



# CDMM Tutorial

Zeeschuimer

Tutorial by  
Asger Harlung  
CDMM, 2025

**CDMM**

Centre for Digital Methods and Media  
<https://cc.au.dk/cdmm/>

# Zeeschuimer Tutorial

Zeeschuimer is a tool for moderate data harvesting from a number social media services, including Instagram and TikTok.

This tutorial shows how to install and use it, with examples from Instagram and TikTok.

This tutorial can also be downloaded as PDF.

## **For samples of videos or visuals:**

The SingleFile browser add-on can preserve Instagram and Tiktok pages, but may not preserve long pages well.

The instagram page used in this tutorial could not be preserved, but all TikTok pages could

<https://cc.au.dk/en/cdmm/tools-and-tutorials/data-collection/singlefile>

Videos from TikTok posts can be preserved by copying their direct links to JDownloader2

<https://cc.au.dk/en/cdmm/tools-and-tutorials/data-collection/jdownloader-2>

## TikTok Warning

**WARNING**: TikTok demands access to large amounts of user data, and is not considered safe.

It may not be used on Aarhus University's devices.

See:

<https://medarbejdere.au.dk/en/administration/it/guides/security/tiktok-faq>

Similar rules may apply at other institutions or workplaces. Examples in this tutorial were thus created on a private computer.

# General Warning

## GENERAL WARNING:

Harvest of data from social media should appear as a human user – not too fast, and not too much.

If systematic harvesting is registered, the account used for the process may be blocked.

To harvest, e.g. several accounts, harvest only one or a few per day, and do not scroll or look around faster than  
a  
human user might do.

# Zeeschuimer Tutorial

Zeeschuimer is a plugin for the browser Mozilla Firefox from Digital Methods Initiative.

Firefox is thus a prerequisite.

Please visit CDMM's Mozilla Firefox page for the version of your choice:

<https://cc.au.dk/en/cdmm/tools-and-tutorials/data-collection/mozilla-firefox>

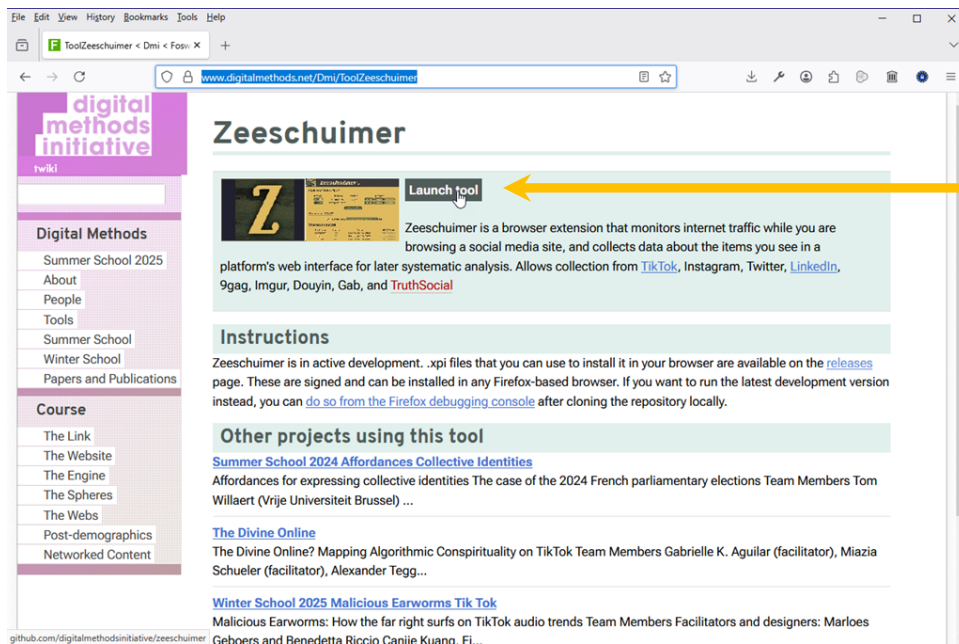
If your browser "remembers" a previous session this can prevent new posts from being registered correctly in a new harvest. It is advised to use a dedicated research browser, and clearing the cache and browser history before reharvesting pages previously visited.

Firefox Developer is a good suggestion for this purpose (see previous slide for CDMM's Mozilla Firefox page).

# Setting up Zeeschuimer

In Firefox, visit the developer page for background

<https://www.digitalmethods.net/Dmi/ToolZeeschuimer>



**“Launch tool”  
will send you to  
the GitHub page  
where the tool  
can be installed**

# Setting up Zeeschuimer

Zeeschuimer  
GitHub page:

The screenshot displays the GitHub repository page for `digitalmethodsinitiative/zeeschuimer`. The repository is public and has 269 stars, 10 watchers, and 23 forks. The main content area shows a file tree with the following items:

File/Folder	Description	Last Commit
fonts	Readying things for the first non-beta release	3 years ago
images	Extra installation instructions	4 months ago
inc	LinkedIn support (beta)	3 years ago
js	Don't run disabled modules on same domain as enabled	4 months ago
modules	Update threads domain	3 months ago
popup	Bump version	3 months ago
tests	Update tests	5 months ago
.gitignore	Ignore .DS_Store	last year
.zenodo.json	Bump version	3 months ago
LICENSE	Bump version	3 years ago

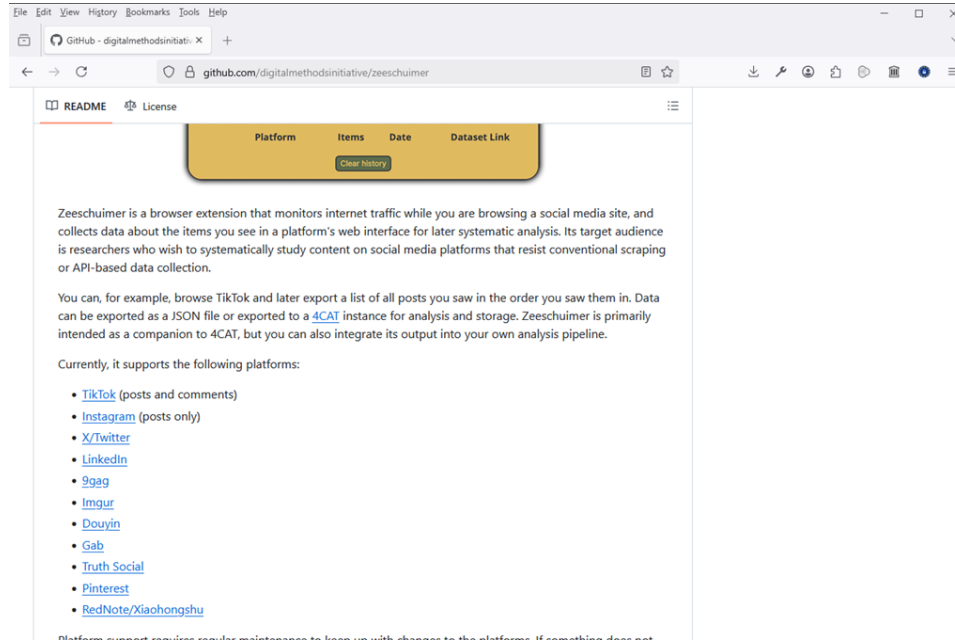
The right sidebar contains the following information:

- About:** A browser extension to collect social media data with.
- Readme:** View license, Activity, Custom properties.
- Stars:** 269 stars, 10 watching, 23 forks.
- Releases:** 35 releases. The latest release is `v1.13.1 threads.net - threads.com` on Apr 25.

It is worth  
scrolling down  
this page for  
information

# Setting up Zeeschuimer

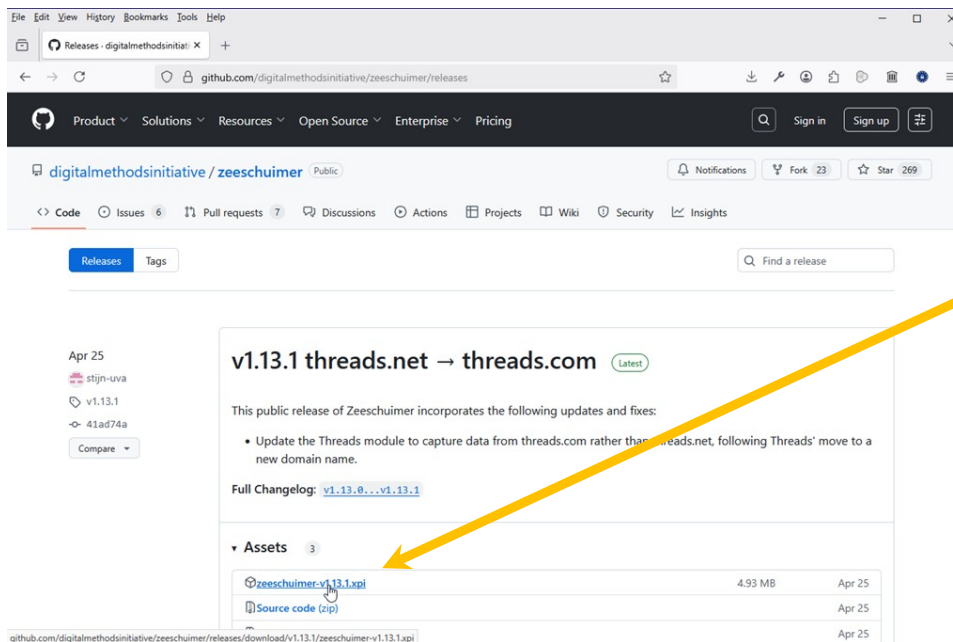
**Zeeschuimer  
GitHub page:**



**On this page you  
will find  
information on  
how it works, and  
of social software  
presently  
supported**

# Setting up Zeeschuimer

The Zeeschuimer GitHub page directs you to the "releases" section. Direct address: <https://github.com/digitalmethodsinitiative/zeeschuimer/releases>

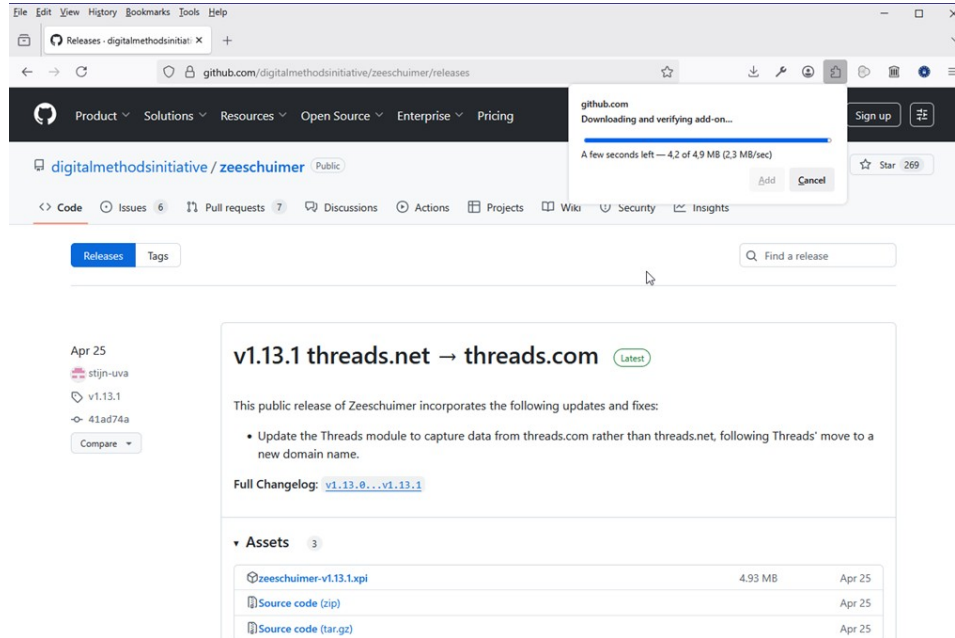


The screenshot shows the GitHub releases page for the repository `digitalmethodsinitiative/zeeschuimer`. The release `v1.13.1` is highlighted as the latest version. The release notes indicate an update to the Threads module. The 'Assets' section contains three items: `zeeschuimer-v1.13.1.xpi` (4.93 MB), `Source code (zip)`, and another asset. A yellow arrow points to the `zeeschuimer-v1.13.1.xpi` link.

