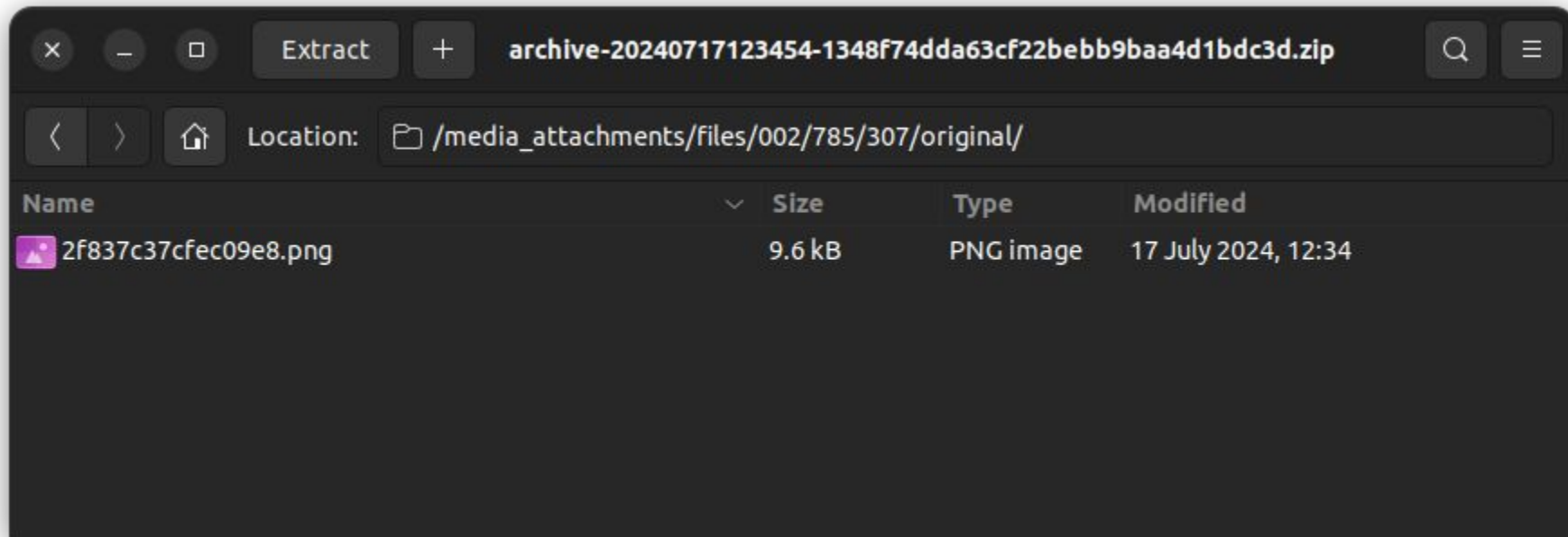
What can you get: Mastodon

Export

Media storage	162 MB	
Posts	6,295	
Follows	261	⬇ CSV
Lists	0	⬇ CSV
Followers	347	
You mute	83	⬇ CSV
You block	66	⬇ CSV
Domain blocks	0	⬇ CSV
Bookmarks	1,029	⬇ CSV

You can request an archive of your **posts and uploaded media**. The exported data will be in the ActivityPub format, readable by any compliant software. You can request an archive every 7 days.

What can you get: Discourse

Complicated: raw CSV.

What can you get: brief takeaways

On average, the big vendors are providing a far better user experience

This matters, because it is the nature of services to wind down, of servers to die, and of backups to fail

The Doings.

What do you want to *do* with your exported data?

- **Library**: save it, have it to search (who was telling me…)
- **Import**: data that's directly usable (e.g., office docs, contacts)
- **Migrate**: take the data into a rival service
- **Repurpose**: do something new on your own

The Doings: FOSS applications

In general, FOSS applications are not designed to work with large imports

- this is not a failing of the applications, it's an unaddressed use case

The Doings: map / location data

```json
{
    "geometry": {
    "coordinates": [
        -6.2353863,
        53.3352424
    ],
    "type": "Point"
    },
    "properties": {
    "date": "2024-03-24T16:15:55Z",
    "google_maps_url": "http://maps.google.com/?cid=112000413610799906245",
    "location": {
        "address": "Haddington Rd, Dublin, D04 E0C9, Ireland",
        "country_code": "IE",
        "name": "National Print Museum"
    }
    },
    "type": "Feature"
    },
```

