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 registrered, the account used for the process may be blocked.

To harvest, e.g. several accoounts, 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.

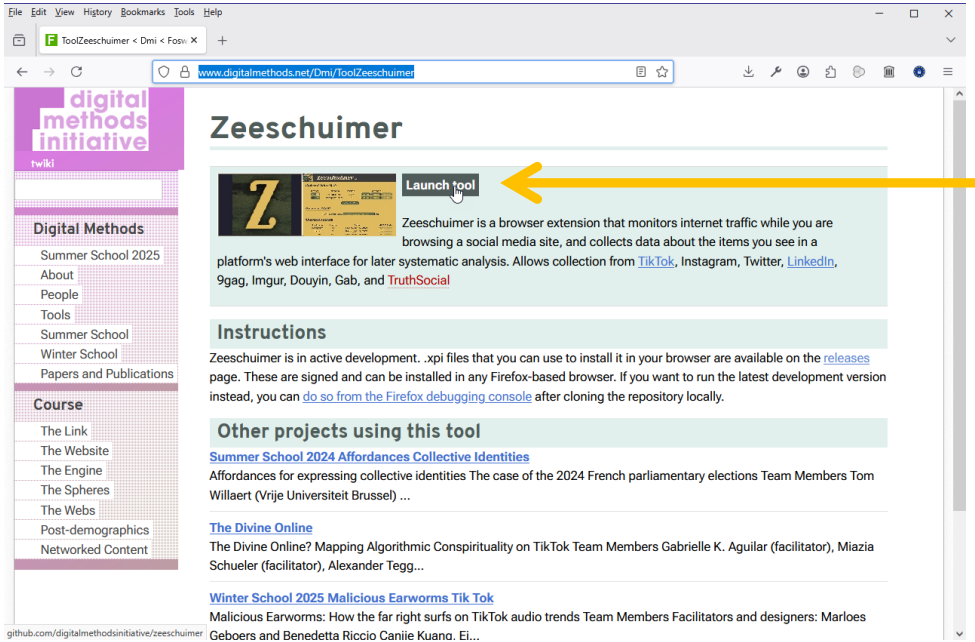
Firefox is thus a prerequisite.

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

Setting up Zeeschuimer

In Firefox, visit the developer page for background

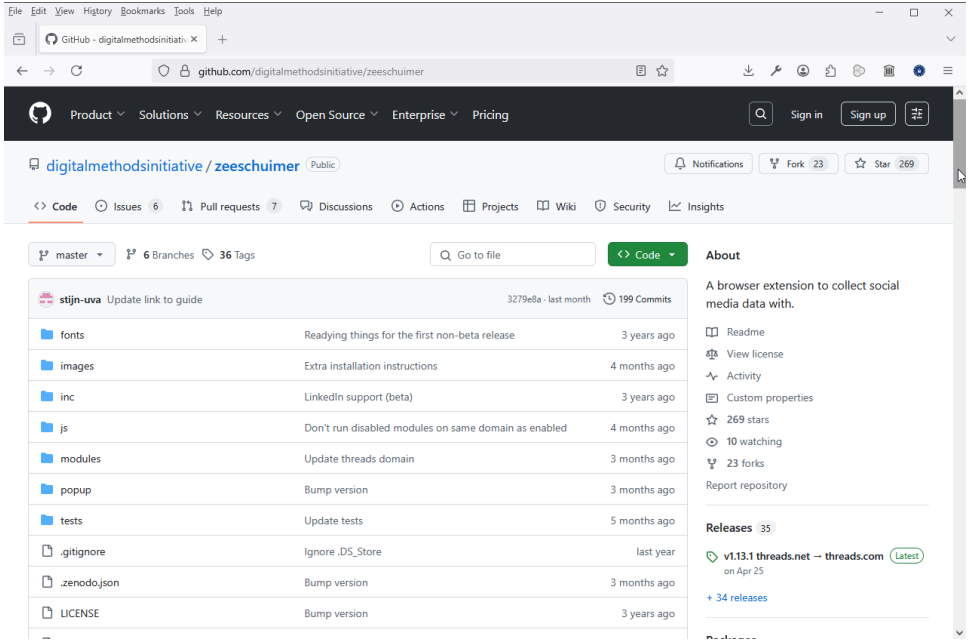
https://www.digitalmethods.net/Dmi/ToolZeeschuimer



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

Setting up Zeeschuimer

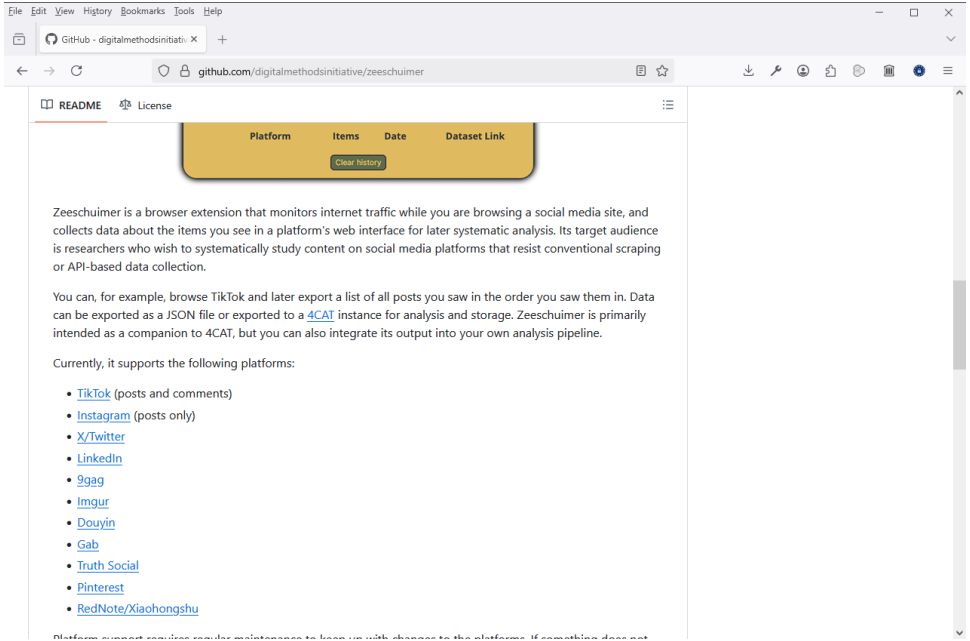
Zeeschuimer GitHub page:



It is worth scrolling down this page for information

Setting up Zeeschuimer

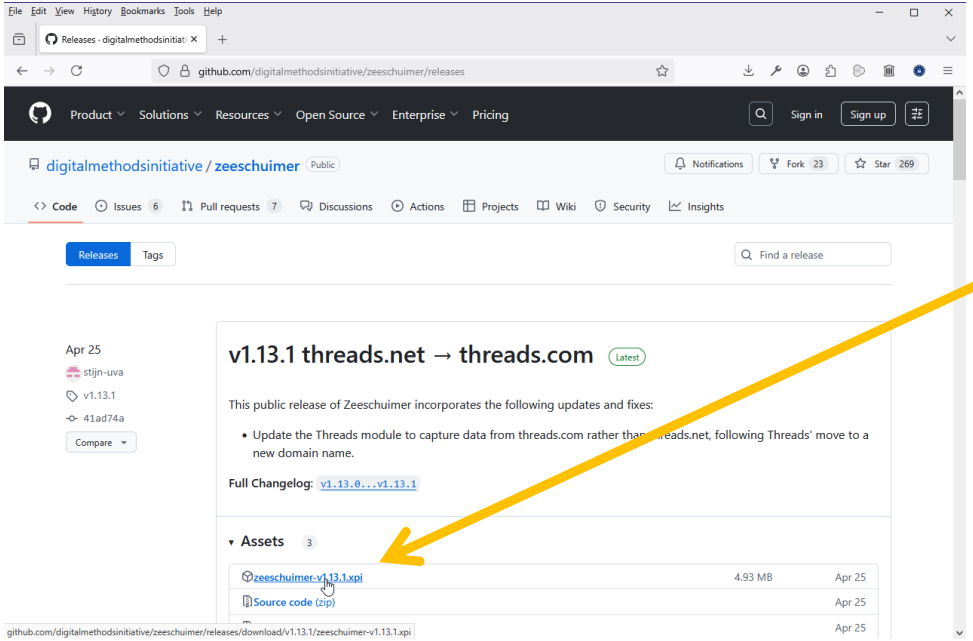
Zeeschuimer GitHub page:



