

Digitale Belege sichern

Internet Archive nutzen
und selbst anlegen

Wie festhalten, was wir im Internet gefunden haben?

Screenshot?

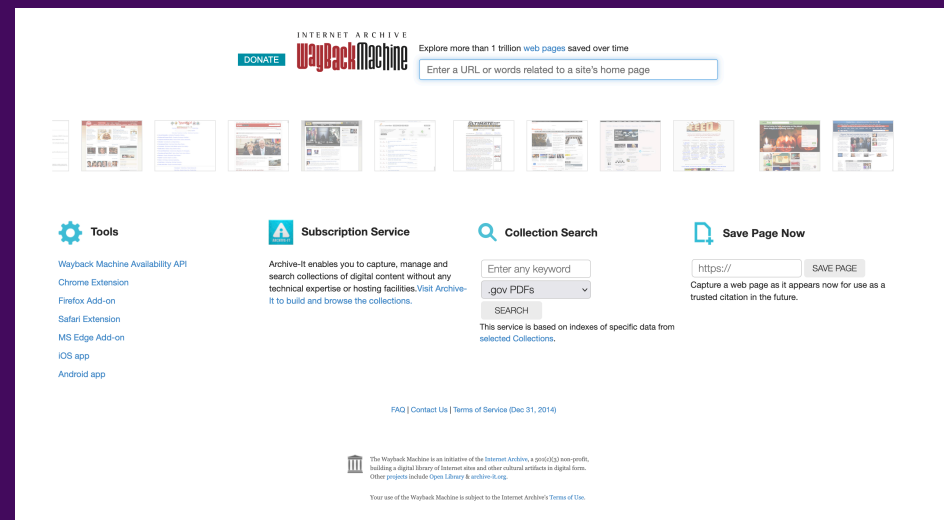
Ein Screenshot fast so schnell gefälscht wie er gemacht wurde.

Welche Internet Archive gibt es?

- Ziemlich viele: Wikipedia "List of web archiving initiatives" ist über 100 Zeilen lang: https://en.wikipedia.org/wiki/List_of_web_archiving_initiatives
- Die meisten gehören zu öffentlichen Bibliotheken/Archiven
- Viele sind "*Dark Archives*": Daten gespeichert, aber nicht öffentlich zugänglich
 - Speicherung ist Vervielfältigung - Urheberrecht
 - DSGVO, Persönlichkeitsrechte

**Die zwei
bekanntesten
öffentlich-
zugänglichen Archive**

Internet Archive (IA)