Asset Name	Size	Created
<a href="#">zeeschuimer-v1.13.1.xpi</a>	4.93 MB	Apr 25
<a href="#">Source code (zip)</a>		Apr 25
		Apr 25

Click on the link for the XPI file for the latest version at the top of the list

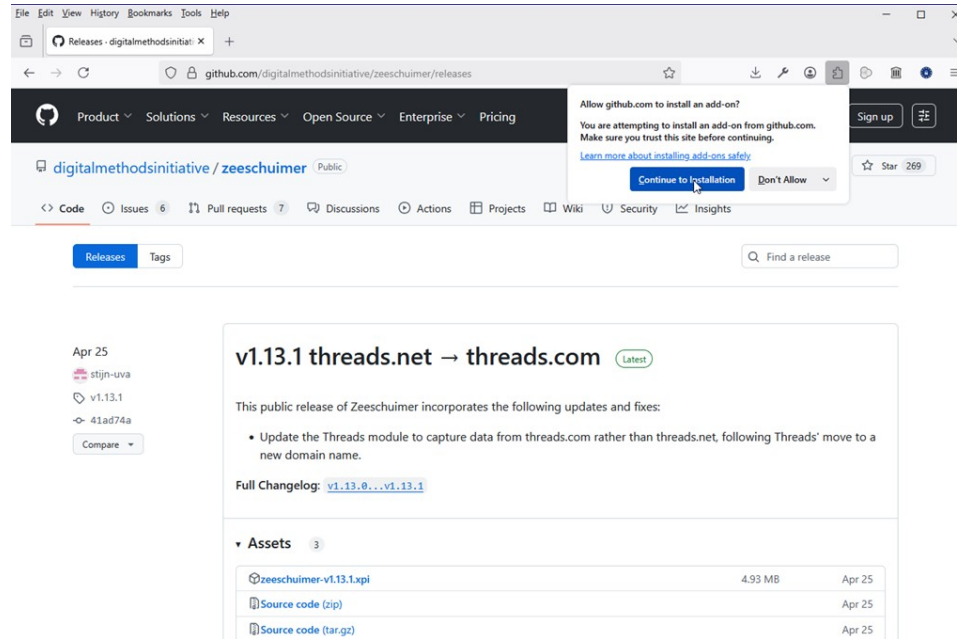
# Setting up Zeeschuimer



The screenshot shows a web browser displaying the GitHub release page for the repository `digitalmethodsinitiative/zeeschuimer`. The page is titled "v1.13.1 threads.net → threads.com" and is marked as the "Latest" release. The release description states that this public release incorporates updates and fixes, including updating the Threads module to capture data from threads.com instead of threads.net. The full changelog is provided as `v1.13.0...v1.13.1`. Under the "Assets" section, three files are listed: `zeeschuimer-v1.13.1.xpi` (4.93 MB), `Source code (zip)`, and `Source code (tar.gz)`, all dated April 25. A download progress bar overlay is visible in the upper right, showing the file is being downloaded and verified at 2.3 MB/sec, with approximately 4.2 MB remaining.

The file starts downloading

# Setting up Zeeschuimer



The screenshot shows a web browser window displaying the GitHub releases page for the repository `digitalmethodsinitiative/zeeschuimer`. The browser's address bar shows the URL `github.com/digitalmethodsinitiative/zeeschuimer/releases`. A notification from Firefox is overlaid on the page, asking to allow `github.com` to install an add-on. The notification text reads: "Allow github.com to install an add-on? You are attempting to install an add-on from github.com. Make sure you trust this site before continuing." Below the notification are two buttons: "Continue to Installation" (highlighted with a mouse cursor) and "Don't Allow".

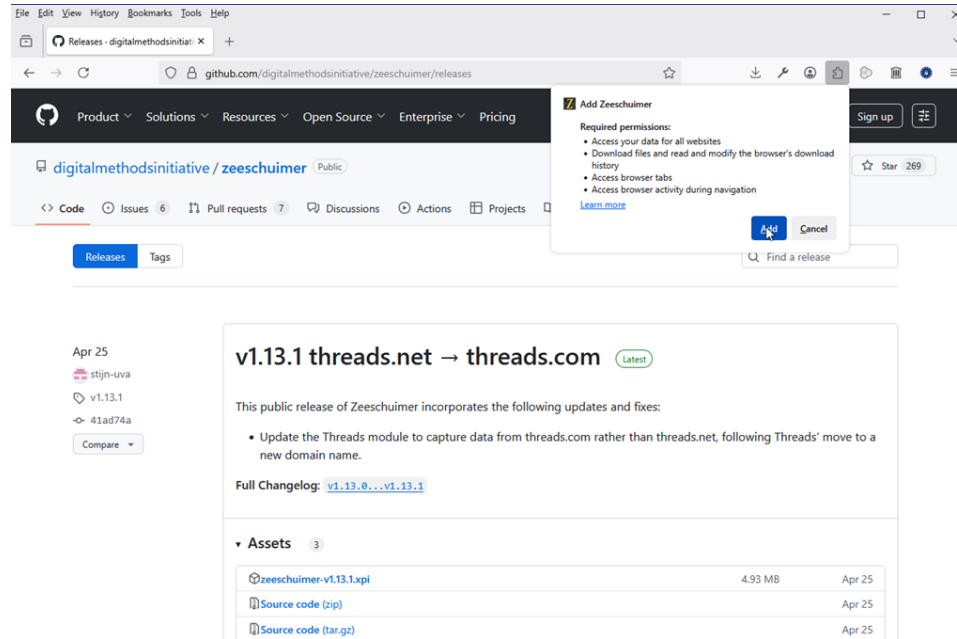
The main content of the page shows the release `v1.13.1 threads.net → threads.com` (Latest) by `stijn-uva`. The release description states: "This public release of Zeeschuimer incorporates the following updates and fixes:" followed by a bullet point: "Update the Threads module to capture data from threads.com rather than threads.net, following Threads' move to a new domain name." The full changelog is `v1.13.0...v1.13.1`.

The Assets section lists three files:

Asset	Size	Created
<a href="#">zeeschuimer-v1.13.1.xpi</a>	4.93 MB	Apr 25
<a href="#">Source code (zip)</a>		Apr 25
<a href="#">Source code (tar.gz)</a>		Apr 25

**When the download is complete, installation of the add-on to Firefox becomes available. Click "Continue to Installation"**

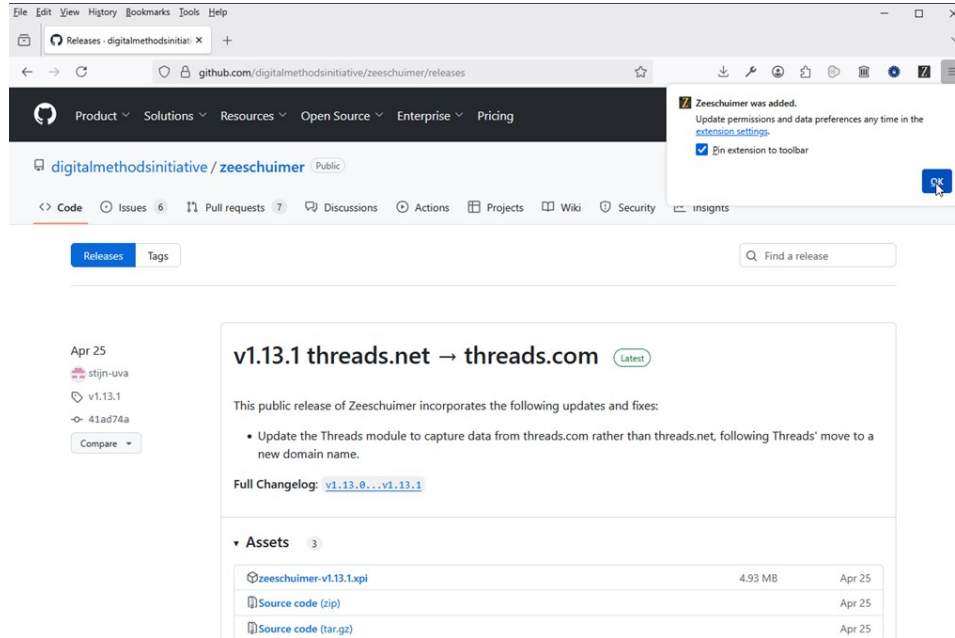
# Setting up Zeeschuimer



The screenshot shows the GitHub releases page for the 'zeeschuimer' repository. The release v1.13.1 is highlighted, with a 'Latest' badge. The release description states: 'This public release of Zeeschuimer incorporates the following updates and fixes: Update the Threads module to capture data from threads.com rather than threads.net, following Threads' move to a new domain name.' The full changelog is linked as v1.13.0...v1.13.1. Under the 'Assets' section, three files are listed: zeeschuimer-v1.13.1.xpi (4.93 MB, Apr 25), Source code (zip) (Apr 25), and Source code (tar.gz) (Apr 25). An 'Add' button is visible in the top right corner of the release card, with a tooltip showing the required permissions: 'Add Zeeschuimer', 'Required permissions:', 'Access your data for all websites', 'Download files and read and modify the browser's download history', 'Access browser tabs', and 'Access browser activity during navigation'. A search bar for releases is also present.

**Click on "Add" to confirm the required access for the add-on**

# Setting up Zeeschuimer



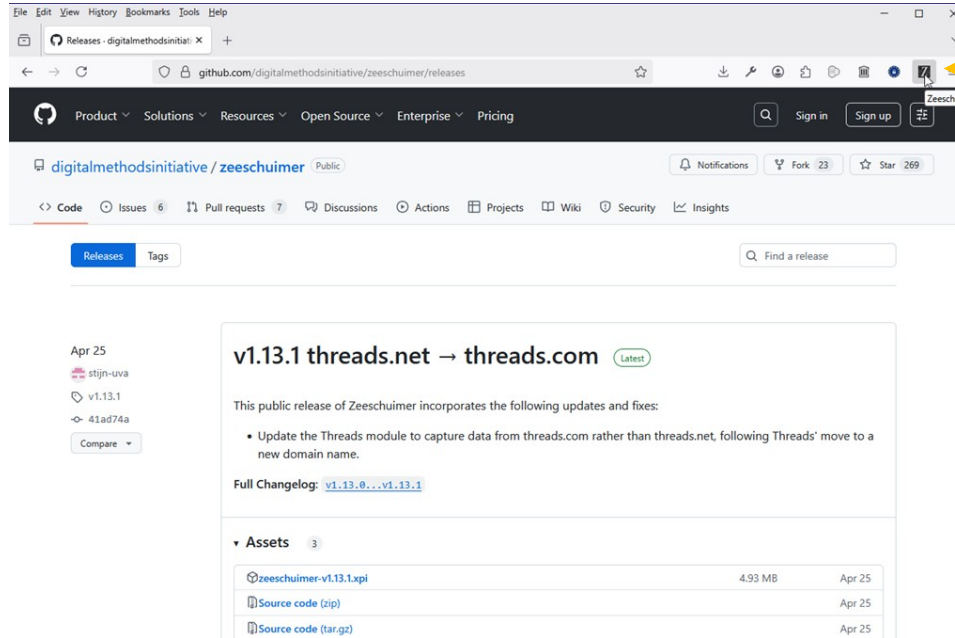
The screenshot shows the GitHub releases page for the repository `digitalmethodsinitiative/zeeschuimer`. The page displays the release `v1.13.1` with the title `v1.13.1 threads.net → threads.com` and a `Latest` badge. The release description states: "This public release of Zeeschuimer incorporates the following updates and fixes:" followed by a bullet point: "Update the Threads module to capture data from threads.com rather than threads.net, following Threads' move to a new domain name." Below the description is a link to the `Full Changelog: v1.13.0...v1.13.1`. The `Assets` section lists three files:

Asset	Size	Date
<a href="#">zeeschuimer-v1.13.1.xpi</a>	4.93 MB	Apr 25
<a href="#">Source code (zip)</a>		Apr 25
<a href="#">Source code (tar.gz)</a>		Apr 25

A notification box in the top right corner of the browser window states: "Zeeschuimer was added. Update permissions and data preferences any time in the [extension settings](#)." Below this message is a checkbox labeled "Pin extension to toolbar" which is checked, and a blue "OK" button.