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

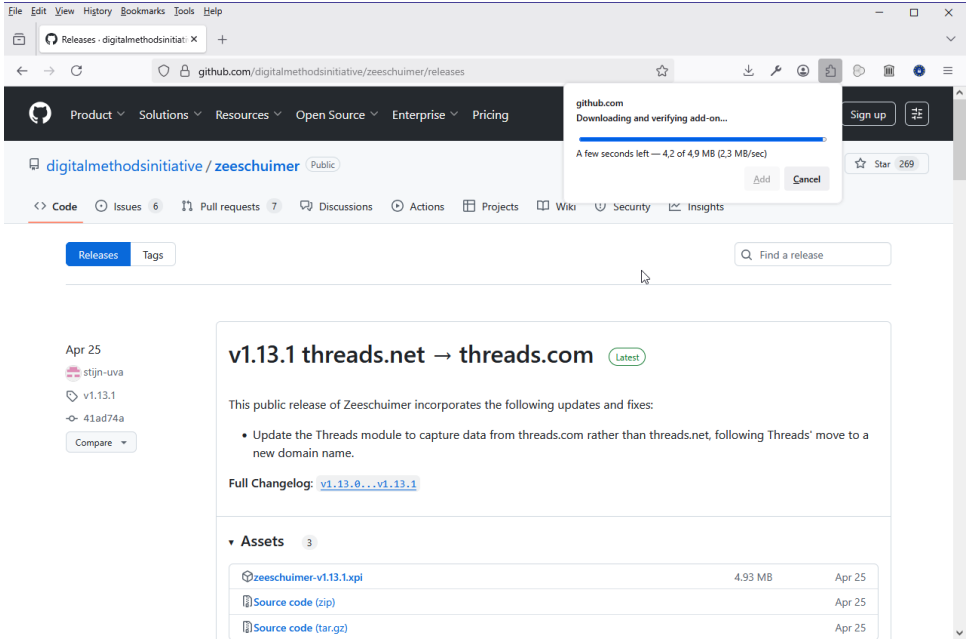
Setting up Zeeschuimer

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



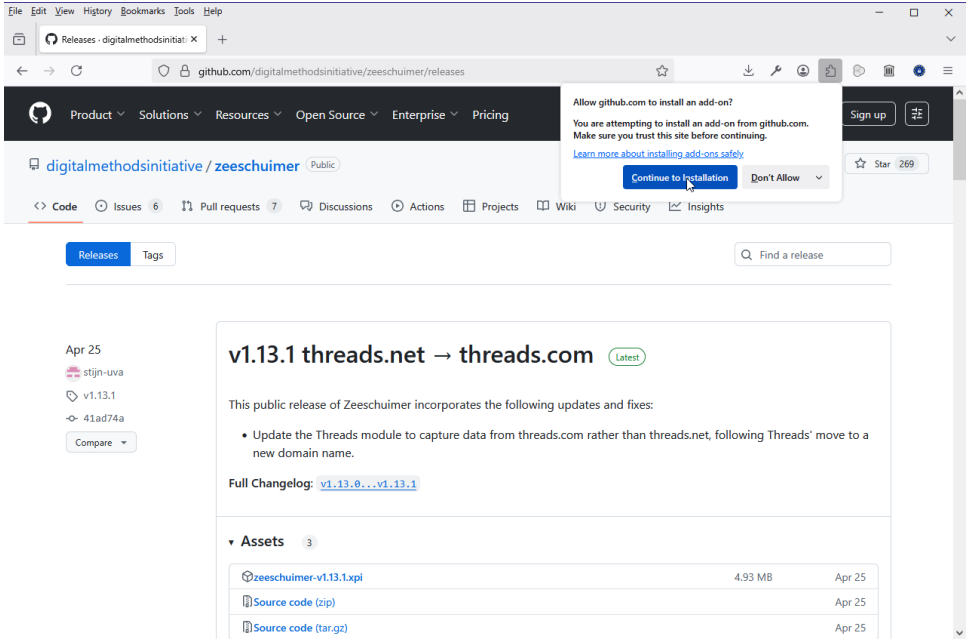
Click on the link for the XPI file for the latest version at the top of the list

Setting up Zeeschuimer



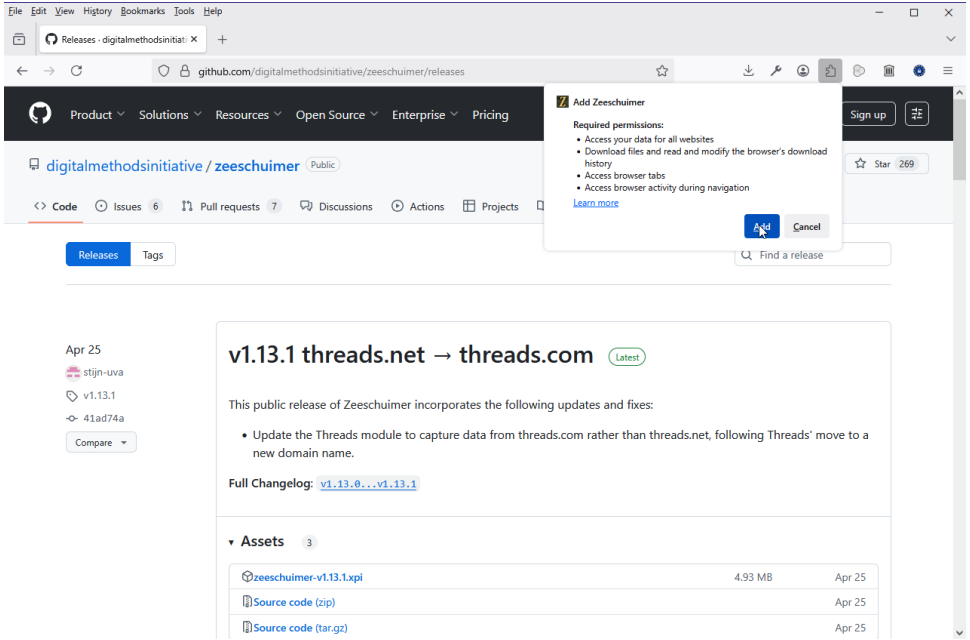
The file starts downloading

Setting up Zeeschuimer



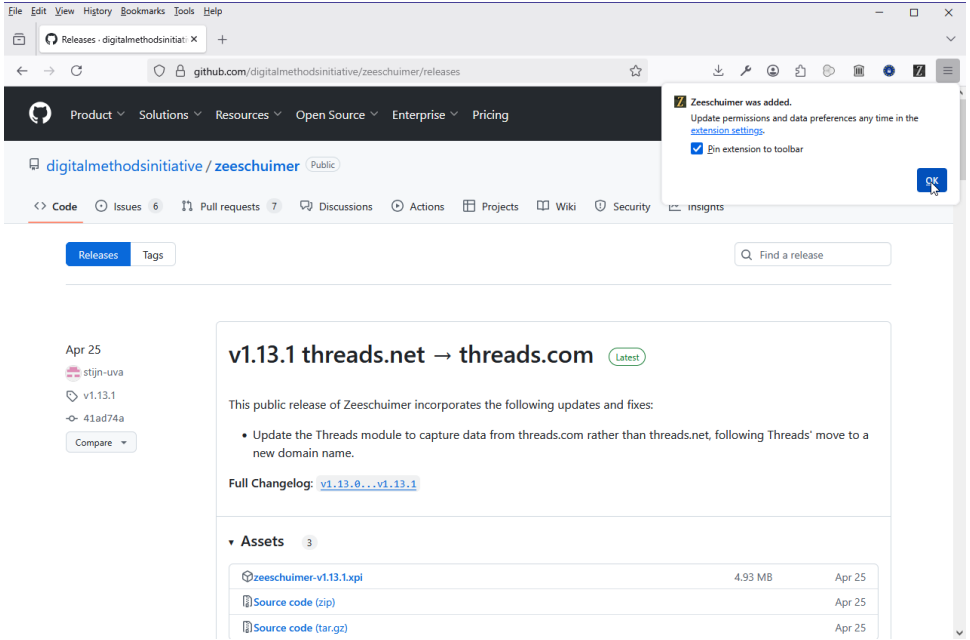
When the download is complete, installation of the add-on to Firefox becomes available. Click ”Continue to Installation”

Setting up Zeeschuimer



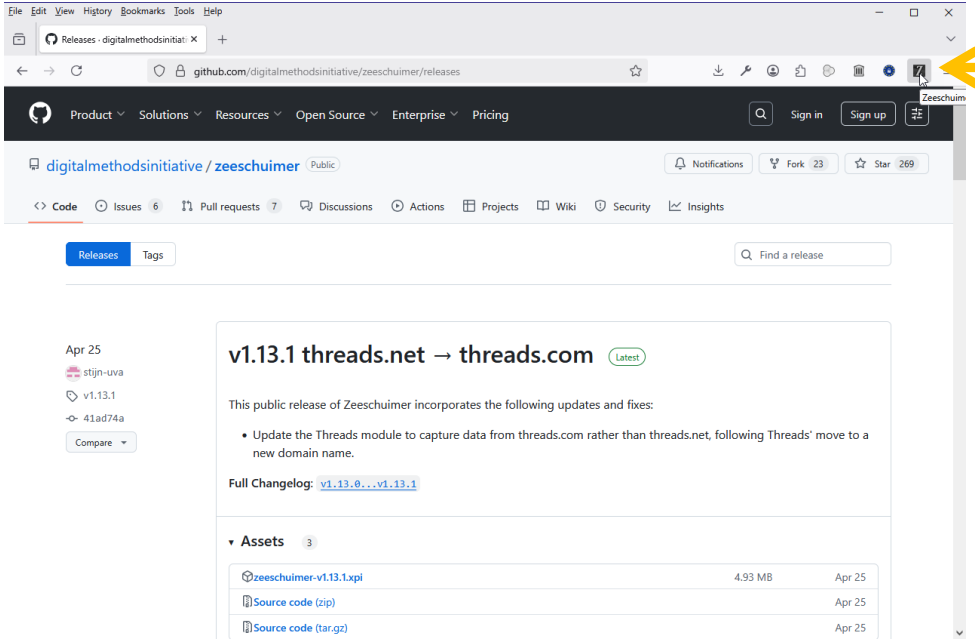
Click on ”Add” to confirm the required access for the add-on

Setting up Zeeschuimer



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

Setting up Zeeschuimer



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 automisation to your workflow. It is not a necessity, but it is used in the examples in this tutorial