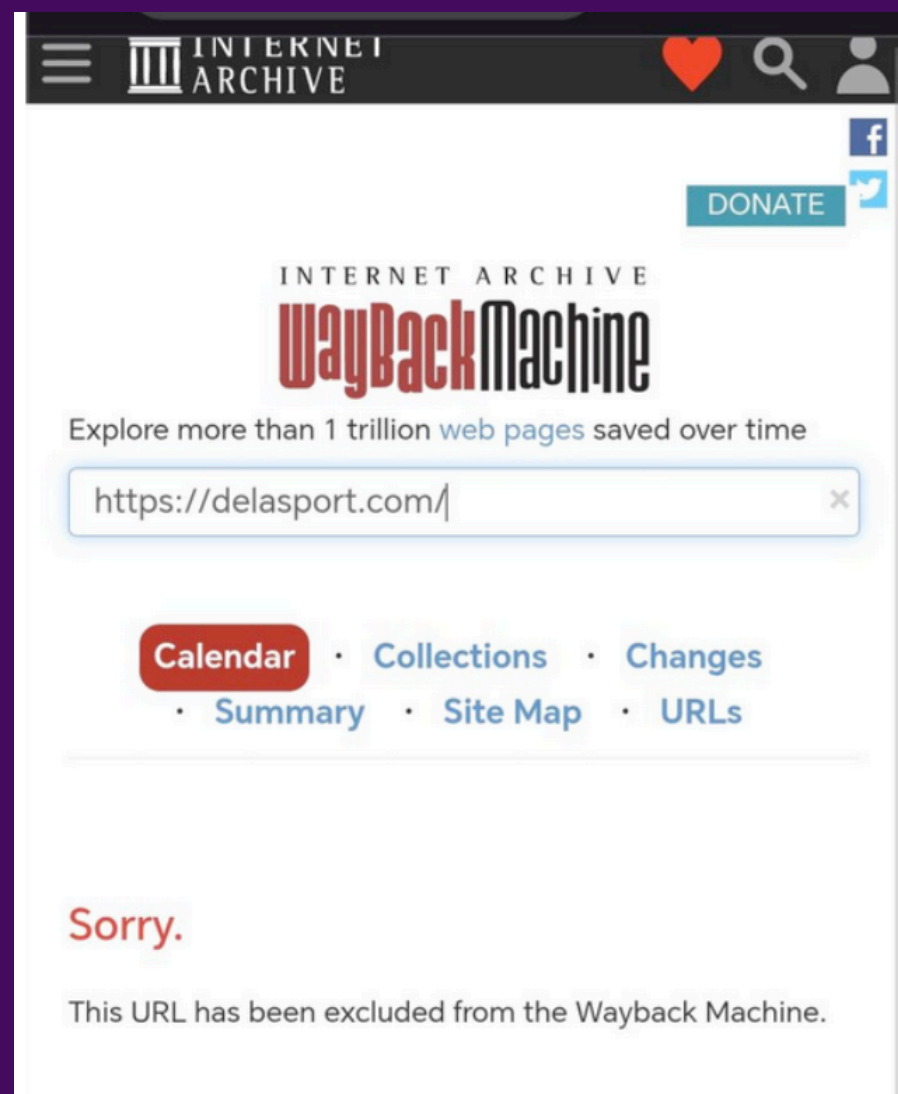
- NGO mit Sitz in Kalifornien
- mit Abstand größtes, öffentlich zugängliches digitales Archiv: Oktober 2025 eine Trillion Webseiten, plus andere digitale Medien
- Sammlung entsteht durch Crawler und einzelnen Speicherungen, Wikipedia-Prinzip: alle können - und sollen - mitmachen
- Sammlung auch durch APIs zugänglich
- Kommerzieller Arm: archive-it, Kunden sind z.B. die EU oder der BR

IA und AI

- NYTimes, Guardian und andere Medien sperren IA-Crawler aus - wegen KI
- IA selbst: "Wir sind Kollateralschaden"
- Offene Briefe unter savethearchive.com