The Doings: map / location data

Google's map data includes several distinct things
● places you have flagged/starred
● named lists of places you have made
● locations you have been
● map sets you have created to share
● ratings and reviews
● images

The Doings: FOSS applications

In general, FOSS applications are not designed to work with large imports

- this is not a failing of the applications, it's an unaddressed use case … or it has rough edges

Starts   10/30/2019 02:30:00 PM
Ends   10/30/2019 05:30:00 PM

Title   Data analysis
Starts   10/28/2019 01:00:00 PM
Ends   10/28/2019 05:00:00 PM

Title   Outline seminar
Starts   10/31/2019 10:30:00 AM
Ends   10/31/2019 01:30:00 PM

Title   Storage analysis
Starts   10/31/2019 03:30:00 PM
Ends   10/31/2019 05:00:00 PM

Title   RRDP: Statistics
Starts   11/15/2019 09:00:00 AM
Ends   11/15/2019 01:00:00 PM

Title   Letterpress induction
Starts   11/19/2019 02:00:00 PM
Ends   11/19/2019 04:00:00 PM

Title   s
Starts   11/18/2019 03:30:00 PM
Ends   11/18/2019 06:30:00 PM

The case study: social network

This is a really important one for people who are unhappy with service providers.

Let's do a deeper dive on Twitter, then extrapolate….

The case study: social network

Before we dive in …

I'm staunchly anti-"deletionist"

The case study: social network

Deletionism:

It's tempting to imagine this as striking back against the sinister owners/operators

Sort by: Best ⌄   🔍 Search Comments

[deleted] · 2y ago ·

Comment deleted by user

⊖   ⬆   ⬇   💬 Reply   ↗ Share

jorissels OP · 2y ago ·

oh wow! defenitly will check it out! Does it work wit Imessage aswell that you know of?

⬆ 0 ⬇   💬 Reply   ⧉ Award   ↗ Share   ···

⊕ 3 more replies

The case study: social network

Before we dive in …

I'm staunchly anti-"deletionist"

There are better alternatives.

The case study: social network

Before we dive in ...

I'm staunchly anti-"deletionist"

There are still good reasons to export!
Outages! EOL! Censors! Etc!

The case study: social network

Twitter export data to consider:
1. The tweets
2. Direct messages
3. The other accounts (follows/followers)
4. Likes / bookmarks

(Our options: Library, Import, Migrate, Repurpose)

Nate Neutrality Willis 🇺🇦
@n8willis

Your archive includes all the account data created up to the moment it was generated.

Date generated · May 17, 2024 at 12:32:09 PM GMT+1 · Estimated size · 187 MB

Home

Account

Tweets

Likes

Direct Messages

Safety

Personalization

Ads

Lists

Moments

Tweets

Tweets | **Replies** | Retweets | Community Tweets

Nate Neutrality Willis 🇺🇦 @n8willis · Mar 21, 2018
"You Can Call Me Baal"
#PaulSimonDeathMetalSongs 🔗

🔁 0   ♡ 1   📤

ID 976539960470843392

View on Twitter 🔗

Nate Neutrality Willis 🇺🇦 @n8willis · Mar 21, 2018
"Urn at the Right Time"
#PaulSimonDeathMetalSongs 🔗

🔁 0   ♡ 0   📤

ID 976538629463068672

View on Twitter 🔗

Nate Neutrality Willis 🇺🇦 @n8willis · Mar 21, 2018
"All Around the World, or the Scythe of Fingerprints"
#PaulSimonDeathMetalSongs 🔗

🔁 0   ♡ 0   📤

ID 976537732888645632

View on Twitter 🔗

Nate Neutrality Willis 🇺🇦 @n8willis · Mar 21, 2018
"50 Ways to Cleave Your Mother"
#PaulSimonDeathMetalSongs 🔗

🔁 0   ♡ 1   📤

ID 976536958464913408

View on Twitter 🔗

Nate Neutrality Willis 🇺🇦 @n8willis · Mar 21, 2018
"Me and Julio Drowned by the Skullyard"
#PaulSimonDeathMetalSongs 🔗
#WorldPoetryDay 🔗

🔁 0   ♡ 0   📤

ID 976536264475336707

Search filters

Clear filters

Show

All Tweets ⦿

Media only ○

Text only ○

Sort by date

Newest first ⦿

Oldest first ○

Dates

From
📅 2024-07-17

To
📅 2024-07-17

paulsimondeathmetal

The case study: social network

Twitter's default archive HTML *might* suffice for the Library option

But its links all go out to the live site....