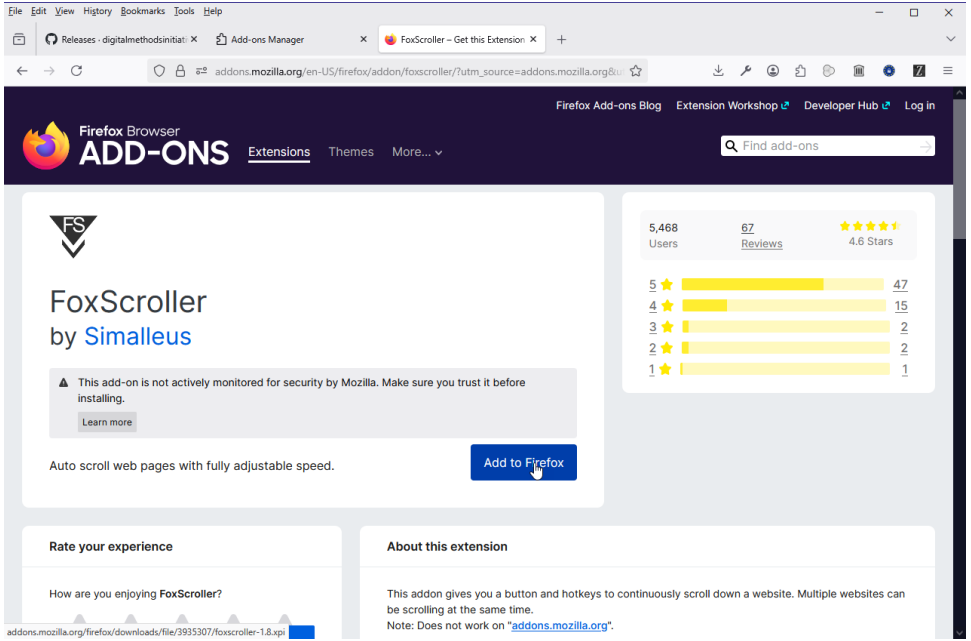
Adding FoxScroller

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/

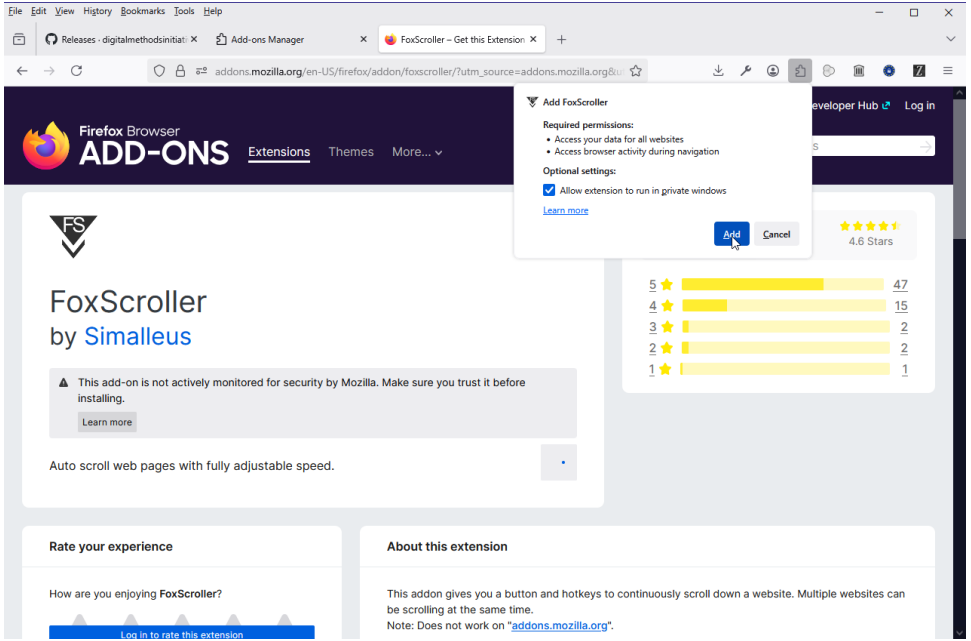
Adding FoxScroller

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



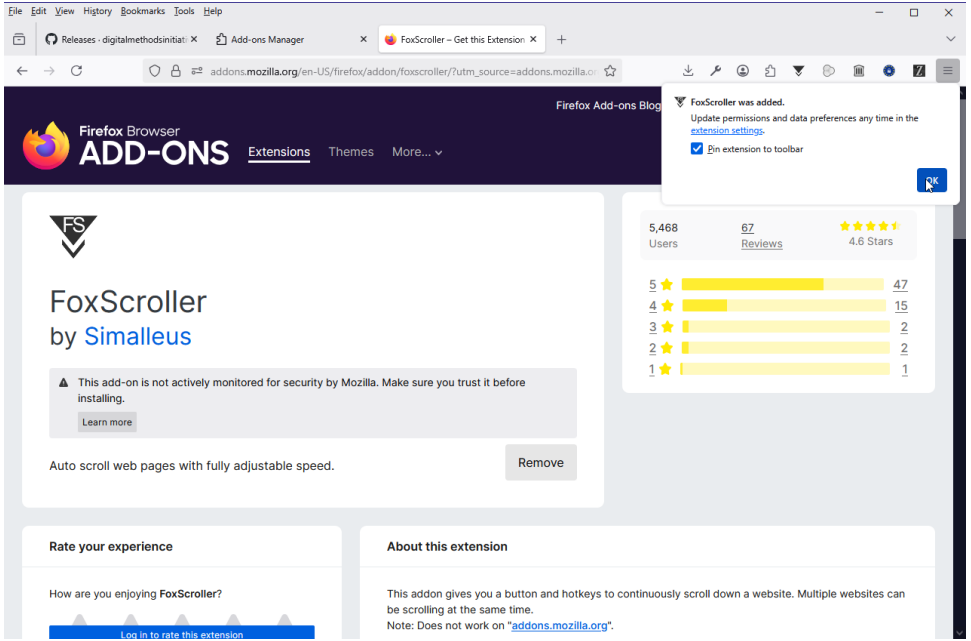
Click on ”Add to Firefox”

Adding FoxScroller



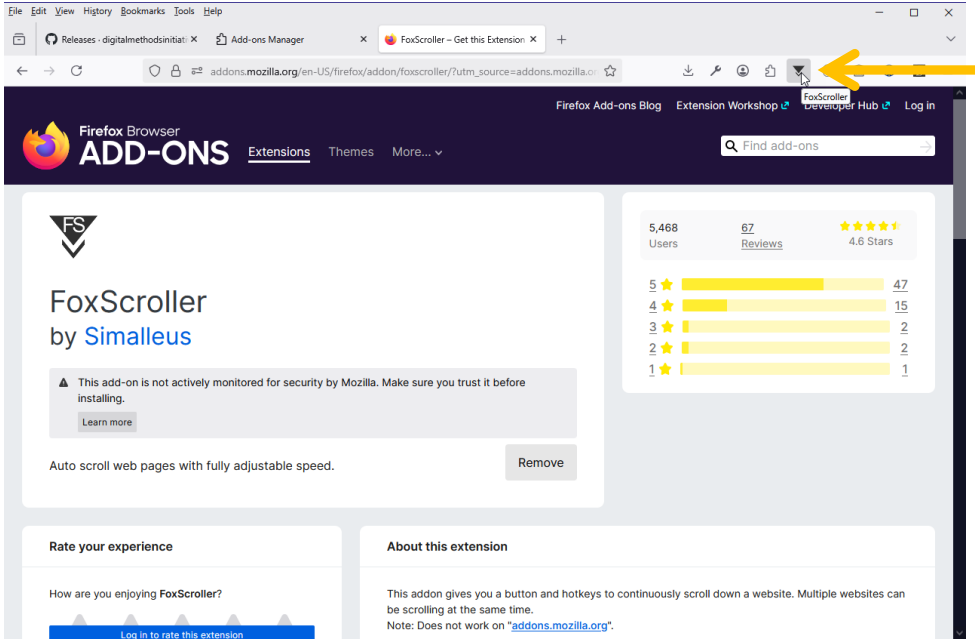
Click on ”Add”

Adding FoxScroller



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

Adding FoxScroller



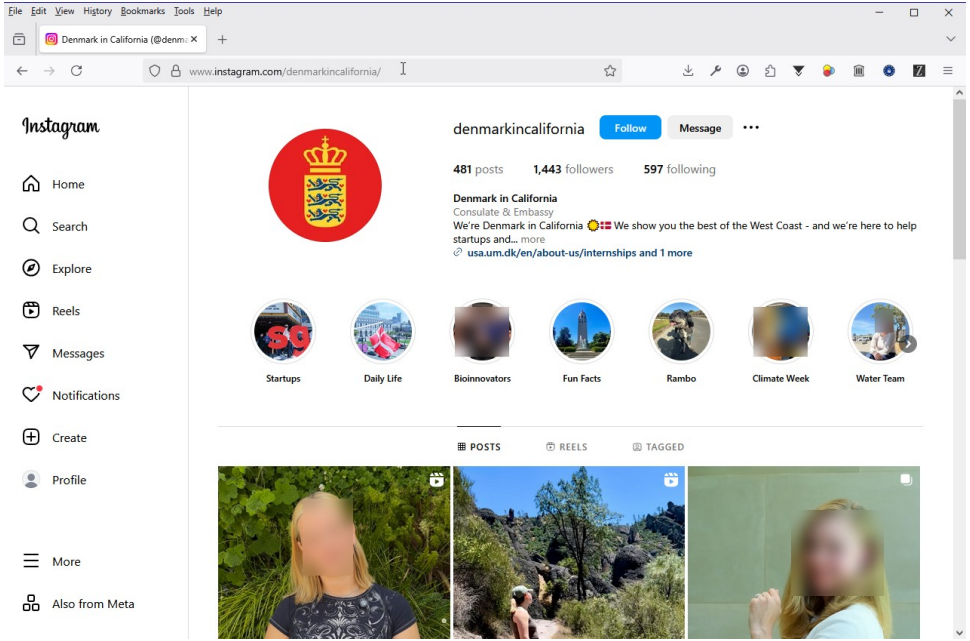
FoxScroller is now added and ready for use