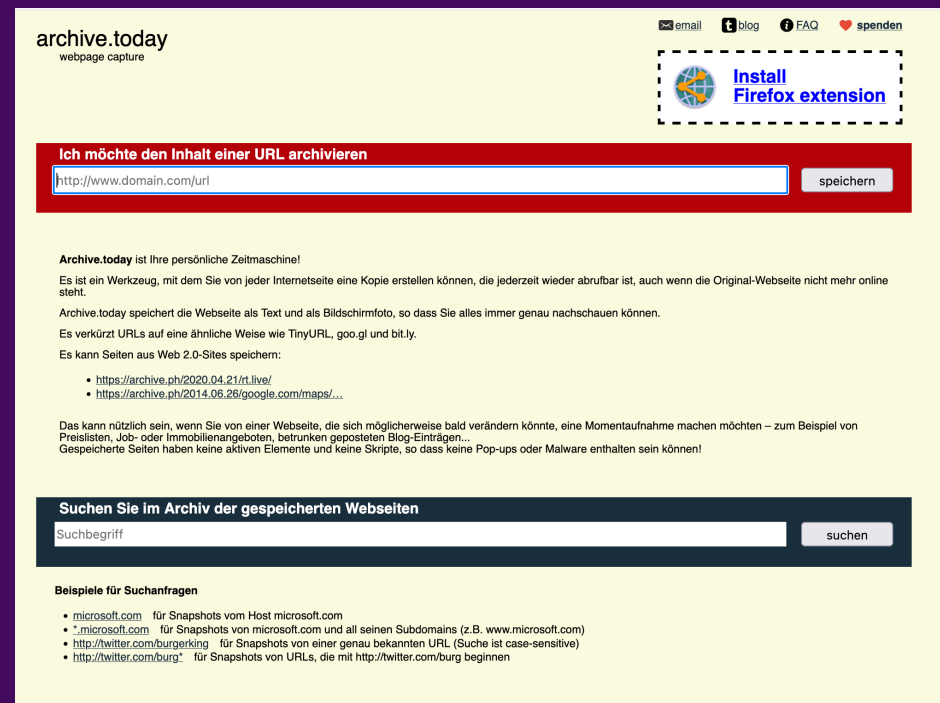


Nicht auf die Wayback Machine verlassen



- Wichtig zu wissen: Man kann sich aus der Wayback Machine entfernen lassen ([FAQ dazu](#))
- Beispiel: Delasport, IT-Dienstleister für Online-Casinos

archive.today



- kann sehr gut Paywalls umgehen
 - Unbekannter und fragwürdiger Betreiber, Spuren nach Russland
 - Vorwürfe:
 - DDOS-Attacke auf Blog einer Person, die Hintergründe recherchierte
 - Hat Snapshots verändert
- Wikipedia hat fast 700 000 Links entfernt

Das Web zu archivieren ist technisch schwierig

- je komplexer Webseiten gebaut sind, desto schwieriger ist es auch, sie zu archivieren
- Je größer die Plattform (Social Media!), desto stärker die Abwehrmechanismen
- eine archivierte Webseite ist keine Kopie, sondern ein Puzzle, von dem möglichst viele Teile gespeichert und wieder zusammengesteckt werden - und das dauert ...
- Funktionen gehen verloren, z.B. die Suche

Kein Archiv kann alles

	ghostarchive	archive.vn	wayback machine	perma.cc
Bluesky	Red	Green	Green	Grey
Facebook	Red	Green	Green	Red
FB-Reel	Red	Red	Red	Grey
Insta-Post	Red	Red	Red	Red
Insta-Reel	Red	Red	Red	Red
LinkedIn	Red	Red	Red	Grey
Mastodon	Green	Green	Red	Green
Reddit	Green	Green	Green	Red
Telegram	Red	Green	Green	Green
TikTok	Red	Red	Red	Green
X	Red	Green	Red	Green
X (Video)	Red	Red	Red	Grey
YT-Short	Red	Red	Red	Red
YT-Vid	Red	Red	Red	Red