It is advised to confirm "Pin extension to toolbar" so the tool may be activated easily

# Setting up Zeeschuimer



The screenshot shows the GitHub releases page for the repository `digitalmethodsinitiative/zeeschuimer`. The page displays the latest release, `v1.13.1 threads.net → threads.com`, dated April 25. The release description states that this public release incorporates updates and fixes, including an update to the Threads module to capture data from `threads.com` instead of `threads.net`. The full changelog is provided as `v1.13.0...v1.13.1`. Under the 'Assets' section, three files are listed: `zeeschuimer-v1.13.1.xpi` (4.93 MB), `Source code (zip)`, and `Source code (tar.gz)`, all dated April 25.

**Zeeschuimer is now added to Firefox and ready for use**

## Adding FoxScroller

An additional tool may prove helpful:

FoxScroller is a browser add-on which can scroll down on most pages automatically.

It cannot handle all types of pages or simulate human interaction such as opening replies to posts, but it can add a level of automation to your workflow. It is not a necessity, but it is used in the examples in this tutorial.

## Adding FoxScroller

In Firefox, go go to “Extensions and Themes” from the application menu (≡) in the top right corner, and search for “FoxScroller”

Or go directly to the extension’s page:

<https://addons.mozilla.org/en-US/firefox/addon/foxscroller/>

# Adding FoxScroller

Thousands of users and mostly positive reviews are good signs. This add-on can be safely installed despite the "not actively monitored" note

Firefox Browser  
**ADD-ONS** Extensions Themes More...  
Find add-ons

Firefox Add-ons Blog Extension Workshop Developer Hub Log in

FS  
**FoxScroller**  
by Simalleus

⚠ This add-on is not actively monitored for security by Mozilla. Make sure you trust it before installing.  
[Learn more](#)

Auto scroll web pages with fully adjustable speed. [Add to Firefox](#)

5,468 Users 87 Reviews 4.6 Stars

5 stars	47
4 stars	15
3 stars	2
2 stars	2
1 star	1

**Rate your experience**  
How are you enjoying FoxScroller?

**About this extension**  
This addon gives you a button and hotkeys to continuously scroll down a website. Multiple websites can be scrolling at the same time.  
Note: Does not work on "addons.mozilla.org".

Click on "Add to Firefox"

# Adding FoxScroller

The screenshot shows the Firefox Add-ons Manager interface. The main content area displays the 'FoxScroller' extension by Simalleus. A warning message states: 'This add-on is not actively monitored for security by Mozilla. Make sure you trust it before installing.' Below this, there is a 'Learn more' button and a description: 'Auto scroll web pages with fully adjustable speed.' A star rating section shows a 4.6-star average with a breakdown: 5 stars (47), 4 stars (15), 3 stars (2), 2 stars (2), and 1 star (1). An 'Add' dialog box is overlaid on the page, showing the extension's name, required permissions (Access your data for all websites, Access browser activity during navigation), and optional settings (Allow extension to run in private windows, which is checked). The dialog has 'Add' and 'Cancel' buttons. A mouse cursor is pointing at the 'Add' button.

Firefox Browser  
**ADD-ONS** Extensions Themes More... ▾

**FoxScroller**  
by Simalleus

⚠ This add-on is not actively monitored for security by Mozilla. Make sure you trust it before installing.  
[Learn more](#)

Auto scroll web pages with fully adjustable speed.

**Rate your experience**

How are you enjoying FoxScroller?  
[Log in to rate this extension](#)

**About this extension**

This addon gives you a button and hotkeys to continuously scroll down a website. Multiple websites can be scrolling at the same time.  
Note: Does not work on "addons.mozilla.org".

**Add FoxScroller**

Required permissions:

- Access your data for all websites
- Access browser activity during navigation

Optional settings:

- Allow extension to run in private windows

[Learn more](#)

**Add** Cancel

4.6 Stars

5 ★	47
4 ★	15
3 ★	2
2 ★	2
1 ★	1

Click on "Add"

# Adding FoxScroller

Firefox Browser  
**ADD-ONS** Extensions Themes More... ▾

Firefox Add-ons Blog

**FoxScroller**  
by Simalleus

⚠ This add-on is not actively monitored for security by Mozilla. Make sure you trust it before installing.  
[Learn more](#)

Auto scroll web pages with fully adjustable speed. [Remove](#)

**Rate your experience**

How are you enjoying FoxScroller?  
[Log in to rate this extension](#)

**About this extension**

This addon gives you a button and hotkeys to continuously scroll down a website. Multiple websites can be scrolling at the same time.  
Note: Does not work on "addons.mozilla.org".

**FoxScroller was added.**  
Update permissions and data preferences any time in the [extension settings](#).  
 Pin extension to toolbar

5,468 Users 87 Reviews 4.6 Stars

Star Rating	Count
5 Stars	47
4 Stars	15
3 Stars	2
2 Stars	2
1 Star	1

Once again it is advised to confirm "pin extension to toolbar"

# Adding FoxScroller

The screenshot shows the Firefox Add-ons page for the FoxScroller extension. The page includes the following information:

- Extension Name:** FoxScroller
- Developer:** Simalleus
- Users:** 5,468
- Reviews:** 87
- Star Rating:** 4.6 Stars
- Review Breakdown:**

Star Rating	Count
5 Stars	47
4 Stars	15
3 Stars	2
2 Stars	2
1 Star	1
- Warning:** This add-on is not actively monitored for security by Mozilla. Make sure you trust it before installing.
- Remove Button:** A button labeled "Remove" is visible.
- Rate your experience:** A section for users to rate their experience with the extension.
- About this extension:** A section describing the extension's functionality: "This addon gives you a button and hotkeys to continuously scroll down a website. Multiple websites can be scrolling at the same time. Note: Does not work on 'addons.mozilla.org'."

**FoxScroller is now added and ready for use**

## Harvesting Instagram Data

With Zeeschuimer and FoxScroller ready, it is time for a walkthrough of the process of an Instagram harvest.

The procedure applies to all harvesting processes on various social media, but there may be variations, especially in the demand for hands-on user interaction. This will be addressed in the TikTok example which follows the first one on Instagram

# Harvesting Instagram Data

The screenshot shows a web browser displaying the Instagram profile for 'denmarkincalifornia'. The browser's address bar shows the URL 'www.instagram.com/denmarkincalifornia/'. The profile header includes the name 'denmarkincalifornia', a 'Follow' button, and a 'Message' button. Below the header, the profile statistics are listed: 481 posts, 1,443 followers, and 597 following. The bio reads: 'Denmark in California', 'Consulate & Embassy', 'We're Denmark in California 🇩🇰🇺🇸 We show you the best of the West Coast - and we're here to help startups and... more', and a link to 'usa.um.dk/en/about-us/internships and 1 more'. A row of seven category icons is displayed: Startups, Daily Life, Bioinnovators, Fun Facts, Rambo, Climate Week, and Water Team. Below the categories, the 'POSTS' tab is selected, showing three image posts. The first post shows a person with blonde hair in a field of green plants. The second post shows a person on a rocky trail. The third post shows a person with blonde hair in a light blue shirt. The left sidebar of the Instagram app is visible, showing navigation options: Home, Search, Explore, Reels, Messages, Notifications, Create, Profile, More, and Also from Meta.

Go to an  
Instagram page  
that you want to  
harvest

# Harvesting Instagram Data

The screenshot shows the Instagram profile for 'denmarkincalifornia'. The browser's developer tools are open, and a red box highlights the 'Zeeschuimer Status' extension icon in the top right corner. A yellow arrow points from the text 'Click on the Zeeschuimer shortcut' to this icon.

Instagram

denmarkincalifornia [Follow](#) [Message](#) [...](#)

481 posts 1,443 followers 597 following

Denmark in California  
Consulate & Embassy  
We're Denmark in California 🇩🇰🇺🇸 We show you the best of the West Coast - and we're here to help startups and... more  
[usa.um.dk/en/about-us/internships](#) and 1 more

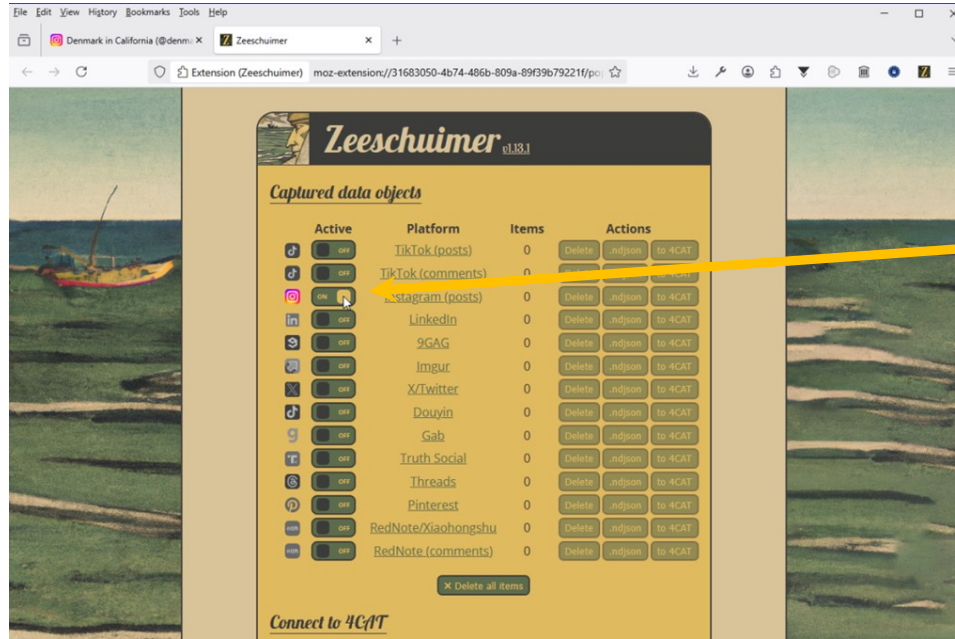
Startups Daily Life Bioinnovators Fun Facts Rambo Climate Week Water Team

POSTS REELS TAGGED

Three post thumbnails are visible at the bottom of the page.