Harvesting Instagram Data

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

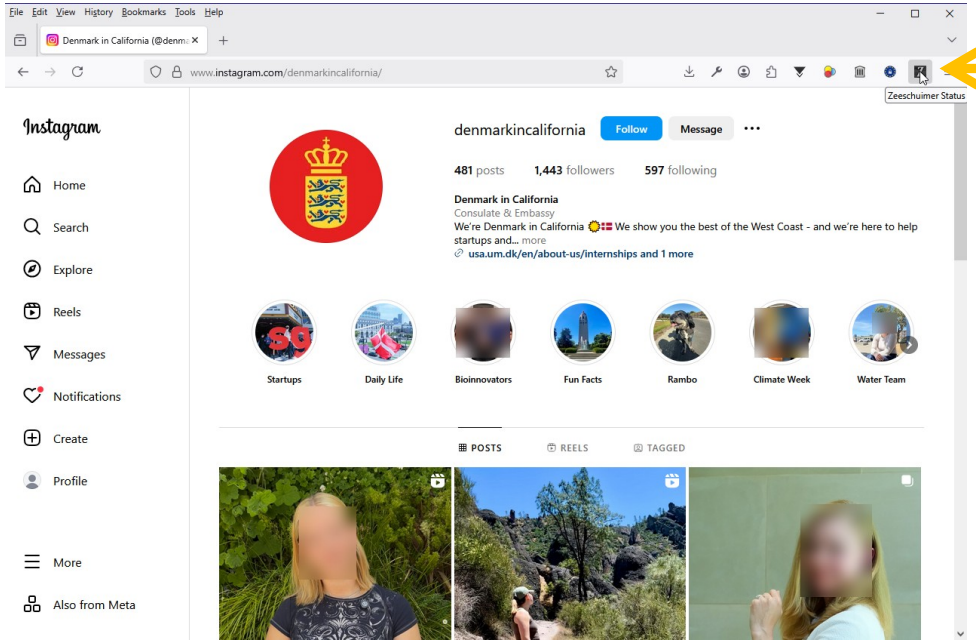
The procedure applies to all harvesting proccesses on various social media, but there may be variations, especially in the demand for hands-on user interaction. This will be addressed in the TikTok example which follows the first one on Instagram

Harvesting Instagram Data



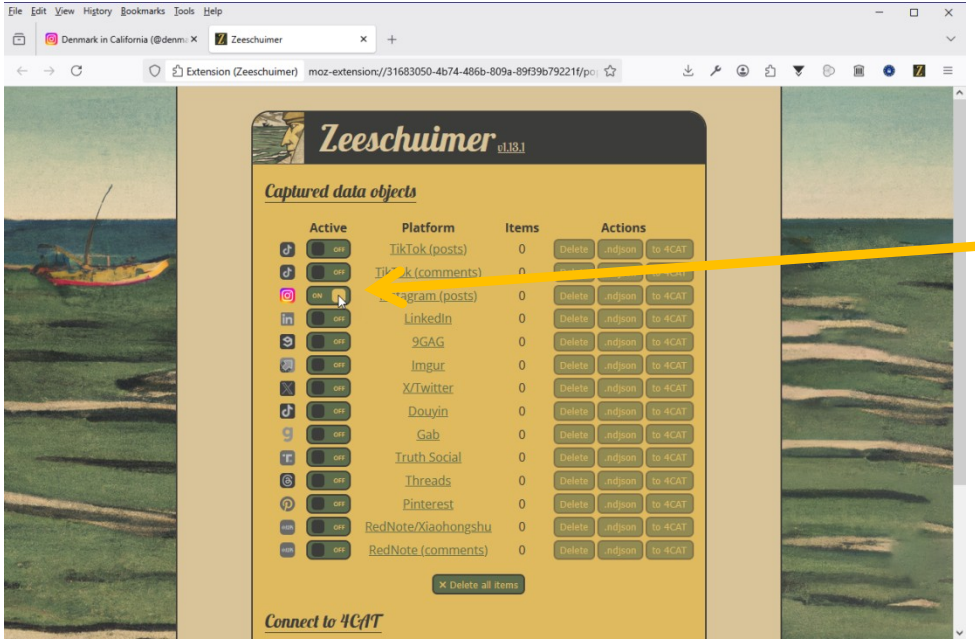
Go to an Instagram page that you want to harvest