Erstellt von Michael Schlegl, BR-Faktenfuchs, Februar 2026

Jetzt beginnt der Praxisteil

Mit Fokus auf das
Internet Archive

Kurze Begriffsklärung

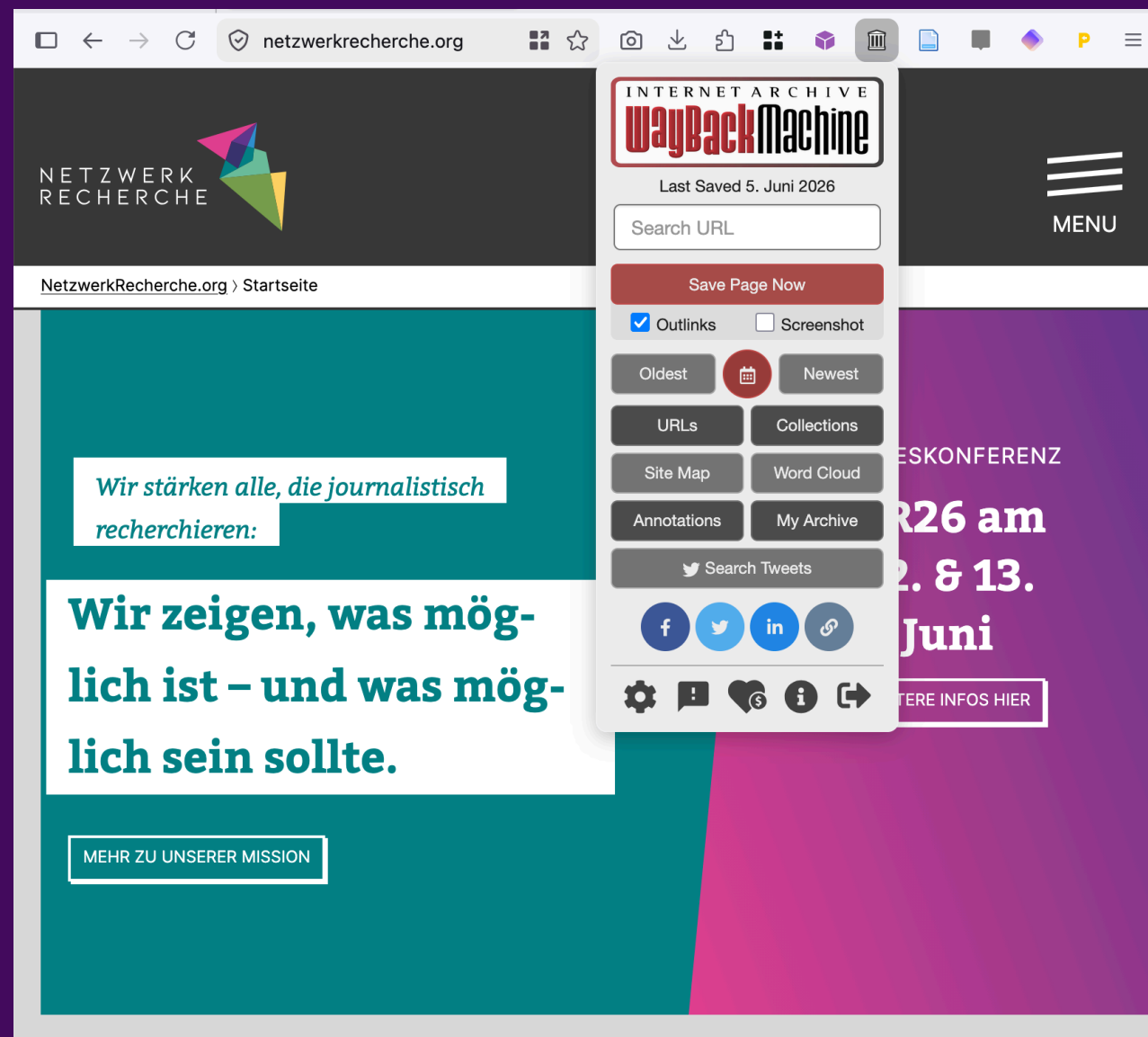
- *Snapshot/Capture*: archivierte Version einer Webseite
- *WARC/WACZ*-Datei: Standard, um Web-Inhalte zu archivieren
- *Command Line/Kommandozeile/CLI*: dieses schwarze Fenster, in das man Befehle eingibt und dann passiert was. Scheu ablegen ermöglicht sehr, sehr viel

Eine Seite finden

- über die Startseiten der Archive wie web.archive.org
- Es gibt mehrere Browser-Erweiterungen, um direkt eine URL in einem oder mehreren Archiven zu finden
- Beispiele:
 - "PastPage" von [Alexander Nabert](#) ([Firefox](#), [Chrome](#))
 - "Web Archives", [Github](#)
 - "Wayback Machine" des Internet Archives ([Firefox](#), [Chrome](#))
 - gibt aus als App für iOS und Android

Eine Seite speichern

"Wayback Machine" Browser-Erweiterung des Internet Archives



Viele Seiten speichern

Beste Variante imo, um viele URLs ins Internet Archive zu schieben:

[overcast07/wayback-machine-spn-scripts](https://overcast07.github.io/wayback-machine-spn-scripts/)

- Save Page Now API mit [Authentication](#)
- parallelisiert
- Logs
- Filtern von Outlinks
- ...