Click on the  
Zeeschuimer  
shortcut

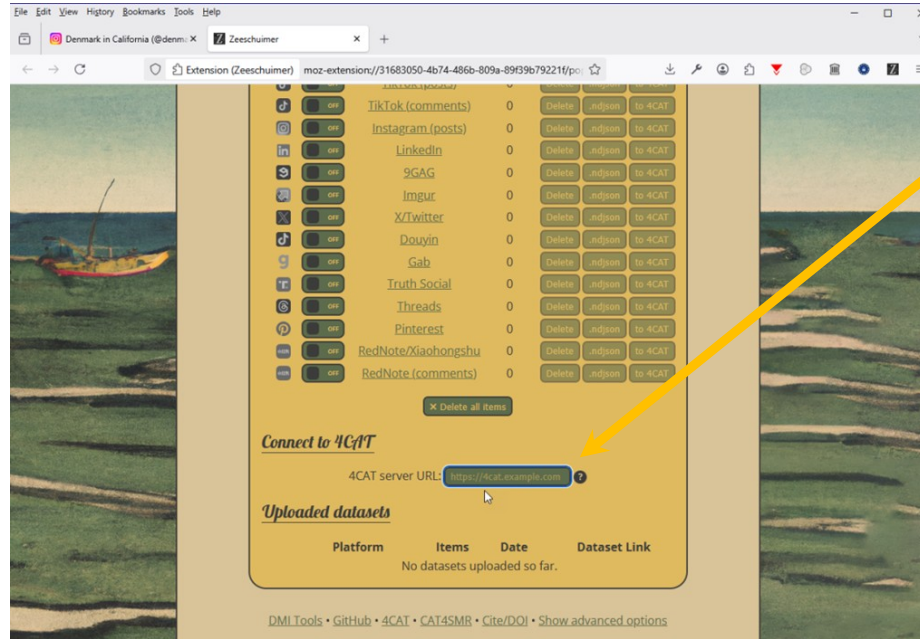
# Harvesting Instagram Data



**Zeeschuimer  
opens in a new  
tab**

**Activate post  
harvesting for  
Instagram**

# Harvesting Instagram Data



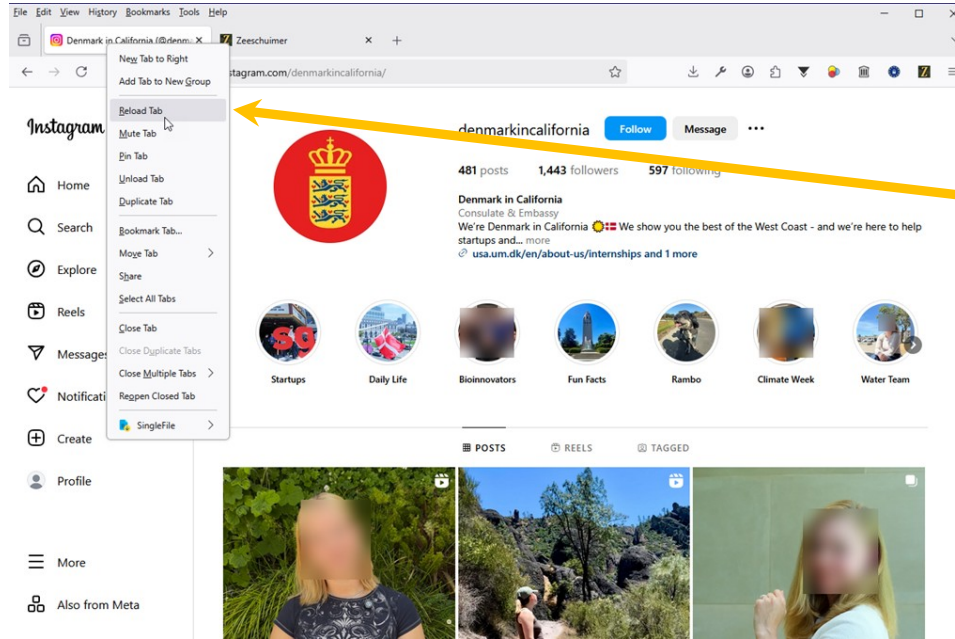
The screenshot shows the 4CAT web interface. At the top, there is a list of social media platforms with their respective icons, names, and item counts. Below this list is a 'Delete all items' button. The 'Connect to 4CAT' section is highlighted with a yellow arrow, showing a '4CAT server URL' input field containing 'https://4cat.example.com'. Below this is an 'Uploaded datasets' table with columns for Platform, Items, Date, and Dataset Link, which currently shows 'No datasets uploaded so far.' At the bottom, there are links for DMI.Tools, GitHub, 4CAT, CAT4SMR, Cite/DOI, and Show advanced options.

Platform	Items	Date	Dataset Link
No datasets uploaded so far.			

**Enter your 4CAT server address here if you have one and prefer using it**

**This tutorial continues with direct data downloads**

# Harvesting Instagram Data

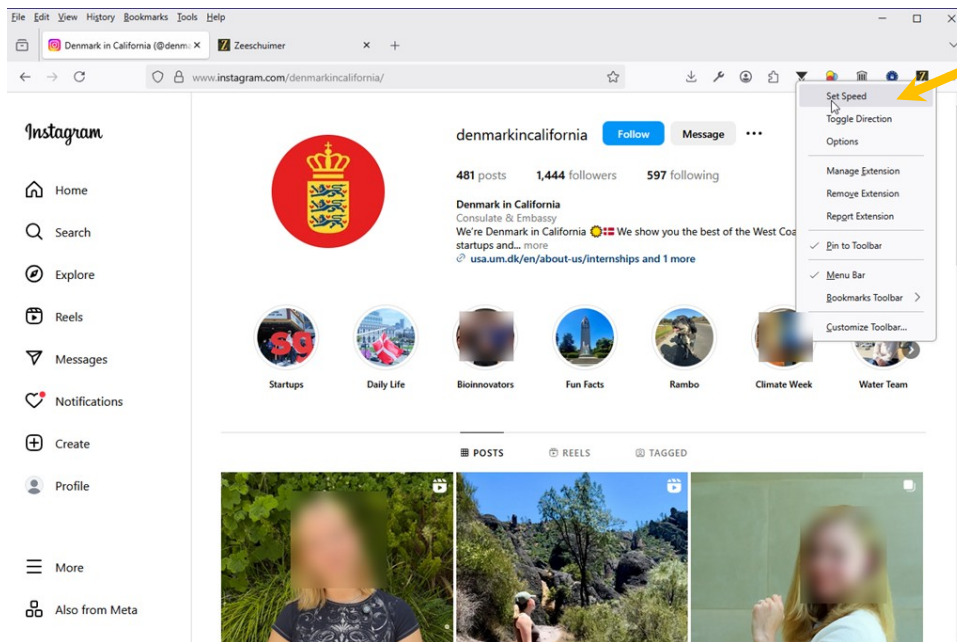


**Go back to the Instagram tab and reload the page**

**After activating Zeeschuimer this is necessary for starting the data capture**

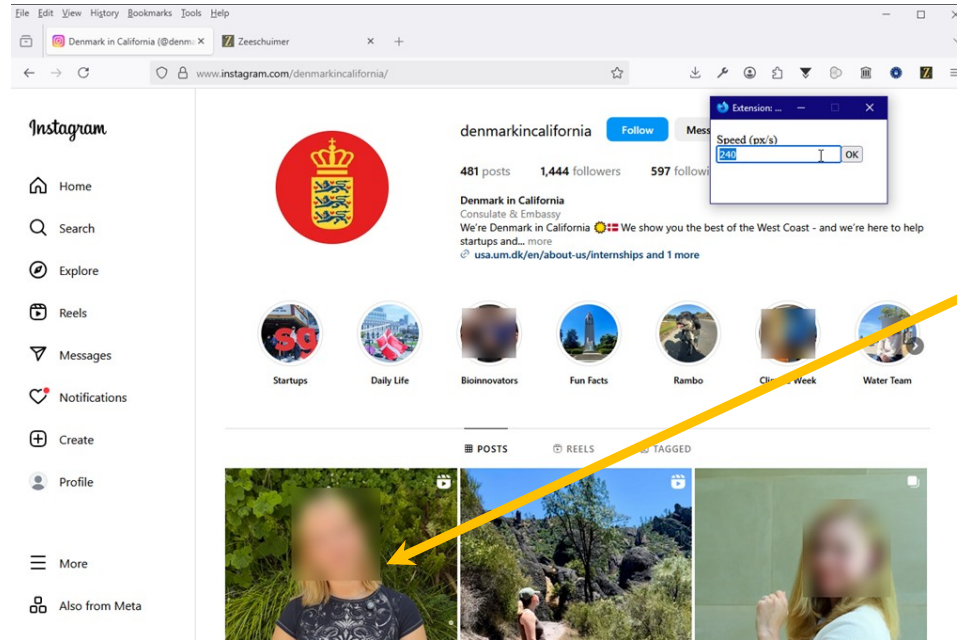
# Harvesting Instagram Data

**FoxScroller will now be used to automate the process. One can set the speed. It is measured in pixels, and will appear faster or slower depending on the screen resolution**



**Right-click (ctrl+click in Mac OS ) on the FoxScroller shortcut if you want to adjust the scrolling speed.**

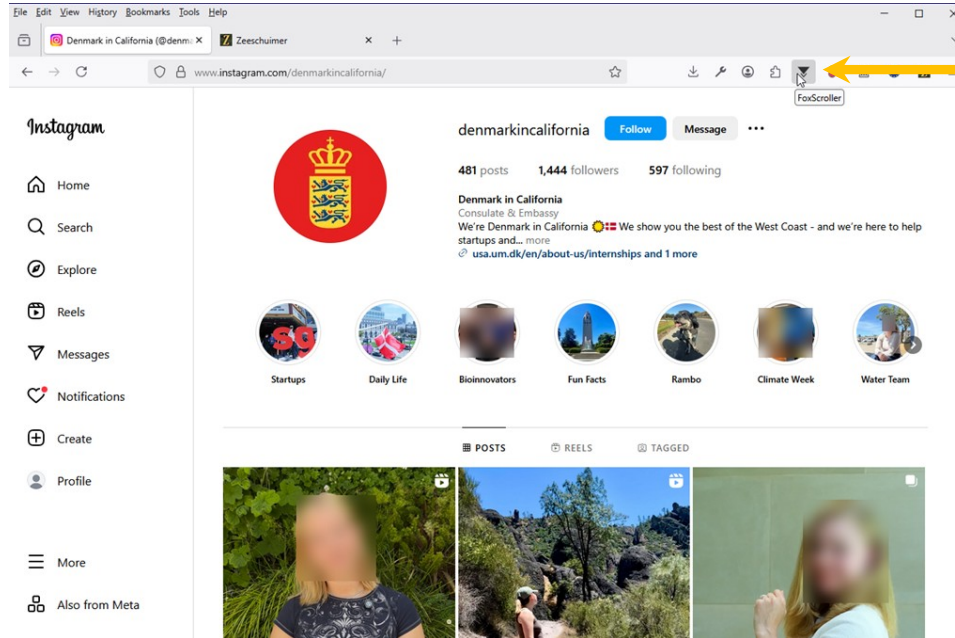
# Harvesting Instagram Data



The screenshot shows a web browser window displaying the Instagram profile for 'denmarkincalifornia'. The browser's address bar shows the URL 'www.instagram.com/denmarkincalifornia/'. The profile page includes the account name, a 'Follow' button, and statistics: 481 posts, 1,444 followers, and 597 following. The bio identifies the account as the 'Denmark in California Consulate & Embassy' and provides a link to 'usa.um.dk/en/about-us/internships and 1 more'. Below the bio are several category icons: Startups, Daily Life, Bioinnovators, Fun Facts, Rambo, Climate Week, and Water Team. A yellow arrow points from the text on the right to the first post in the grid, which shows a person's face blurred. An 'Extension...' dialog box is open, showing a 'Speed (px/s)' input field with the value '240' and an 'OK' button.