Although single-site, in-browser JS search is a poor experience.
- What about all my hilarious Facebook comments

The case study: social network

A lot of utilities are there to be explored:
- https://dogsheep.github.io/
- https://github.com/FGRibreau/import-tweets-to-mastodon
- https://github.com/alkihis/twitter-archive-reader
- https://helgenberger.net/alice/tumelune
- https://github.com/dangoldin/twitter-archive-analysis
- https://github.com/Webklex/tbm
- https://github.com/selfawaresoup/twitter-tools
- https://github.com/mshea/Parse-Twitter-Archive

The case study: social network

More usable: reformat the tweet archive, fix the links

`https://github.com/timhutton/twitter-archive-parser`

- creates flat HTML pages; fixes links to relative
- fetches original media
- converts follow/follower lists (by ID) to full account links
- is *non*-destructive on the downloaded archive, yay!

Your twitter archive

RT @Juan_Kafka: We have scholarships for every class @TypeElectives ✨✨✨✨ The priority deadline is this Sunday, January 8. More details on o...

🐦 Thu Jan 05 21:43:13 +0000 2023

RT @UniRdg_OpenRes: Overcome name confusion by claiming all YOUR research outputs with the unique and free ORCID iD! Take a look at our Lib...

🐦 Mon Jan 16 11:01:33 +0000 2023

Wellllllllllllllllllllll that's fun.

Matplotlib and SVG don't agree with each other about how color gradients work, so when you export an SVG _from_ Matplotlib, it says the opposite of what you mean for it to.

#ThisIsWhyYouCantHaveNiceData
#DontBlameMeWhenYourScienceIsBackwards

The case study: social network

`https://github.com/timhutton/twitter-archive-parser`
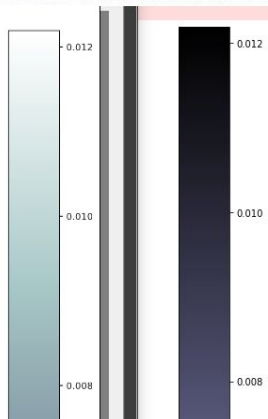
- auto-splits tweets into whole-month-at-a-time files
- much plainer look
- No Twitter-owned JS
- does *not* solve the half-swan problem
- but also creates Markdown

The case study: social network

Migrate to Mastodon? A dead end

- bulk import is not supported
- one third-party utility:
  `github.com/FGRibreau/import-tweets-to-mastodon`
- server admins generally oppose this:
  `github.com/mastodon/mastodon/issues/12423`
- manual database inserts:
  `github.com/mastodon/mastodon/issues/10745`

The case study: social network

Import

Import type *

You can import data that you have exported from another server, such as a list of the people you are following or blocking.

Following list

Data *

CSV file exported from another Mastodon server

Browse…   No file selected.

◉ Merge
    Keep existing records and add new ones

◯ Overwrite
    Replace current records with the new ones

Upload

The case study: social network

Migrate to a personal Mastodon: Overhead

This matters, because if FOSS is only available to people who are already Unix server sysadmins, or who have the spare time to become one, it's got no growth in its future.

But an awkward fit for historical data anyway....