Liste von URLs

urls.txt:

```
https://netzwerkecherche.org  
https://www.ndr.de
```

Befehl

```
./spn.sh -a accesskey:secret -f ./logs ./urls.txt
```

```
== Starting spn.sh ==
```

```
Using existing data folder: ./logs
```

```
2026-06-05 13:46:56 [Job submitted] https://netzwerkecherche.org
```

```
    The same snapshot had been made 38 minutes ago. You can make new capture of this URL after 1 hour.
```

```
2026-06-05 13:46:57 [Job submitted] https://www.ndr.de
```

```
█
```

Archival OSINT

Zielgenau Infos im

Internet Archive finden

Alte Preislisten finden

- Wir kennen die URLs von Firmen
- Wir wissen, dass Preislisten immer als PDFs veröffentlicht wurden
- Aber wir kennen die konkreten URLs nicht und die können auch unterschiedlich sein.

Lösung:

- Internet Archive CDX API pro Domain untersuchen und dann alle pdf's filtern
- Checken, ob alte Preisliste; meist schon durch URL erkennbar

CDX-API-Abfrage:

```
curl -o urls.json "https://web.archive.org/cdx/search/cdx?url=DOMAIN.de*&output=json"
```

waybackurls: durchsucht Internet Archive, Common Crawl, Virus Total

urls.txt:

```
br.de  
netzwerkecherche.org  
auswaertiges-amt.de
```

```
cat urls.txt | waybackurls > wayback_urls.txt
```

Anderes Beispiel aus dem readwrite-Newsletter:

[Fetching websites from the Wayback Machine](#)

Veränderungen finden

- Wann wurde etwas auf einer Webseite hinzugefügt oder entfernt?
- Viel händische Arbeit, warten, klicken, suchen, neu suchen... 🤯
- Wann hat die Bundesregierung Reise- und Sicherheitshinweise für die USA ergänzt um extra Aussagen für trans Personen?

[Auswärtiges Amt: USA/Vereinigte Staaten: Reise- und Sicherheitshinweise](#)

Werkzeug memento-cli: Verbindet Snapshots mit Volltextsuche für spezielle URL

```
memento bisect --text "Geschlechtseintrag" \  
https://web.archive.org/web/20241211111436/https://www.auswaertiges-amt.de/de/service/  
laender/usa-node/usavereinigtestaatensicherheit-201382
```

Ergebnis:

Found your archive snapshot:

```
https://web.archive.org/web/20250311184444/  
https://www.auswaertiges-amt.de/de/service/laender/usa-node/usavereinigtestaatensicherheit-201382
```

[link](#)

Eigene nicht- öffentliches Archive

- Zotero
- Webrecorder/Browsertrix
- Social Media

Zotero

- Literaturverwaltungssoftware, die Snapshots von Webseiten mit SingleFile erstellen kann
- Speichern mit Browser-Erweiterung, suchen und finden in Anwendung
- Nutzbar im Team
- Kommt bei Social Media an die Grenzen

The screenshot shows a Zotero record for a webpage. The record is titled "USA/Vereinigte Staaten: Reise- und Sicherheitshinweise". The metadata includes the following information:

- Zitierschlüssel:** amtUSAvereinigteStaaten
- Eintragsart:** Webseite
- Titel:** USA/Vereinigte Staaten: Reise- und Sicherheitshinweise
- Autor:** Amt, Auswärtiges
- Titel der Website:** Auswärtiges Amt
- Art der Webseite:** (empty)
- Datum:** (empty)
- Verlag:** (empty)
- Ort:** (empty)
- DOI:** (empty)
- URL:** <https://www.auswaertiges-amt.de/de/service/laender/usa-node/usaverinigtestaatensicherheit-201382>
- Heruntergeladen am:** 11.6.2026, 12:40:32
- Kurztitel:** USA/Vereinigte Staaten
- Sprache:** de
- Lizenz:** (empty)
- Extra:** (empty)
- Hinzugefügt am:** 11.6.2026, 12:40:32
- Geändert am:** 11.6.2026, 12:40:38

Below the metadata, there is a section for "Zusammenfassung" (Summary) with the text: "Letzte Änderungen: Aktuelles – Fußballweltmeisterschaft 2026 Redaktionelle Änderungen".