**A speed setting of 240 to 500 should be fine. Do not scroll to fast (see explanation on slide 5 in the introduction)**

# Harvesting Instagram Data

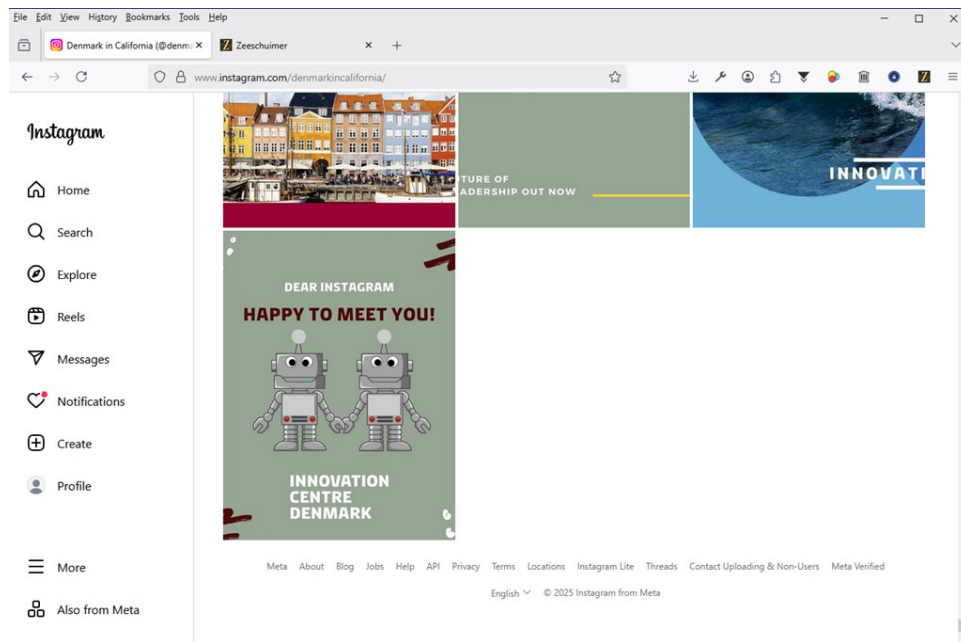


**Now click on the FoxScroller shortcut to start automatic scrolling**

# Harvesting Instagram Data

**FoxScroller will scroll down until it reaches the bottom of the page**

**On some pages user interaction is required to load more content, whereupon scrolling has to be restarted**



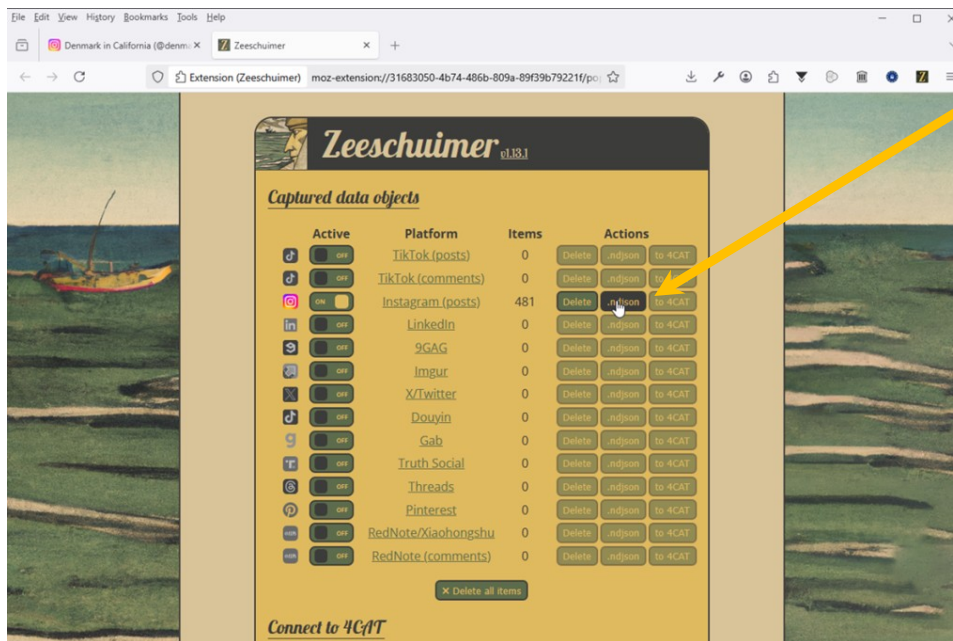
**For very large pages, data overload in the browser may occur.**

**See slide 58-60 on Data Overload for comments on this problem**

# Harvesting Instagram Data

Now go back to the Zeeschuimer tab

Under "Items",  
481 harvested posts are now harvested



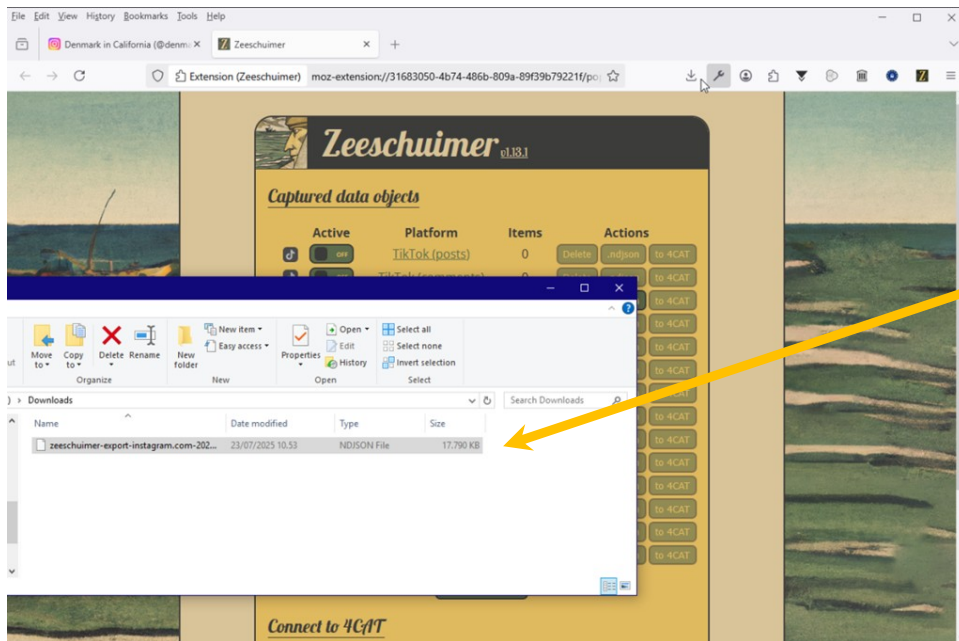
The screenshot shows the Zeeschuimer extension interface. At the top, it says "Zeeschuimer v1.13.1". Below that, it says "Captured data objects". There is a table with the following columns: Active, Platform, Items, and Actions. The table lists various social media platforms and their harvested item counts. A yellow arrow points to the ".ndjson" button in the Actions column for the "Instagram (posts)" row.

Active	Platform	Items	Actions
<input checked="" type="checkbox"/>	TikTok (posts)	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	TikTok (comments)	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Instagram (posts)	481	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	LinkedIn	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	9GAG	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Imgur	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	X/Twitter	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Douyin	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Gab	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Truth Social	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Threads	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	Pinterest	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	RedNote/Xiaohongshu	0	Delete .ndjson to 4CAT
<input checked="" type="checkbox"/>	RedNote (comments)	0	Delete .ndjson to 4CAT

At the bottom of the table, there is a button that says "Delete all items". Below the table, it says "Connect to 4CAT".

Click on the .ndjson button to download the data

# Harvesting Instagram data



**The data will download and can now be found in your Downloads folder**

**It may be a good idea to add cue words to the file name, e.g. name of the harvested page at the beginning**

As described on CDMMs tutorial page, JSON Extraction to CSV or Excel

<https://cc.au.dk/en/cdmm/tools-and-tutorials/data-cleaning/translate-to-english-json-extraction-to-csv-or-excel>,

JSON files can be difficult to work with, and difficult to transform to the CSV format that allows import to a spreadsheet

# Zeehaven

Fortunately, this service can transform the NDJSON files from Zeeschuimer for you:

<https://publicdatalab.github.io/zeehaven/>

The Zeehaven service was created by Public Data Lab

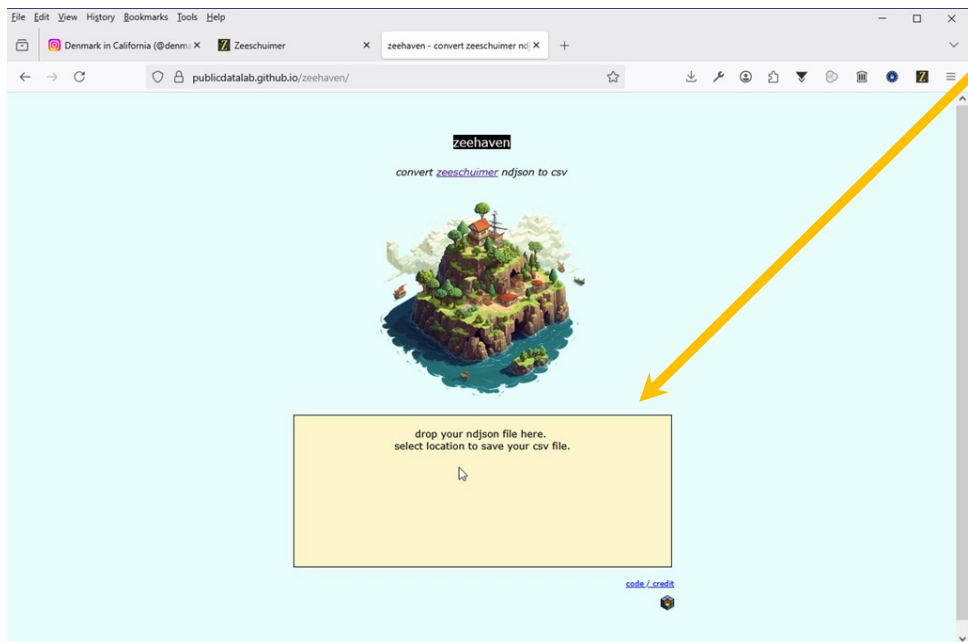
<https://publicdatalab.org/>,

an interdisciplinary network dedicated to collective inquiry with and about digital data, digital methods and digital infrastructures. Details here:

<https://publicdatalab.org/2023/12/18/zeehaven-social-media-data/>

# Converting the Data

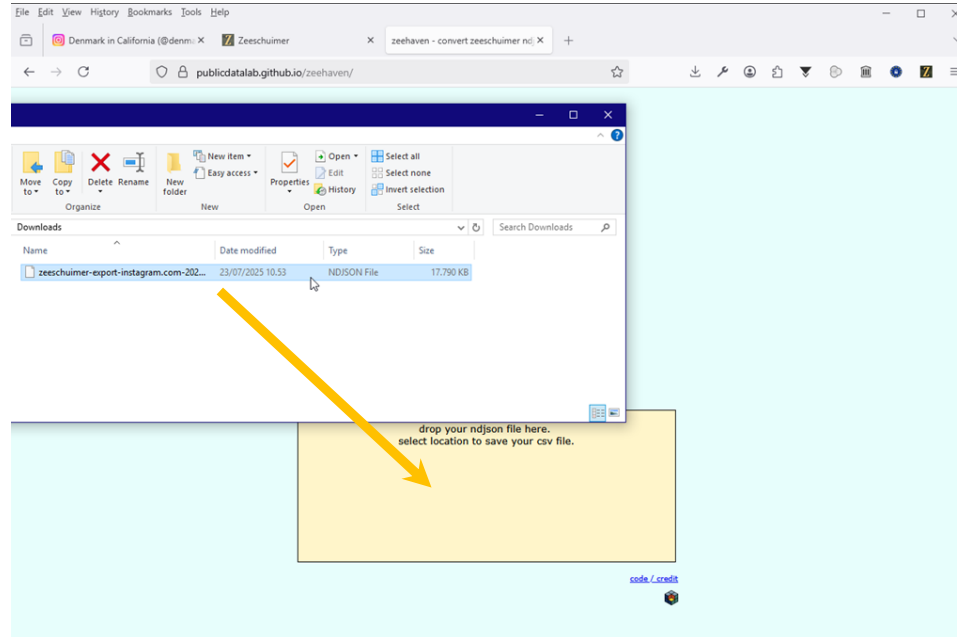
Go to the  
Zeehaven page at  
<https://publicdatalab.github.io/zeehaven/>



Notice the drag  
and drop field

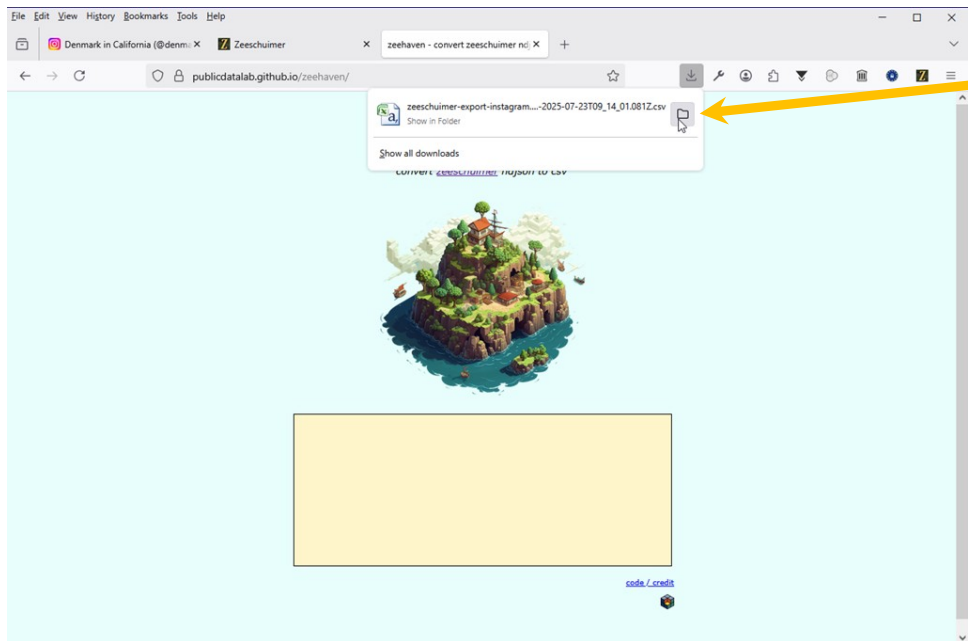
# Converting the Data

**Drag your  
NDJSON data file  
to the drag and  
drop conversion  
field**



# Converting the Data

The "drop your  
.ndjson file here"  
text disappears,



and a converted  
CSV file is  
downloaded.