Harvesting Instagram Data



Click on the Zeeschuimer shortcut

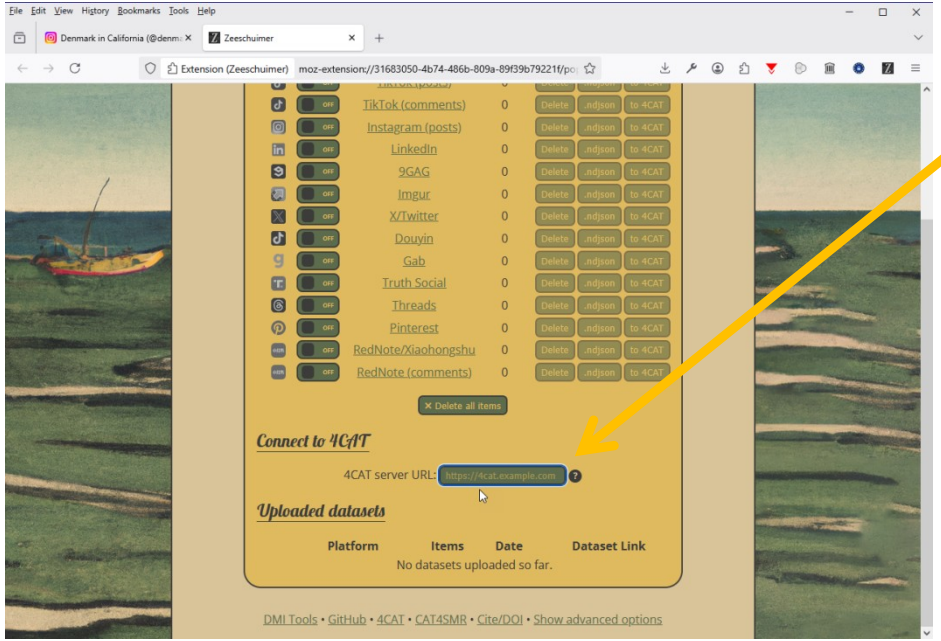
Harvesting Instagram Data



Zeeschuimer opens in a new tab

Activate post harvesting for Instagram

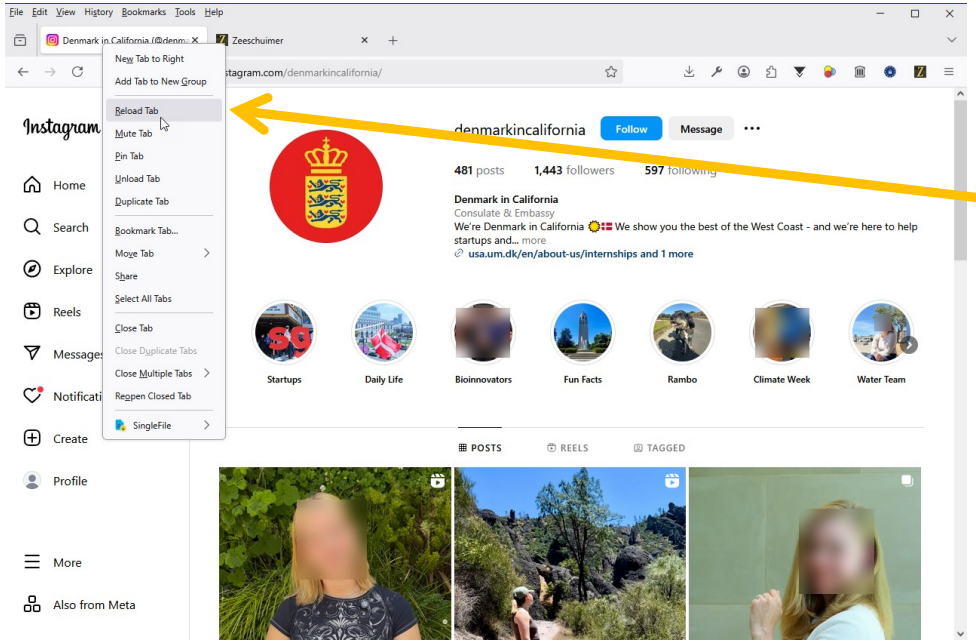
Harvesting Instagram Data



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

This tutorial continues with direct data downloads

Harvesting Instagram Data

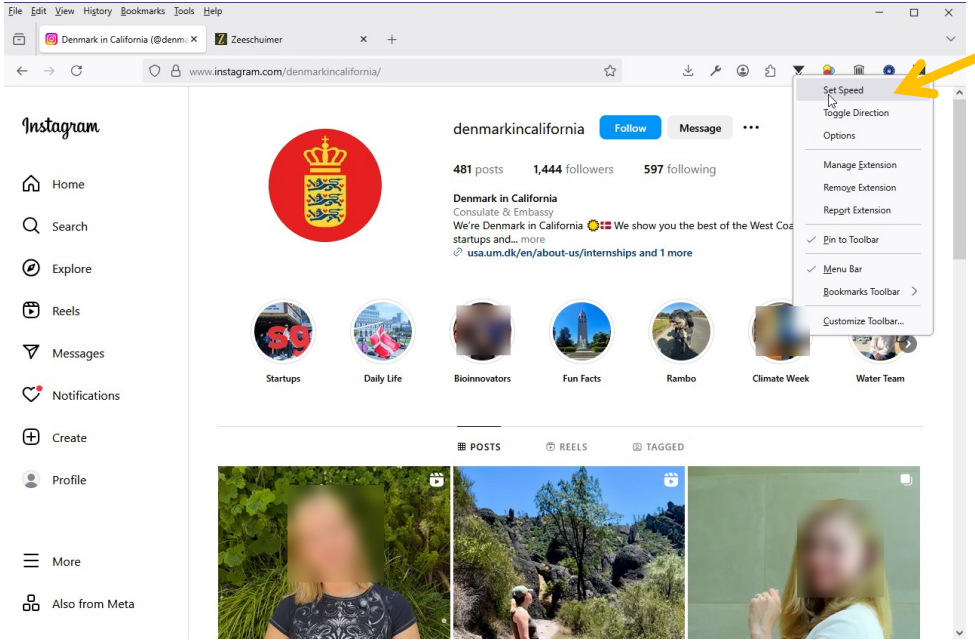


Go back to the Instagram tab and reload the page

After activating Zeeschuimer this is necessary for starting the data capture

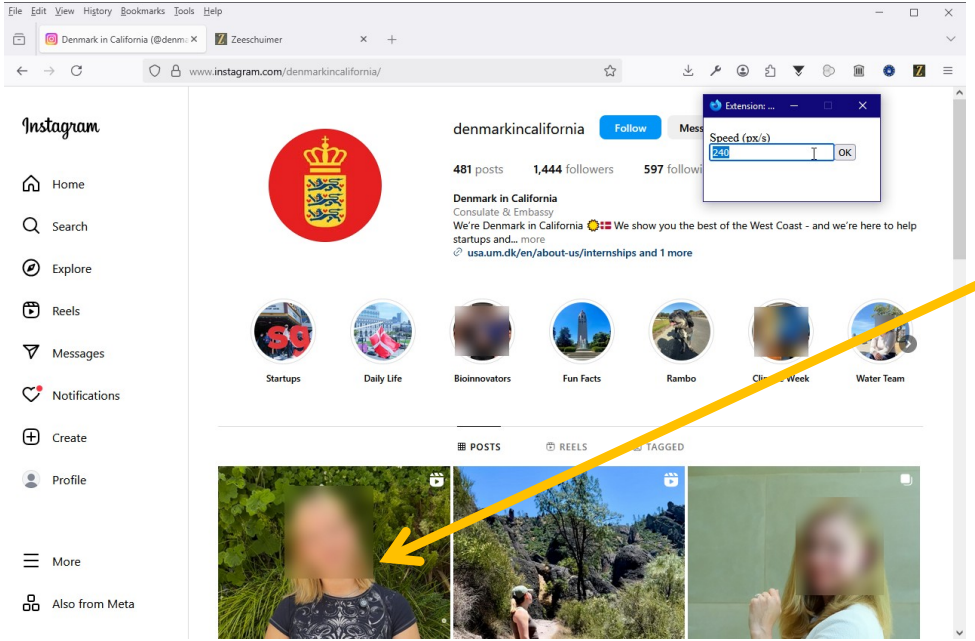
Harvesting Instagram Data

ForScroller 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



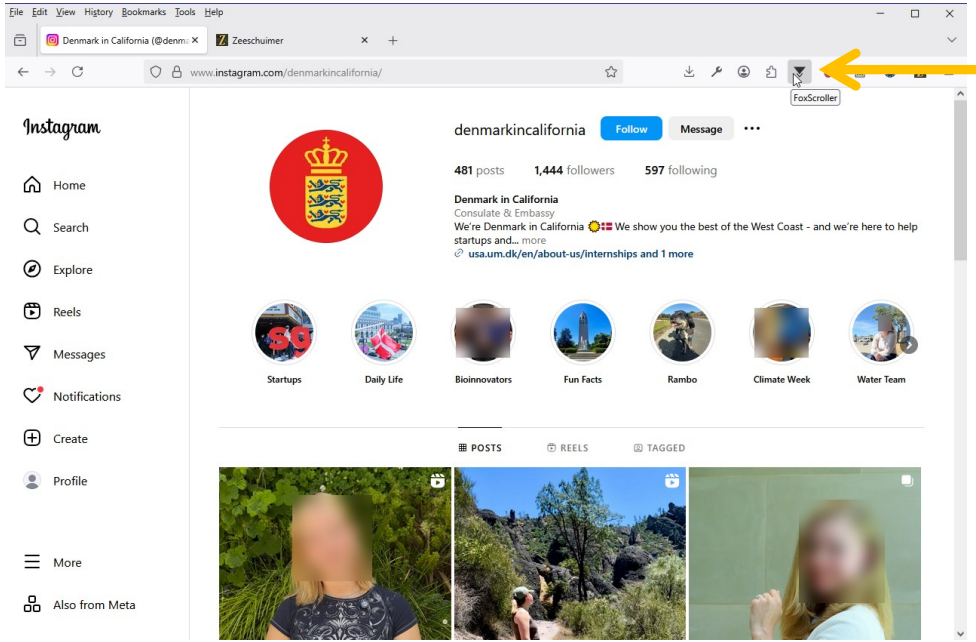
Right-click (ctrl+click in Mac OS) on the FoxScroller shortcut if you want to adjust the scolling speed.

Harvesting Instagram Data



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

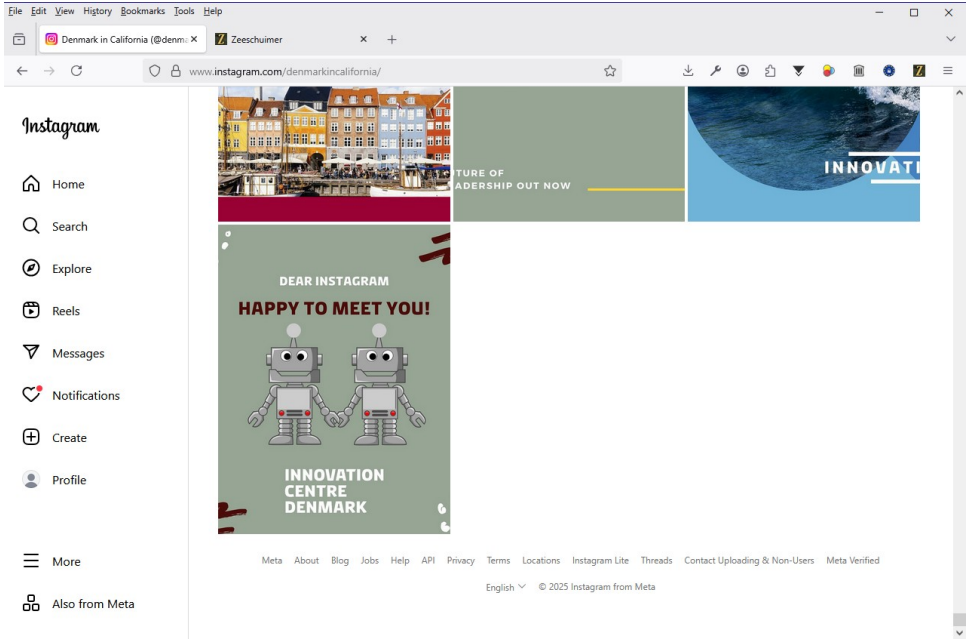


Now click on the FoxScroller shortcut to start automatic scrolling

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



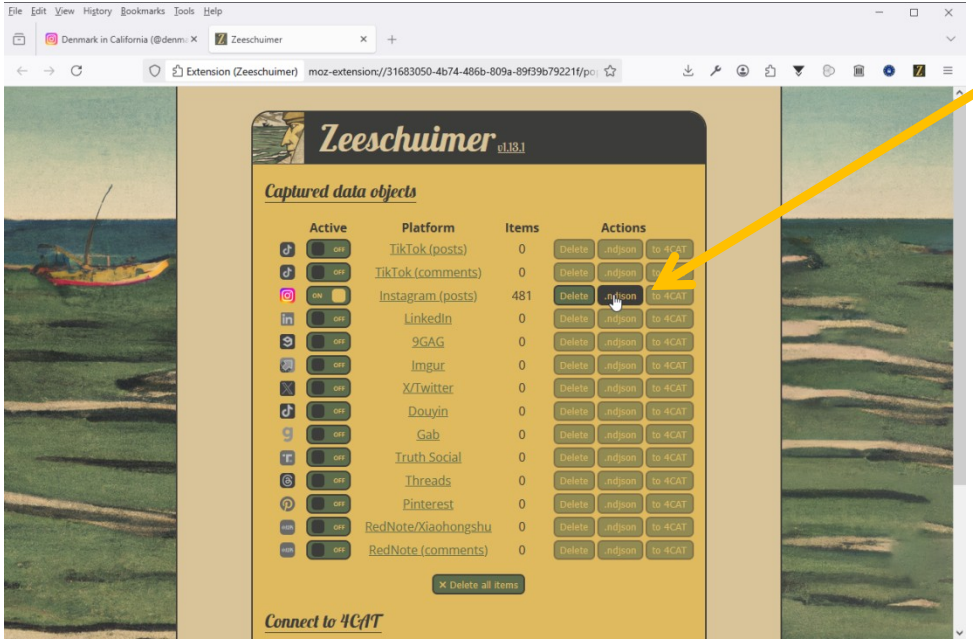
For very large pages, data overload in the browser may occur.