The case study: social network

Repurpose: host tweets on your own site

```
https://indieweb.org/POSSE
```
- Post (on your) Own Site, Syndicate Elsewhere

The case study: social network

Repurpose: host tweets on your own site

`https://indieweb.org/POSSE`
- Post (on your) Own Site, Syndicate Elsewhere

...but in reverse. `FUGITIVE`

    *acronym still in development phase

The case study: social network

As autonomous as it can get: your posts, your site

- Pelican conversion of the converted Markdown files

- Remap: twitter.com/n8willis/status/1234567890
  --> twitter.freesoftwhere.org/1234567890

The case study: social network

🐍 I'm archiving all of my tweets on my own domain at
twitter.freesoftwhere.org

#FUGITIVE #Snowmanning #IceCastling

... all gonna be huge trends

The case study: social network

> @nationalrailenq westbound Elizabeth Line to LHRt5 is stuck waiting at Ealing Brwy\. Can you tell us what the status/outlook is please? 10\+ min and counting

 [Fri Apr 05 05:31:43 +0000 2024](https://twitter.com/n8willis/status/1776120510725685265)

> RT @OpenMFProject: If you're going to be at @texaslinuxfest in \#Austin this weekend \- don't miss the @OpenMFProject \#Linux Distribution \#Wo…

 [Wed Apr 10 01:47:32 +0000 2024](https://twitter.com/n8willis/status/1777876032206582196)

> RT @texaslinuxfest: A huge thank you goes out to @awscloud for being a Gold Sponsor of \#TXLF 2024\!
>
> Join us at Texas Linux Fest THIS WEEK o…

 [Thu Apr 11 01:36:15 +0000 2024](https://twitter.com/n8willis/status/1778235583329890718)

The case study: social network

```
Title: Tweeted at Thu Apr 11 01:36:15 +0000 2024
Date: Thu Apr 11 01:36:15 +0000 2024
Modified: Thu Apr 11 01:36:15 +0000 2024
Category: Twitter
Tags: Retweet
Slug: 1778235583329890718
Authors: Nathan Willis
Summary: RT @texaslinuxfest: A huge thank you goes out to @awscloud for being a Gold Sponsor of \#TXLF
2024\! Join us at Texas Linux Fest THIS WEEK o...

RT @texaslinuxfest: A huge thank you goes out to @awscloud for being a Gold Sponsor of \#TXLF 2024\! Join
us at Texas Linux Fest THIS WEEK o...
```

Tagged as : **Retweet**

Tweeted at Sat Sep 02 08:48:12 +0000 2023

Sat 02 September 2023

By Nathan Willis

Replying to @qole I assume that 10 mo out of the year most wrecks in Canada end with both vehicles being flung into snowdrifts. Softer impact. Of course that means more St Bernards per capita are required for freeing the passengers, but the stats don't appear to have measured that.

Read more ➡

Tagged as : **Reply**

Tweeted at Thu Aug 24 11:07:27 +0000 2023

Thu 24 August 2023

By Nathan Willis

RT @KyivPost: Happy Independence Day🇺🇦

Read more ➡

Tagged as : **Retweet**

Tweeted at Tue Aug 22 09:43:29 +0000 2023

Tue 22 August 2023

By Nathan Willis

Thanks for nuthin', "Translated by". Now I'll never know how to fit in. #failuremodes

Read more ➡

The case study: social network

```
Title: Tweeted at Thu Apr 11 01:36:15 +0000 2024
Date: Thu Apr 11 01:36:15 +0000 2024
Modified: Thu Apr 11 01:36:15 +0000 2024
Category: Twitter
Tags: Retweet
Slug: 1778235583329890718
Authors: Nathan Willis
Summary: RT @texaslinuxfest: A huge thank you goes out to @awscloud for being a Gold Sponsor of \#TXLF
2024\! Join us at Texas Linux Fest THIS WEEK o...

RT @texaslinuxfest: A huge thank you goes out to @awscloud for being a Gold Sponsor of \#TXLF 2024\! Join
us at Texas Linux Fest THIS WEEK o...
```

The case study: social network

Repurposing has its limits
- Thousands (potentially) of files
- Retweets / boosts
- Replies
- Permission?

But this is still better than deletionism....

The case study: social network

Twitter export data to consider:
1. The tweets ✓
2. Direct messages
3. The other accounts (follows/followers)
4. Likes / bookmarks

(Our options: Library, Import, Migrate, Repurpose)

The case study: half-pivot to link management

Twitter export data to consider:
1. The tweets ✓
2. Direct messages
3. The other accounts (follows/followers)
4. **Likes / bookmarks**

(Our options: Library, Import, Migrate, Repurpose)

The half-pivot: link management

A lot of online services have some type of "saved", "favorited", "starred", "followed", "watched", "subscribed", "bookmarked", or "liked" links....