It is now located  
in the Downloads  
folder with the  
NDJSON file

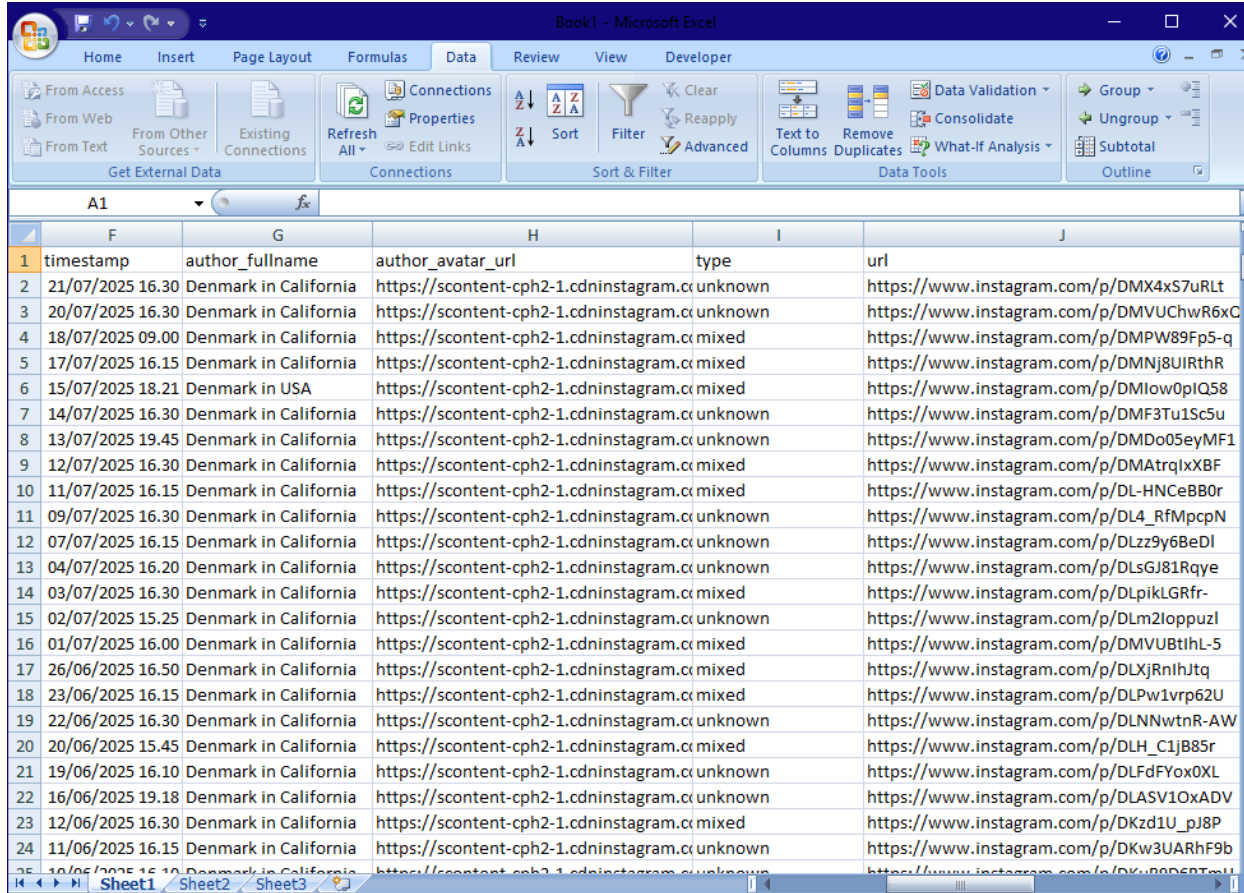
## Converting the Data

The last thing to do is to import the CSV file to a spreadsheet.

Please consult CDMM's page “CSV Import to Excel” for advise on import to an Excel spreadsheet

<https://cc.au.dk/en/cdmm/tools-and-tutorials/data-cleaning/csv-import-to-excel>

# Converting the Data



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	F	G	H	I	J
1	timestamp	author_fullname	author_avatar_url	type	url
2	21/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DMX4xS7uRLt
3	20/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DMVUChwR6xC
4	18/07/2025 09.00	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DMPW89Fp5-q
5	17/07/2025 16.15	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DMNj8UIRthR
6	15/07/2025 18.21	Denmark in USA	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DMLow0PiQ58
7	14/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DMF3Tu1Sc5u
8	13/07/2025 19.45	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DMDo05eyMF1
9	12/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DMAtrqixXBF
10	11/07/2025 16.15	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DL-HNCeBB0r
11	09/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DL4_RfmppcN
12	07/07/2025 16.15	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLz29y6BeDI
13	04/07/2025 16.20	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLsGJ81Rqye
14	03/07/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DLpikLGRfr
15	02/07/2025 15.25	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLm2IoppuzI
16	01/07/2025 16.00	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DMVUBtIhL-5
17	26/06/2025 16.50	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DLXjRnIhJtq
18	23/06/2025 16.15	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DLPw1vrp62U
19	22/06/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLNNwtnR-AW
20	20/06/2025 15.45	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DLH_C1JB85r
21	19/06/2025 16.10	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLFdFYox0XL
22	16/06/2025 19.18	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DLASV1OxADV
23	12/06/2025 16.30	Denmark in California	https://scontent-cph2-1.cdninstagram.com/mixed		https://www.instagram.com/p/DKzd1U_pJ8P
24	11/06/2025 16.15	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DKw3UARhF9b
25	10/06/2025 15.10	Denmark in California	https://scontent-cph2-1.cdninstagram.com/unknown		https://www.instagram.com/p/DKw3UARhF9b

You now have a data sheet for the harvested posts with post ID, author, text, metadata such as timestamp, geolocation, number of likes, hashtags, content URLs, etc.

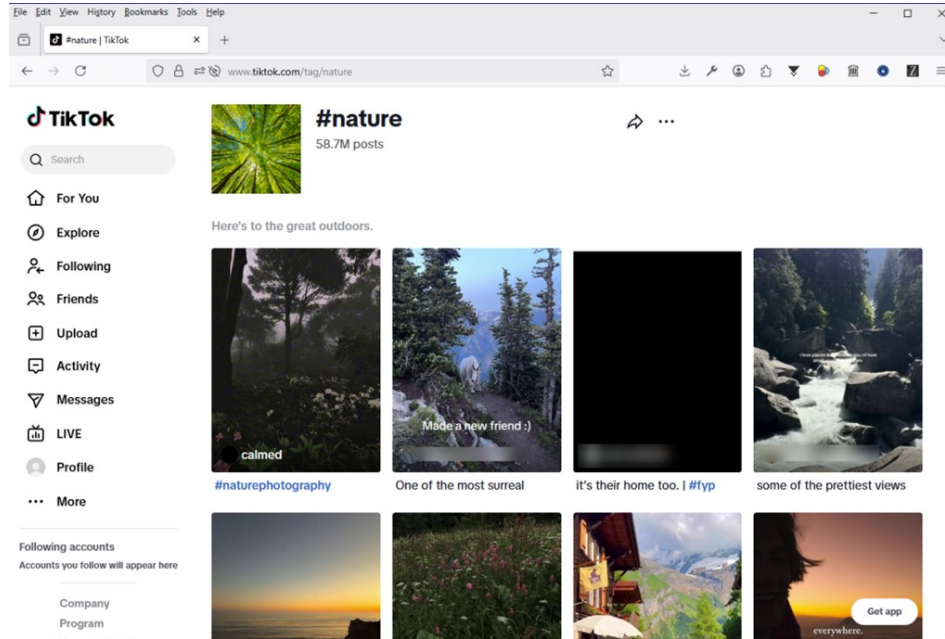
# Harvesting TikTok Data

The methods demonstrated for Instagram apply to all harvest projects for social media supported by Zeeschuimer.

A less detailed walkthrough of a TikTok harvest will provide examples of variations and limitations that may be encountered

# Harvesting TikTok Data

This is the list of main TikTok posts from a search for the hashtag #nature

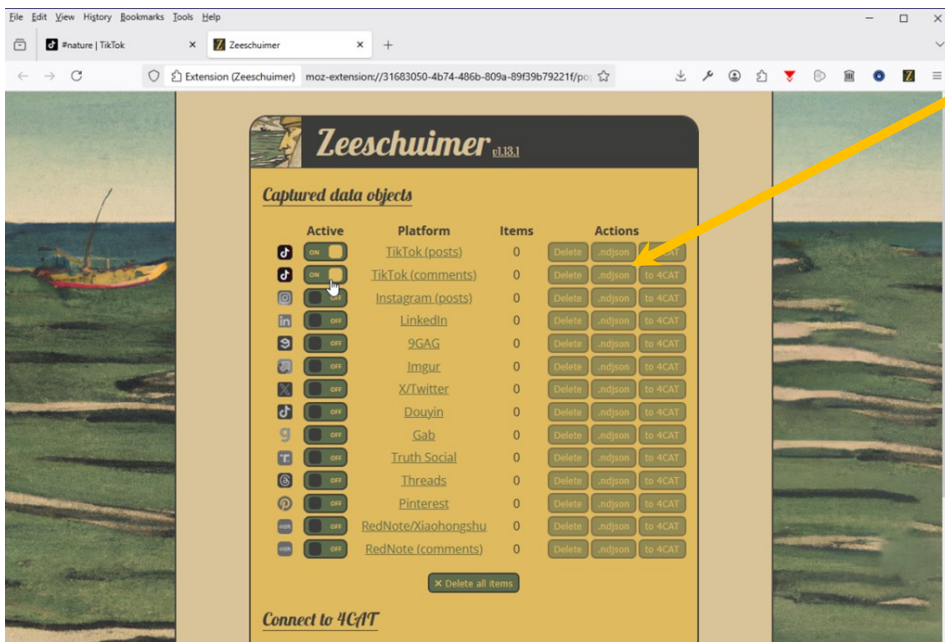


Remember the TikTok warning from slide 4

All tests with TikTok shown here were thus conducted on a private computer

# Harvesting TikTok Data

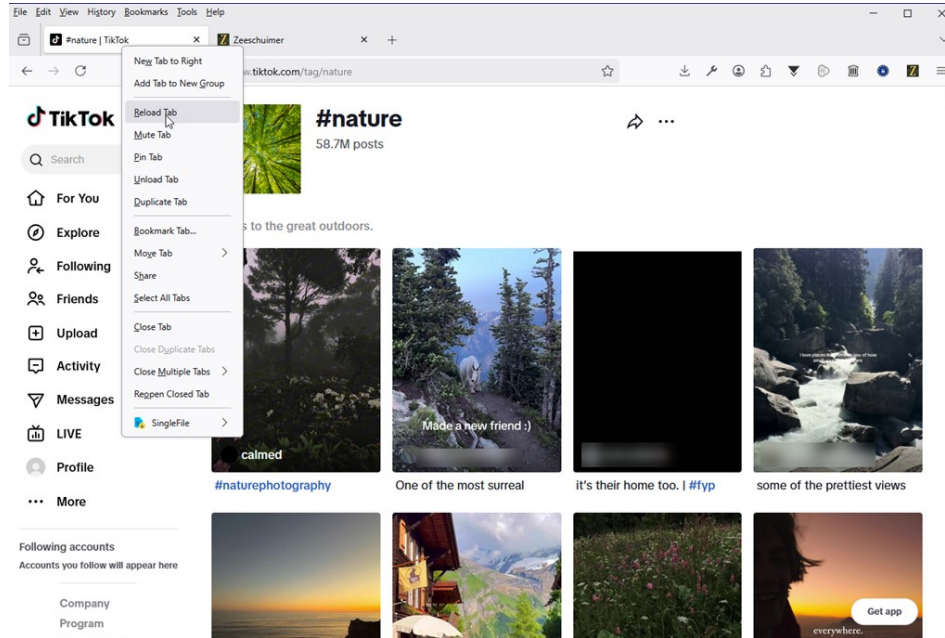
In Zeeschuimer both the harvest function for posts and for comments will be used, so both functions are activated here