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

Harvesting Instagram Data

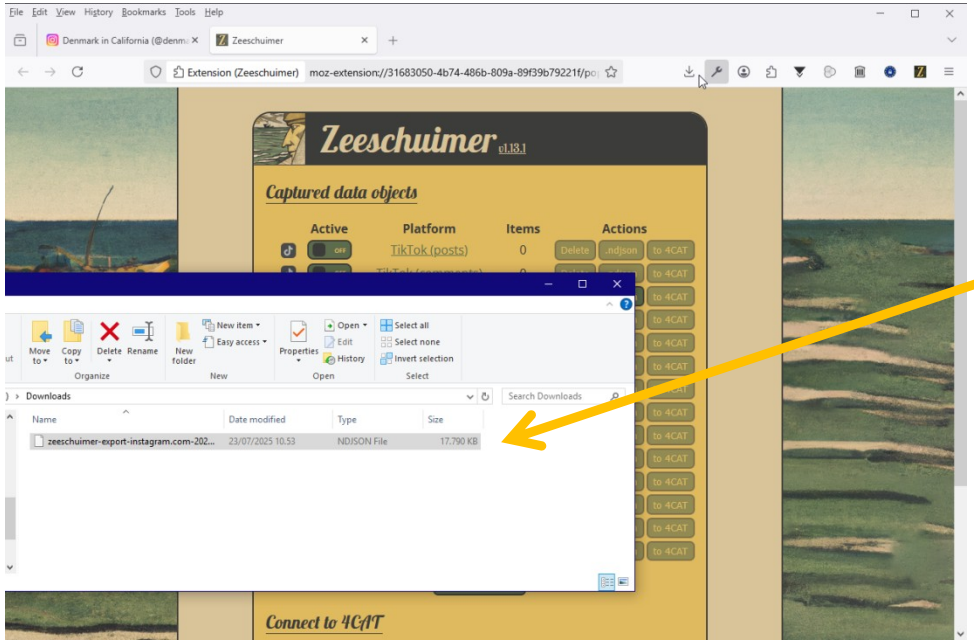
Now go back to the Zeeschuimer tab

Under ”Items”, 481 harvested posts are now harvested



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

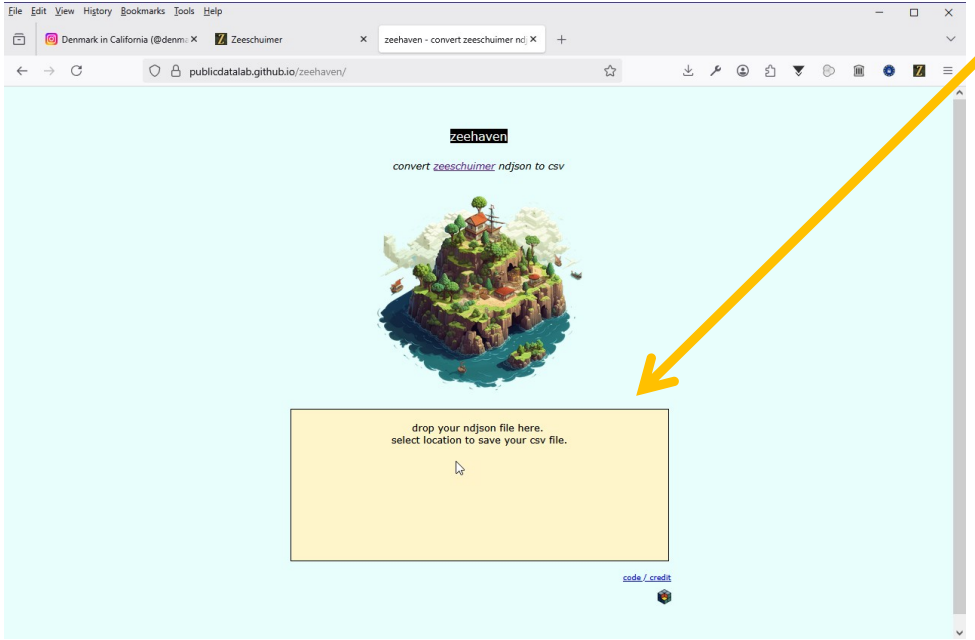
Zeehaven

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

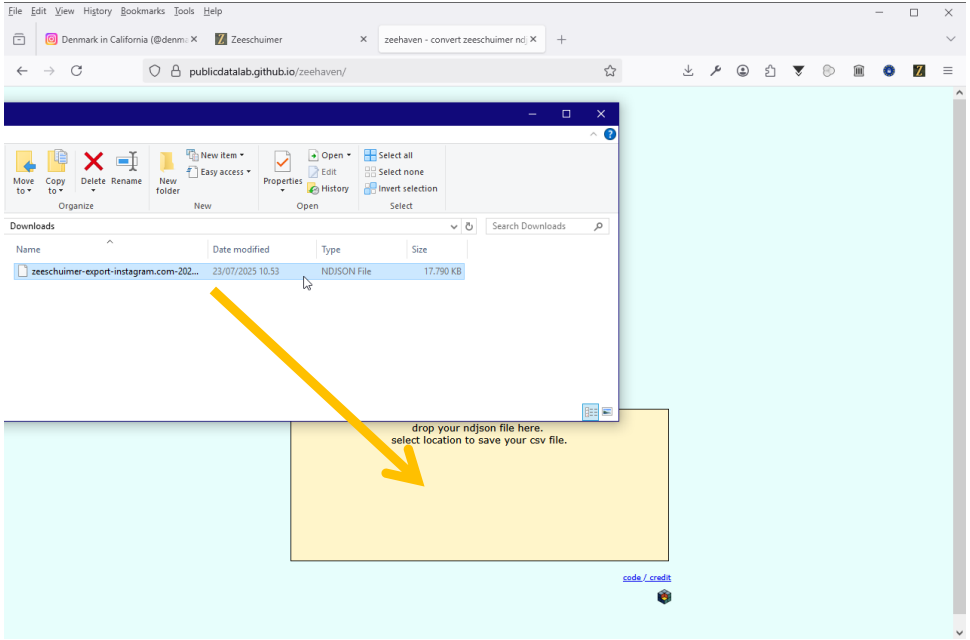
Go to the Zeehaven page at https://publicdatalab.github.io/zeehaven/