What would we do with those?

(Our options: Library, Import, Migrate, Repurpose)

The half-pivot: link management

A lot of online services have some type of "saved", "favorited", "starred", "followed", "watched", "subscribed", "bookmarked", or "liked" links....

What would we do with those?

(Our options: Library, Import, Migrate, Repurpose)

The half-pivot: link management

Importing link collections is non-trivial
- Formats vary: CSV, Netscape-Bookmark HTML, XBEL, OPML, JSON, etc.
- Formats often lack metadata
- Exports often provide only internal URIs, requiring API use?
- Bookmark managers rarely optimize management

The half-pivot: link management

Applications I test drove:
- Zotero
- Buku
- LinkAce
- Linkding
- Servas

List (406)   Create   Add Filter ▾   10 items ▾

Entry

👁 ✏ 🗑 Reddit saved post g286xs
https://www.reddit.com/r/selfhosted/comments/g286xs/moving_away_from_google_voice/

netloc:www.reddit.com   reddit

👁 ✏ 🗑 Reddit saved post 11s8dms
https://www.reddit.com/r/selfhosted/comments/11s8dms/guys_please_recommend_some_interesting_docker/

netloc:www.reddit.com

👁 ✏ 🗑 Reddit saved post f2d8at
https://www.reddit.com/r/cyberpunk2020/comments/f2d8at/looking_for_a_copy_of_cyberpunk_update_1_the/

netloc:www.reddit.com

👁 ✏ 🗑 Reddit saved post yjvpp6
https://www.reddit.com/r/selfhosted/comments/yjvpp6/deleted_by_user/

netloc:www.reddit.com

👁 ✏ 🗑 Reddit saved post 10mh51o
https://www.reddit.com/r/selfhosted/comments/10mh51o/selfhosted_captcha_solution/

netloc:www.reddit.com

Bookmarks

Search for words or #tags

Tags

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon liked post

#mastodon

Today | View Edit Archive Remove

Mastodon

Reddit

The case study: half-pivot to link management

Twitter export data to consider:
1. The tweets ✓
2. **Direct messages**
3. **The other accounts (follows/followers)**
4. Likes / bookmarks **?**

(Our options: Library, Import, Migrate, Repurpose)

The case study: half-pivot to accounts/contacts

In theory, follows/followers are contacts: a clear case for the Import option.

In practice, most data exports remove private information and many try to be unhelpful. Manual cleanup is required.

Service-specific info is not standard VCARD fare.

The case study: half-pivot to private messages

There is no universal "private message" format to combine exported messages into....

Fun story: I briefly attempted to force messages into a format for use with `notmuch`. Maildir does not want to be used for this.

Gap analysis

What are the takeaways?

Gap analysis

1. Free-software services do a poor job providing useful exports

Gap analysis

1. Free-software services do a poor job providing useful exports

2. FOSS applications are not designed to work with import data sets

Gap analysis

1. Free-software services do a poor job providing useful exports

2. FOSS applications are not designed to work with import data sets

3. We have only piecemeal tools for migrating and repurposing data exports

What do we do to improve this for users?

1. Talk about it: try repurposing your Twitter/etc account and write it up. Tag me if you like!

What do we do to improve this for users?

1. Talk about it: try repurposing your Twitter/etc account and write it up. Tag me if you like!

2. Fix a small bug in one of the piecemeal tools

What do we do to improve this for users?

1. Talk about it: try repurposing your Twitter/etc account and write it up. Tag me if you like!

2. Fix a small bug in one of the piecemeal tools

3. Help me figure out where to "centralize" some information or discussions.... A wiki? An "`awesome-`" list?

What do we do to improve this for users?

1. Talk about it: try repurposing your Twitter/etc account and write it up. Tag me if you like!

2. Fix a small bug in one of the piecemeal tools

3. Help me figure out where to "centralize" some information or discussions.... A wiki? An "`awesome-`" list?

4. Request some data exports and see what happens

Thank you for listening!

BANGERS — Vernon Adams

Michroma — Vernon Adams

Gentium Basic — Victor Gaultney

nate@glyphography.com
@n8@mastodon.social
@n8willis