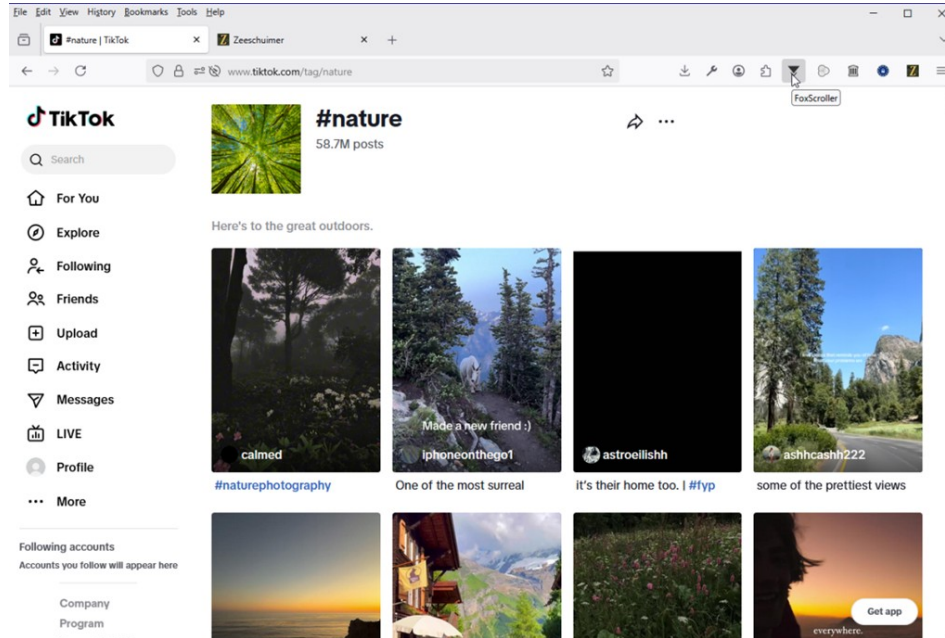
Two data sets will be generated, one for posts and one for comments

# Harvesting TikTok Data

Go back to the  
TikTok page and  
reload it



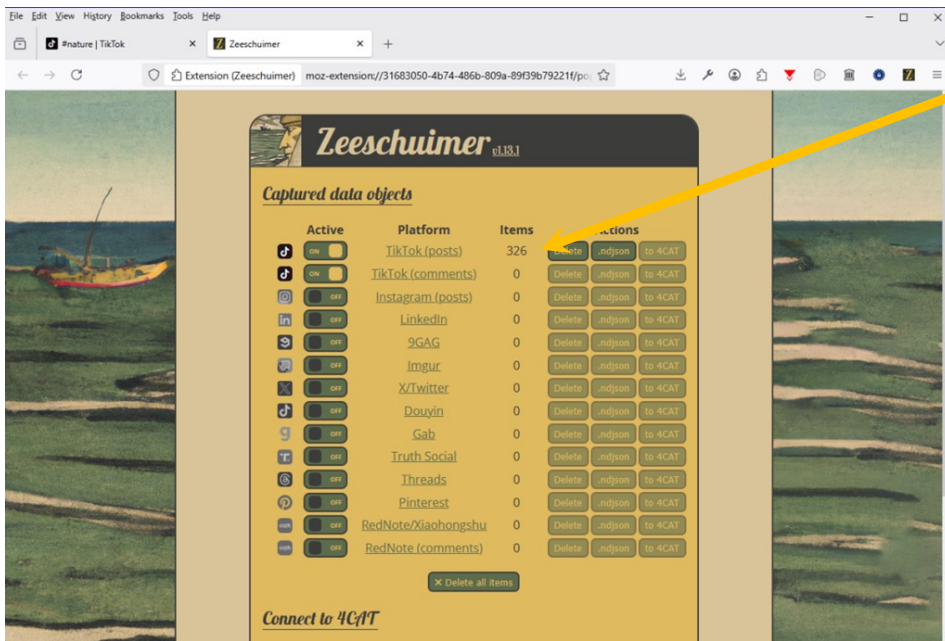
# Harvesting TikTok Data



**Start FoxScroller**

# Harvesting TikTok Data

After FoxScroller has finished scrolling down the page with search results...



The screenshot shows the Zeeschuimer extension interface overlaid on a browser window. The interface displays a table of captured data objects. A yellow arrow points to the 'Items' column for the first row, which shows '326' items for 'TikTok (posts)'. The table lists various social media platforms and their corresponding item counts.

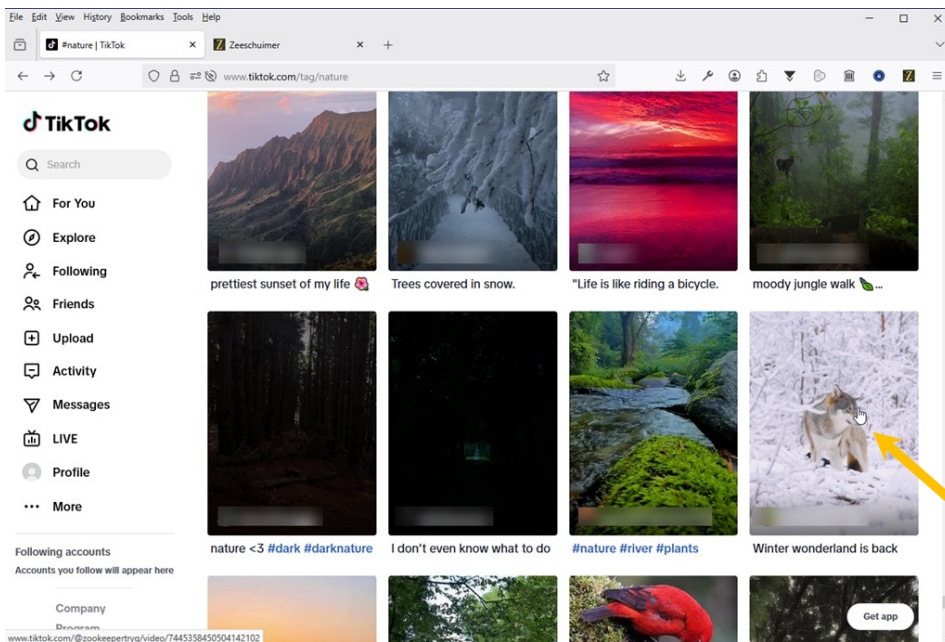
Active	Platform	Items	Actions
<input checked="" type="checkbox"/>	TikTok (posts)	326	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	TikTok (comments)	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Instagram (posts)	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	LinkedIn	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	9GAG	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Imgur	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	X/Twitter	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Douyin	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Gab	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Truth Social	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	TThreads	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Pinterest	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	RedNote/Xiaohongshu	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	RedNote (comments)	0	Delete, ndjson, to 4CAT

Connect to 4CAT

...data for 326 posts have been captured

# Harvesting TikTok Data

In order to capture data for comments one has to enter posts

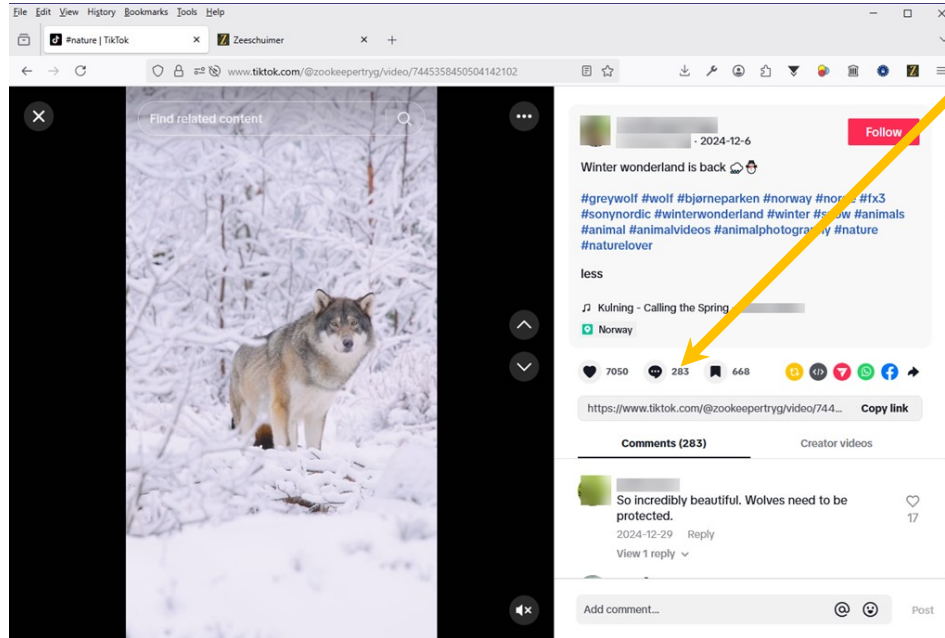


Unfortunately, post harvesting needs to be done manually, by entering the relevant posts one wishes to get the data from

A post is selected here

# Harvesting TikTok Data

The selected post has been entered



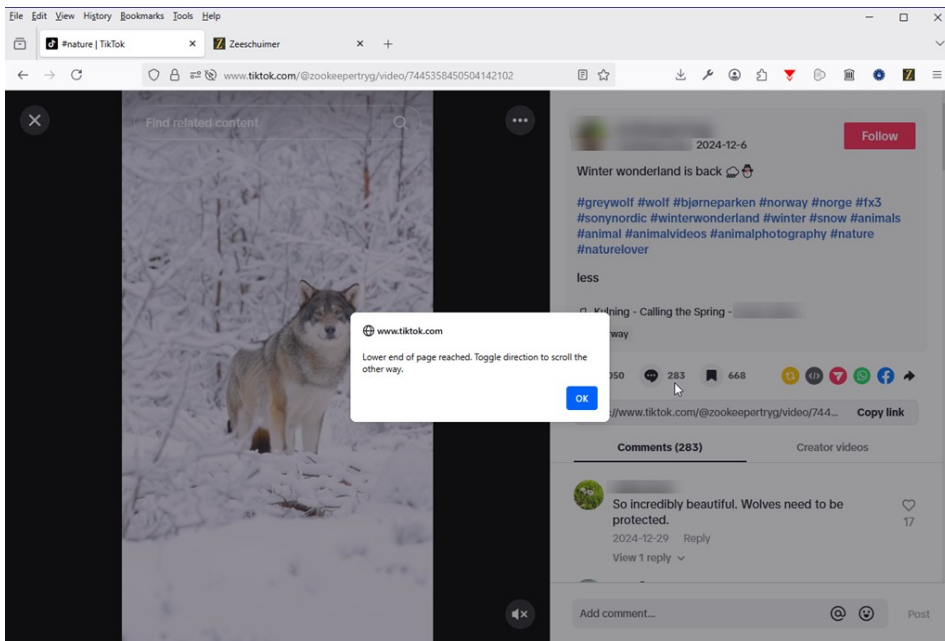
283 comments are listed

The count includes comments and replies

It also includes comments that have been removed again

# Harvesting TikTok Data

**At this point, FoxScroller registers the main post as the page content. With nothing below it registers the page as fully scrolled down**



**This can be solved by maximising the browser window. The website has responsive design, and comments will then be placed under the posted video**

# Harvesting TikTok Data

The screenshot displays the Zeeschuimer extension interface. At the top, it says "Zeeschuimer v1.13.1". Below that, it lists "Captured data objects" in a table format. The table has four columns: "Active", "Platform", "Items", and "Actions".