There is also a section for "Anhang" (Attachment) with 1 attachment. The attachment is a PDF file titled "Auswärtiges Amt" with a language selector set to "DE".

The bottom part of the screenshot shows a preview of the webpage content, including a navigation menu with the following items: "Auswärtiges Amt", "Außen- und Europapolitik", "Sicher Reisen", "INHALT", "Karriere", and "News".

Webrecorder

- Erst aufnehmen, dann abspielen (record, replay)
- Browser-Erweiterung (Chromium-basiert), um Webseiten dynamisch zu speichern
- Ergebnis ist WARC/WACZ-Datei
- Kann auch Social Media (Accounts, Logins sichtbar)
- Skalieren mit browsertrix

Archiving: 4 URLs pending, please wait before loading a new page. ? 🏠

Currently archiving to: ■ Stop Archiving

[View Archived Pages](#)

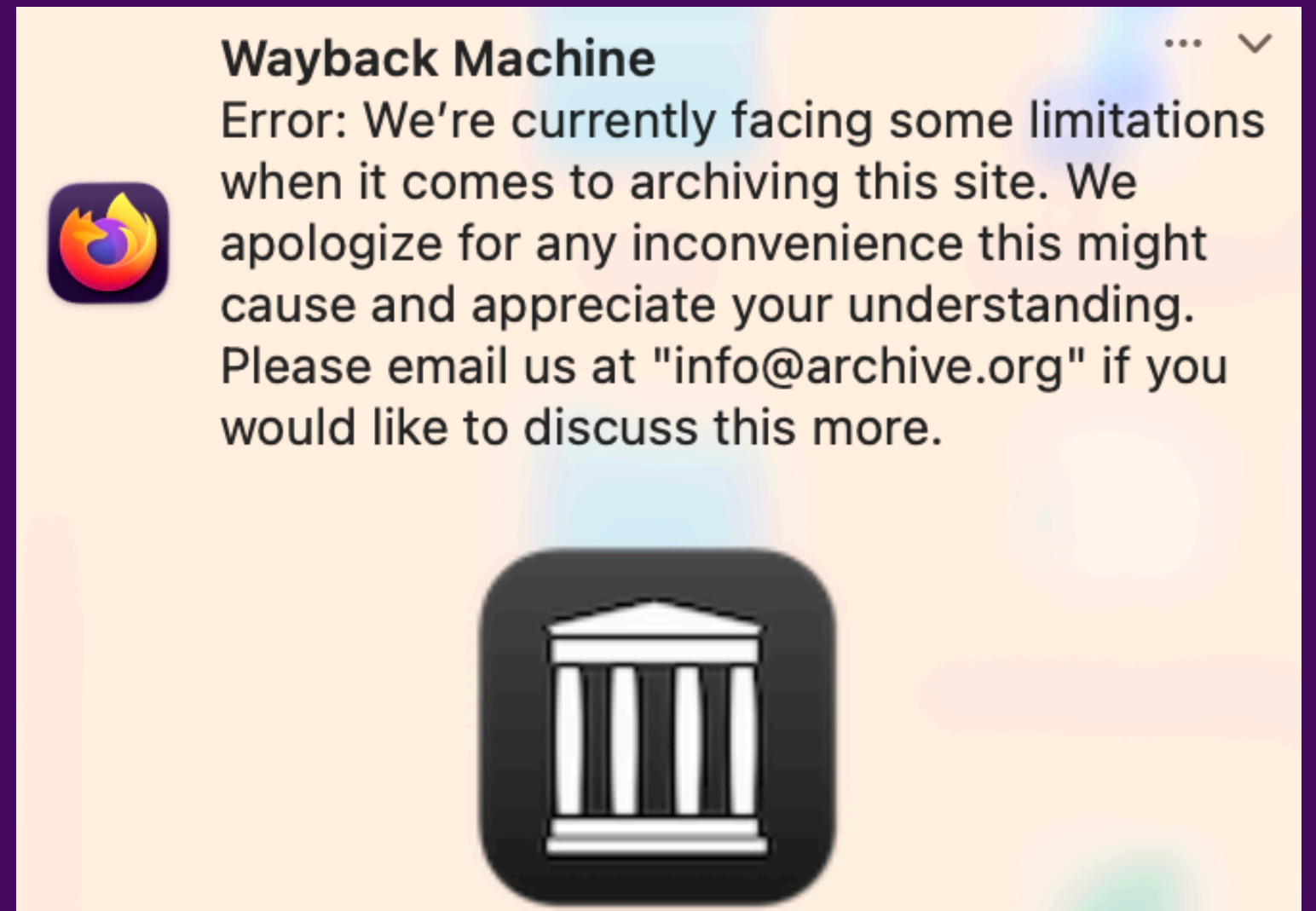
Autopilot: Waiting for page to load...