Notice the drag and drop field

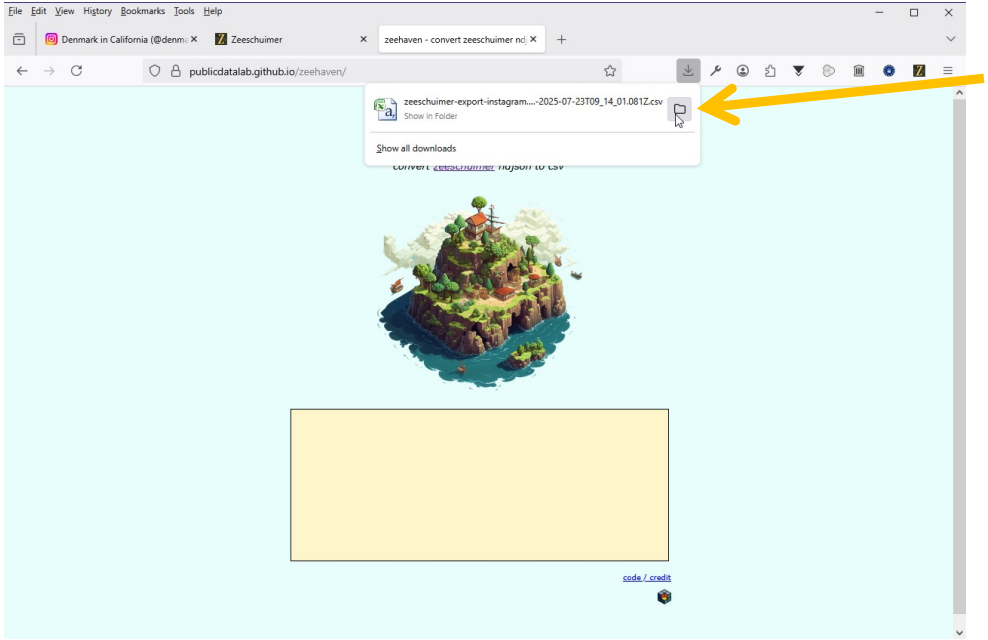
Converting the Data

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

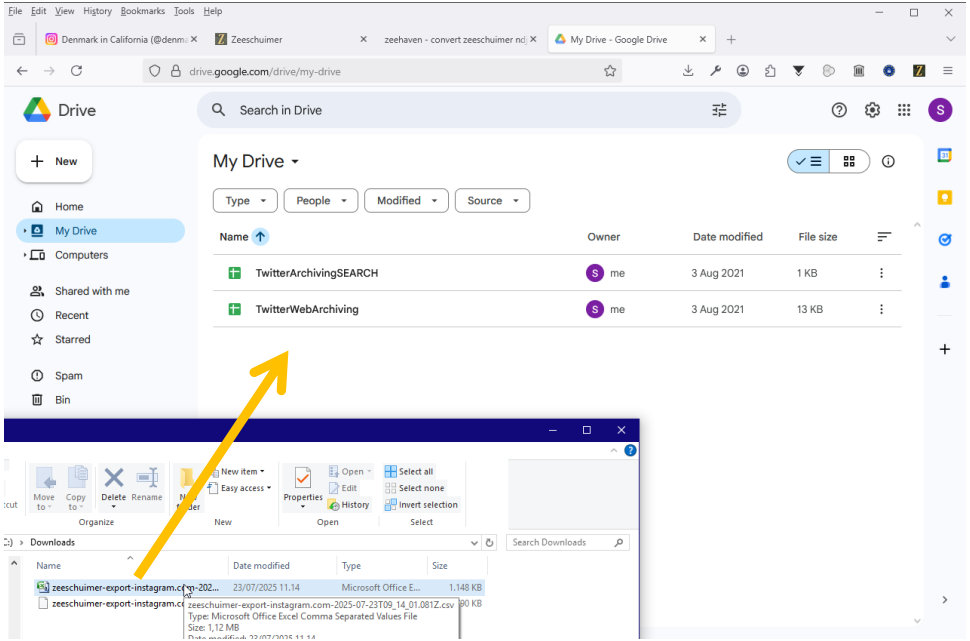
There are several ways to import a CSV file to a spreadsheet. Please consult CDMM's page “CSV Import to Excel” for advise on import to an Excel spreadsheet
https://cc.au.dk/en/cdmm/tools-and-tutorials/data-cleaning/csv-import-to-excel

A quick alternative spreadsheet conversion method will be demonstrated here.
It may not work for all CSV files, but it does for the file size and format from Zeehaven (up to 100 MB)

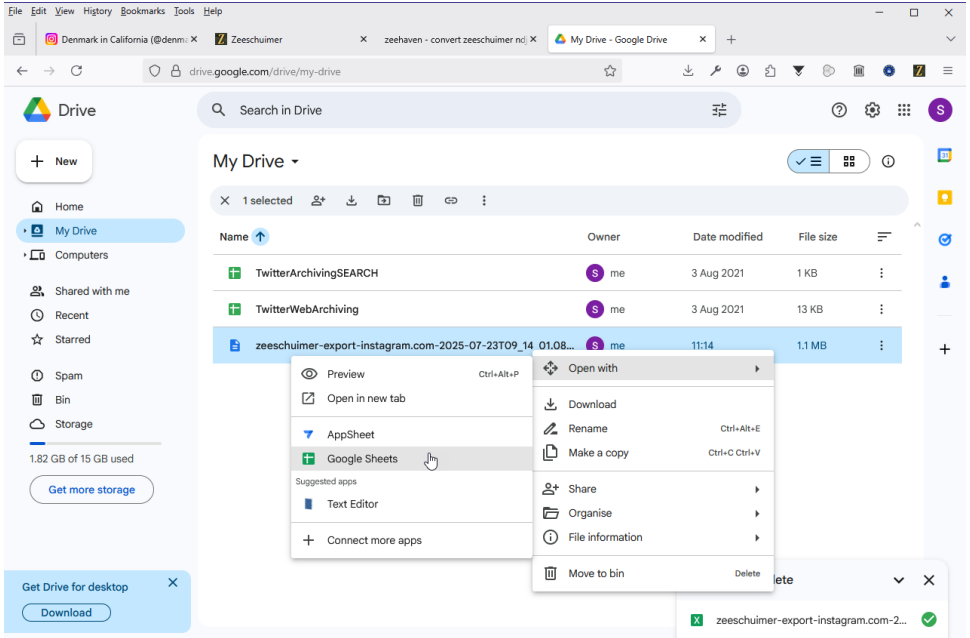
Converting the Data

Open your Google Drive (assuming you have one)

Drag your NDJSON data file to the My drive section

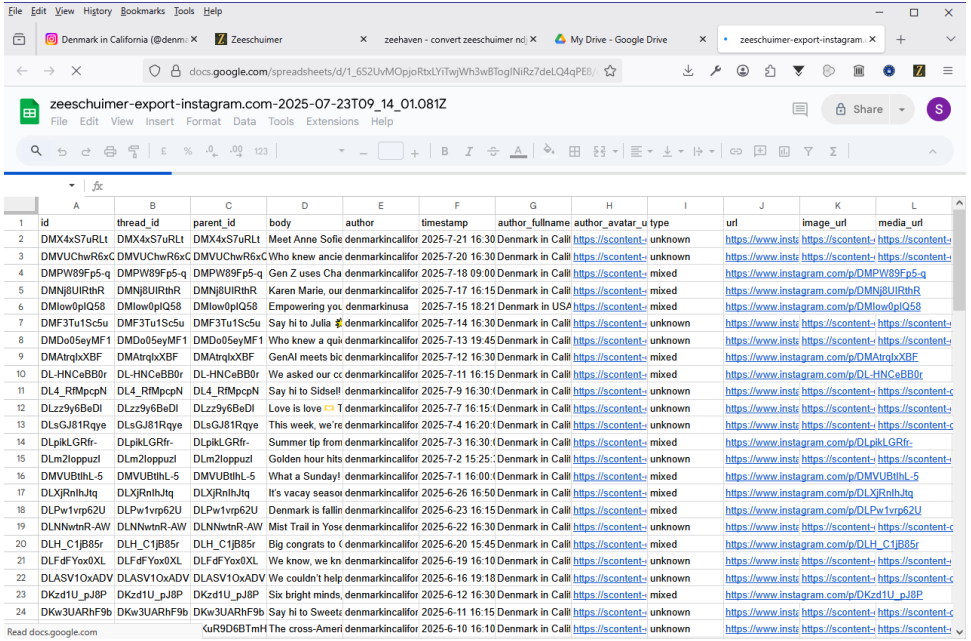


Converting the Data



Right-click (ctrl+click in Mac OS) on the file, choose ”Open with” and then ”Google Sheets”

Converting the Data



	A	B	C	D	E	F	G	H	I	J	K	L
1	id	thread_id	parent_id	body	author	timestamp	author_fullname	author_avatar_u	type	url	image_url	media_url
2	DMX4xS7uRLt	DMX4xS7uRLt	DMX4xS7uRLt	Meet Anne Sofie	denmarkincalifor	2025-7-21 16:30	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
3	DMVUChwR6xC	DMVUChwR6xC	DMVUChwR6xC	Who knew ancie	denmarkincalifor	2025-7-20 16:30	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
4	DMPW89Fp5-q	DMPW89Fp5-q	DMPW89Fp5-q	Gen Z uses Cha	denmarkincalifor	2025-7-18 09:00	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DMPW89Fp5-q		
5	DMNj8UIRthR	DMNj8UIRthR	DMNj8UIRthR	Karen Marie, our	denmarkincalifor	2025-7-17 16:15	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DMNj8UIRthR		
6	DMIow0pIQ58	DMIow0pIQ58	DMIow0pIQ58	Empowering you	denmarkinusa	2025-7-15 18:21	Denmark in USA	https://scontent-	mixed	https://www.instagram.com/p/DMIow0pIQ58		
7	DMF3Tu1Sc5u	DMF3Tu1Sc5u	DMF3Tu1Sc5u	Say hi to Julia 🐝	denmarkincalifor	2025-7-14 16:30	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
8	DMDo05eyMF1	DMDo05eyMF1	DMDo05eyMF1	Who knew a qui	denmarkincalifor	2025-7-13 19:45	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
9	DMAtrqlxXBF	DMAtrqlxXBF	DMAtrqlxXBF	GenAI meets bic	denmarkincalifor	2025-7-12 16:30	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DMAtrqlxXBF		
10	DL-HNCeBB0r	DL-HNCeBB0r	DL-HNCeBB0r	We asked our cc	denmarkincalifor	2025-7-11 16:15	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DL-HNCeBB0r		
11	DL4_RfMpcpN	DL4_RfMpcpN	DL4_RfMpcpN	Say hi to Sidsell	denmarkincalifor	2025-7-9 16:30 (	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
12	DLzz9y6BeDI	DLzz9y6BeDI	DLzz9y6BeDI	Love is love 🏳️ T	denmarkincalifor	2025-7-7 16:15 (	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
13	DLsGJ81Rqye	DLsGJ81Rqye	DLsGJ81Rqye	This week, we're	denmarkincalifor	2025-7-4 16:20 (	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
14	DLpikLGRfr-	DLpikLGRfr-	DLpikLGRfr-	Summer tip from	denmarkincalifor	2025-7-3 16:30 (	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DLpikLGRfr-		
15	DLm2loppuzI	DLm2loppuzI	DLm2loppuzI	Golden hour hits	denmarkincalifor	2025-7-2 15:25 :	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
16	DMVUBtlhL-5	DMVUBtlhL-5	DMVUBtlhL-5	What a Sunday!	denmarkincalifor	2025-7-1 16:00 (	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DMVUBtlhL-5		
17	DLXjRnlhJtq	DLXjRnlhJtq	DLXjRnlhJtq	It's vacay seaso	denmarkincalifor	2025-6-26 16:50	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DLXjRnlhJtq		
18	DLPw1vrp62U	DLPw1vrp62U	DLPw1vrp62U	Denmark is fallin	denmarkincalifor	2025-6-23 16:15	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DLPw1vrp62U		
19	DLNNwtnR-AW	DLNNwtnR-AW	DLNNwtnR-AW	Mist Trail in Yos	denmarkincalifor	2025-6-22 16:30	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
20	DLH_C1jB85r	DLH_C1jB85r	DLH_C1jB85r	Big congrats to (	denmarkincalifor	2025-6-20 15:45	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DLH_C1jB85r		
21	DLFdFYox0XL	DLFdFYox0XL	DLFdFYox0XL	We know, we kn	denmarkincalifor	2025-6-19 16:10	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
22	DLASV1OxADV	DLASV1OxADV	DLASV1OxADV	We couldn't help	denmarkincalifor	2025-6-16 19:18	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
23	DKzd1U_pJ8P	DKzd1U_pJ8P	DKzd1U_pJ8P	Six bright minds,	denmarkincalifor	2025-6-12 16:30	Denmark in Calif	https://scontent-	mixed	https://www.instagram.com/p/DKzd1U_pJ8P		
24	DKw3UARhF9b	DKw3UARhF9b	DKw3UARhF9b	Say hi to Sweet	denmarkincalifor	2025-6-11 16:15	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-
			uR9D6BTmH	The cross-Ameri	denmarkincalifor	2025-6-10 16:10	Denmark in Calif	https://scontent-	unknown	https://www.inst	https://scontent-	https://scontent-

Read docs.google.com

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.