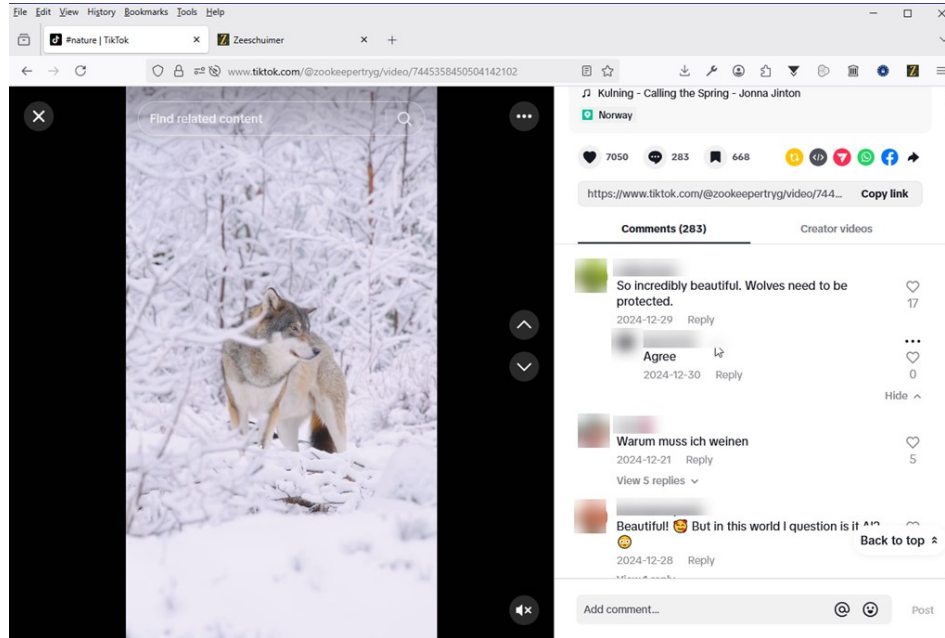
Active	Platform	Items	Actions
<input checked="" type="checkbox"/>	TikTok (posts)	326	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	TikTok (comments)	259	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Instagram (posts)	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	LinkedIn	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	9GAG	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Imgur	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	X/Twitter	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Douyin	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Gab	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Truth Social	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Threads	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	Pinterest	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	RedNote/Xiaohongshu	0	Delete, ndjson, to 4CAT
<input type="checkbox"/>	RedNote (comments)	0	Delete, ndjson, to 4CAT

At the bottom of the interface, there is a button that says "Delete all items" and a link that says "Connect to 4CAT".

After scrolling to the bottom, data for 259 out of the listed 283 posts is captured

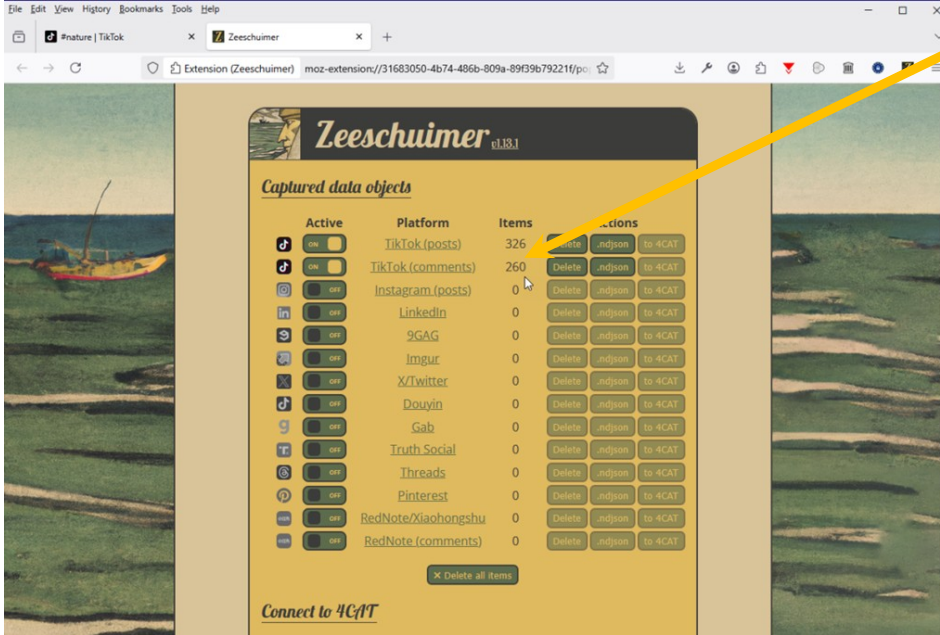
If replies need to be included in the data collection, they have to be opened manually

# Harvesting TikTok Data



**After going back to the top, a single reply has been opened**

# Harvesting TikTok Data

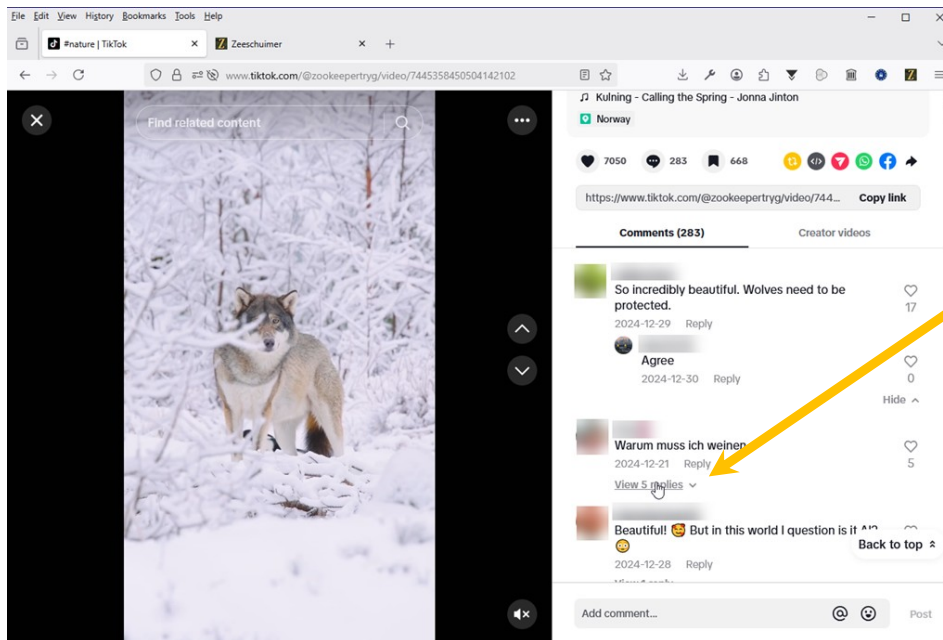


The screenshot shows the Zeeschuimer extension interface. The title bar indicates the browser is on the 'Zeeschuimer' extension page. The main content area is titled 'Zeeschuimer v1.33.1' and 'Captured data objects'. It displays a table of captured data objects across various platforms. The 'Items' column shows the count for each platform, with 'TikTok (posts)' at 326 and 'TikTok (comments)' at 260. A yellow arrow points to the '326' value in the 'Items' column for 'TikTok (posts)'. The table also includes columns for 'Active' status, 'Platform', and 'Actions' (Delete, ndjson, to 4CAT). At the bottom, there is a 'Connect to 4CAT' button and a 'Delete all items' button.

Active	Platform	Items	Actions
<input checked="" type="checkbox"/>	TikTok (posts)	326	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	TikTok (comments)	260	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Instagram (posts)	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	LinkedIn	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	9GAG	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Imgur	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	X/Twitter	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Douyin	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Gab	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Truth Social	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	TThreads	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	Pinterest	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	RedNote/Xiaohongshu	0	Delete, ndjson, to 4CAT
<input checked="" type="checkbox"/>	RedNote (comments)	0	Delete, ndjson, to 4CAT

**Going back to Zeeschuimer confirms that the reply just opened has been captured. The count has gone up to 260 from the initial 259**

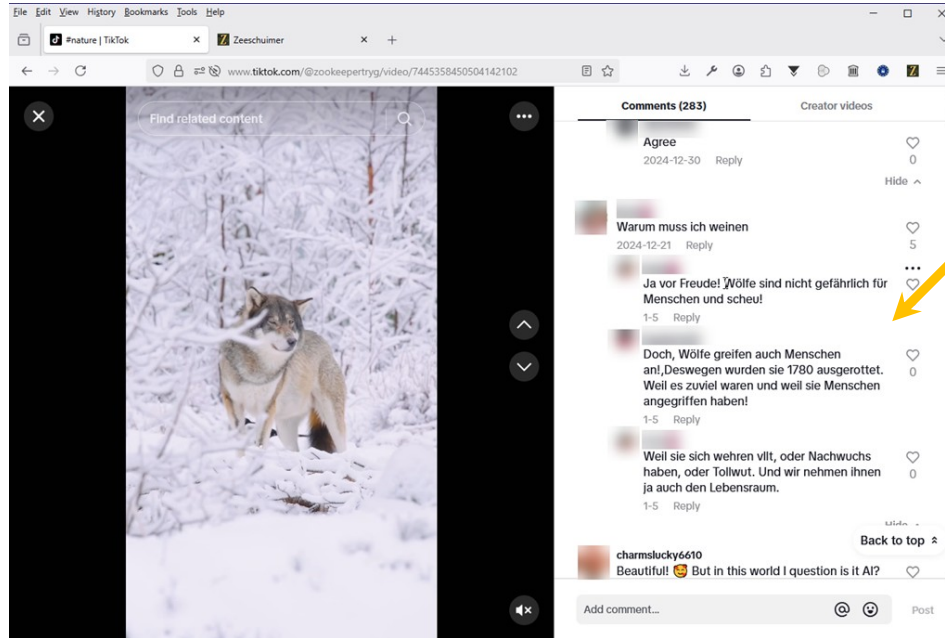
# Harvesting TikTok Data



Here the replies to another comment are about to be opened

Note that five replies are listed

# Harvesting TikTok Data

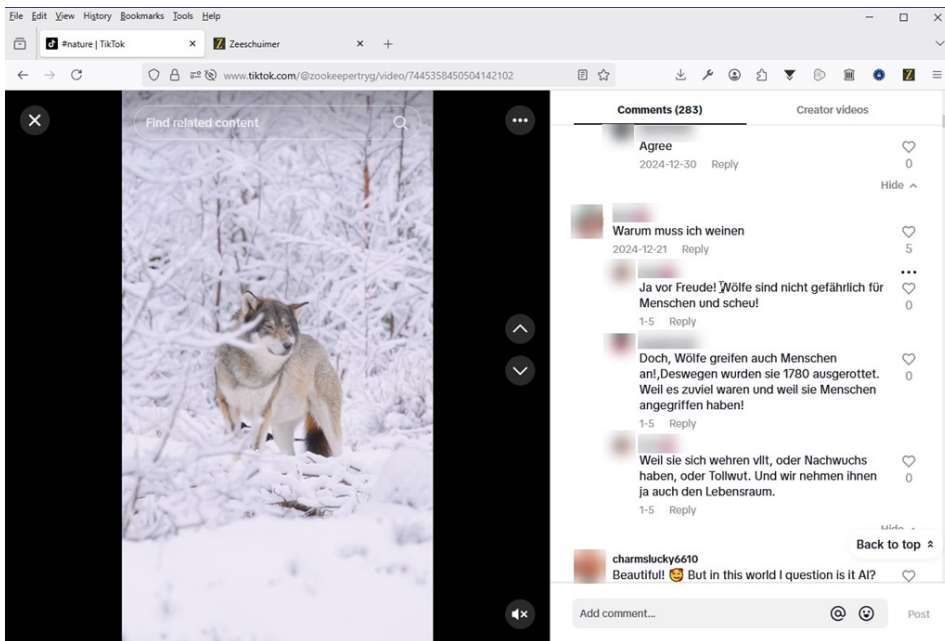


**Five replies were listed, but there are only three. Two of the counted replies have thus been deleted again.**

**The three remaining comments are added to the data**

# Harvesting TikTok Data

**Only comments that are still available can be added to the data. More posts may be added by going to them**



**When finished, the datasets for posts and comments can be downloaded and converted in the same manner as shown in the Instagram example**

# Data Overload

Before finishing, a few words on a problem that may occur:

For very long pages, a huge load of data from hundreds of images and other content can at some point cause the browser to overload.

Further scrolling can become impossible, and content may not load correctly (e.g. greyed out images)

## Data Overload

For problematic large pages this may help, but it is no guarantee for a full and complete harvest:

- Avoid working with other browser windows open,
- Use a powerful computer with high memory capacity (e.g. 32 or 64 GB RAM)
- Do not scroll down too fast, try lowering the speed

## Data Overload

Further advice for very large pages:

Restart the browser and try to collect more data

- by sections such as topics and playlists,
- if possible, by using different sorting options.

This can provide more datasets that may then be combined and cleaned of duplicates - but getting complete datasets for very large pages may not be possible with a hands-on approach

## Final Words

Zeeschuimer works well for moderate data collection

The process can to some extent be automated, but all data collection is limited to what is actively opened

It is your decision depending on your research question(s) how much manual harvesting should be done

CDMM hopes that you will find this tutorial useful