Archived in this tab [Replay Current Page](#)

SIZE STORED:	22.3 MB
SIZE LOADED:	23 MB
PAGES:	1
URLS:	127

Social Media

- Anstatt die Webseite zu speichern, über interne APIs auf Inhalte und Metadaten zugreifen
- entweder lokal oder v.a. Daten über Dienstleister, z.B. apify oder rapidapi



Beispiel Instagram

- Instaloader als Werkzeug:
 - Holt Bilder und Videos direkt über interne APIs
 - immer mit Wegwerf-Accounts: Man wird regelmäßig gesperrt
 - man bekommt nicht nur die Bilder und Videos, sondern auch Kommentare und viele weitere Metadaten, die man auswerten kann
 - Alle Captions
 - Alle Accounts, die kommentiert haben
 - Alle getaggten Accounts
 - ...

Beispiel Youtube

- öffentliches Archiv: [preservetube.com](https://www.preservetube.com)
- [yt-dlp](#):
 - Videos
 - Transkripte

und außerdem . . .

- Bellingcat Auto Archiver:
 - fortgeschrittenes Werkzeug, um Webseiten, Social Media zu speichern
 - gewisser Setup-Aufwand
- evidx: Suchen Redaktionen, die testen wollen
 - "Capture, cryptographically sign, and preserve web content — creating tamper-evident archives that hold up under scrutiny. On your infrastructure, under your control."
 - gerade "Prototype Fund"-Projekt zu Zeitstempeln ohne Blockchain

Links

- Eurovision News Spotlight Newsletter: [Advanced Wayback Machine and archival OSINT](#)
- [Awesome Web Archiving](#)
- [GLAM Workbench](#): Beschreibung und Code
- Shafayat Ahmed Alif - Linked: [The Ultimate Wayback Machine Cheat Sheet for OSINT, Cybersecurity, and Archival Research](#)

Mein Tipp: Tendenziell eher an academia halten als an OSINT/Hacker Social Media

Mehr Anleitungen zu digitalen Archiven

katharinabrunner.de (Abonnieren mit RSS )

mastodon.social/@cutterkom

Ex-AfD-Mitarbeiter bekommt Schadenersatz

Kontext verliert auch beim BGH

Die Wochenzeitung „Kontext“ veröffentlichte Chats, die ihr ein Hacker zuspielte. Ihre Echtheit konnte sie aber nicht beweisen.

3.6.2026 20:02 Uhr

 teilen

[Hintergründe in der taz](#)

Belegen, dass der Inhalt genau so online war

- **Minimum:** zusätzlich Zeitstempel
- **Maximum:** Hash ("Fingerabdruck") einer Datei speichern und Existenz der Hashes zu einem bestimmten Zeitpunkt dokumentieren (zum Beispiel in der Blockchain)
-