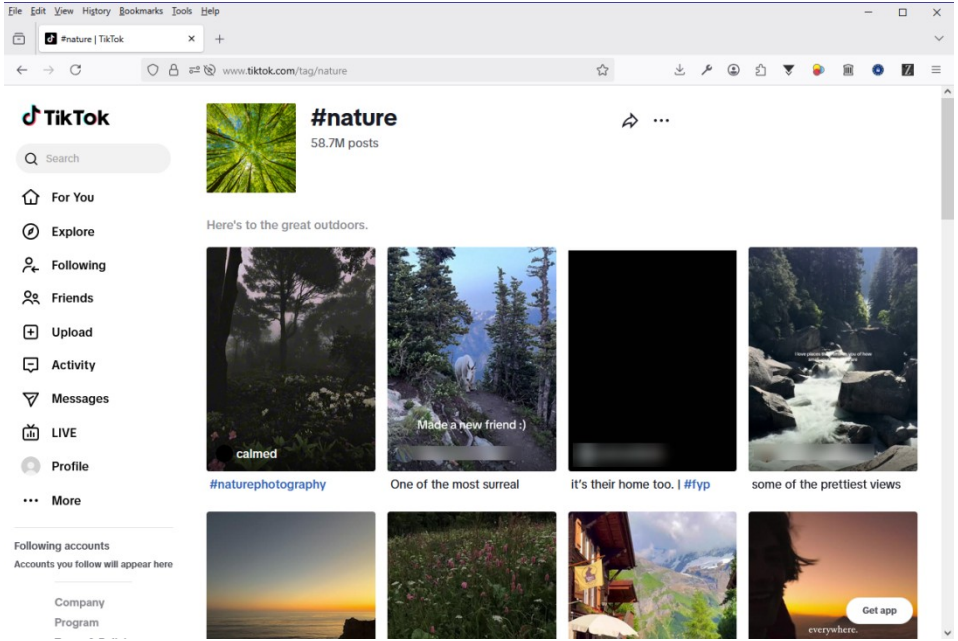
Harvesting TikTok Data

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

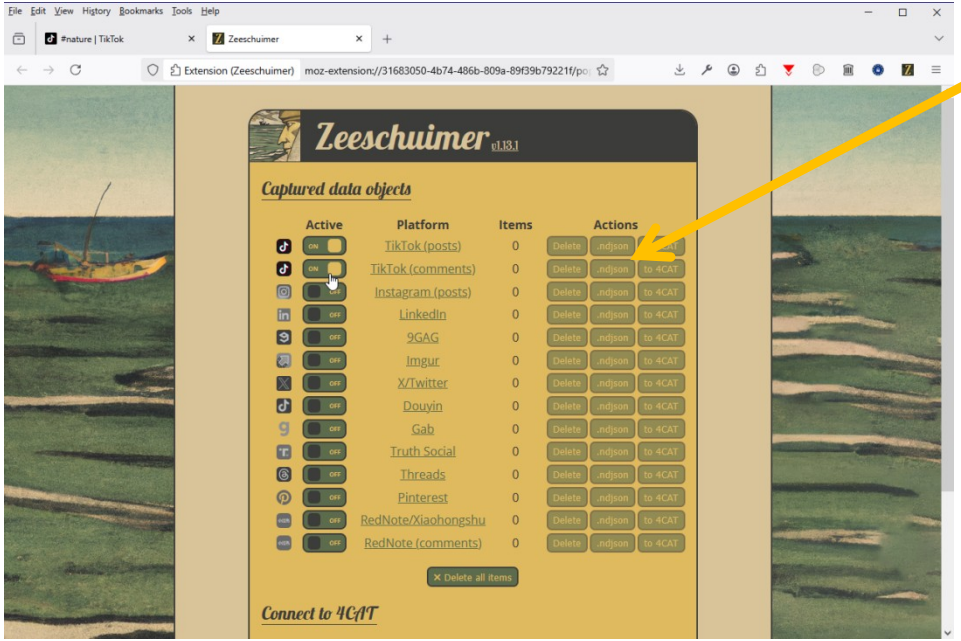


Remember the TikTok warning from slide 4

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

Harvesting TikTok Data

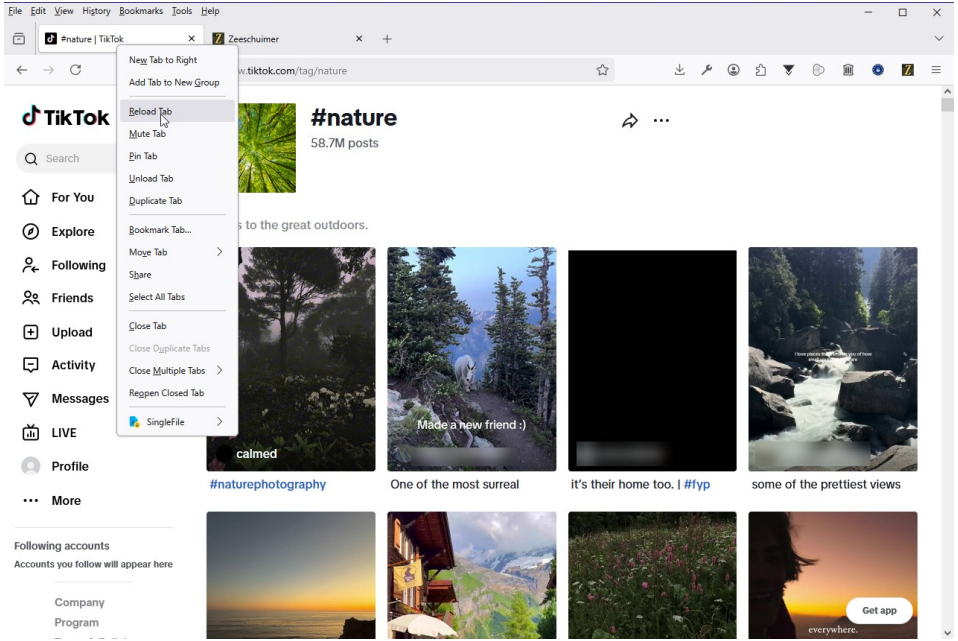
In Zeeschuimer both the harvest function for posts and for comments will be used, so both functions are activated here



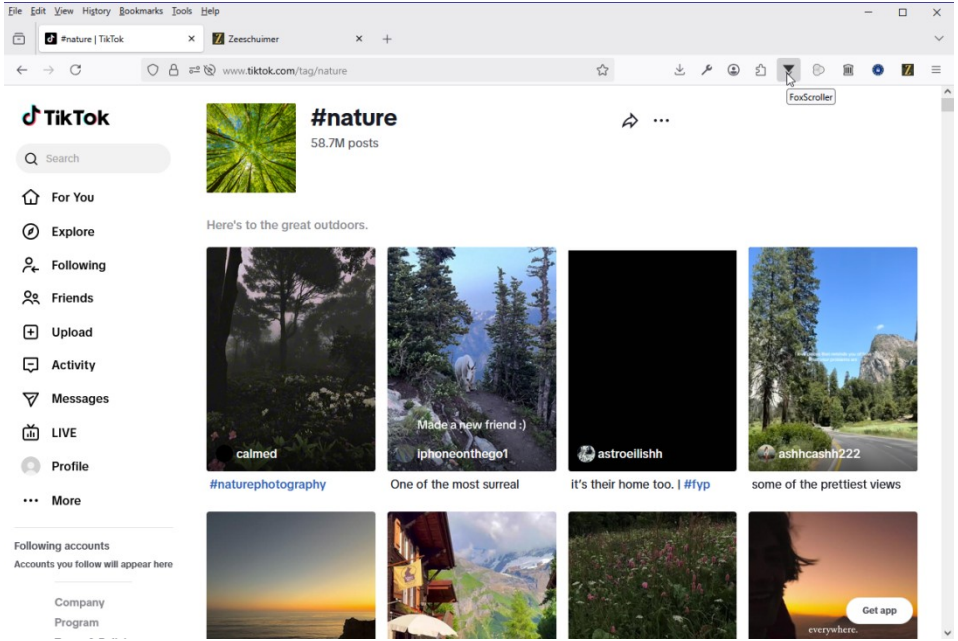
Two data sets will be generated, one for posts and one for comments

Harvesting TikTok Data

Go back to the TikTok page and reload it



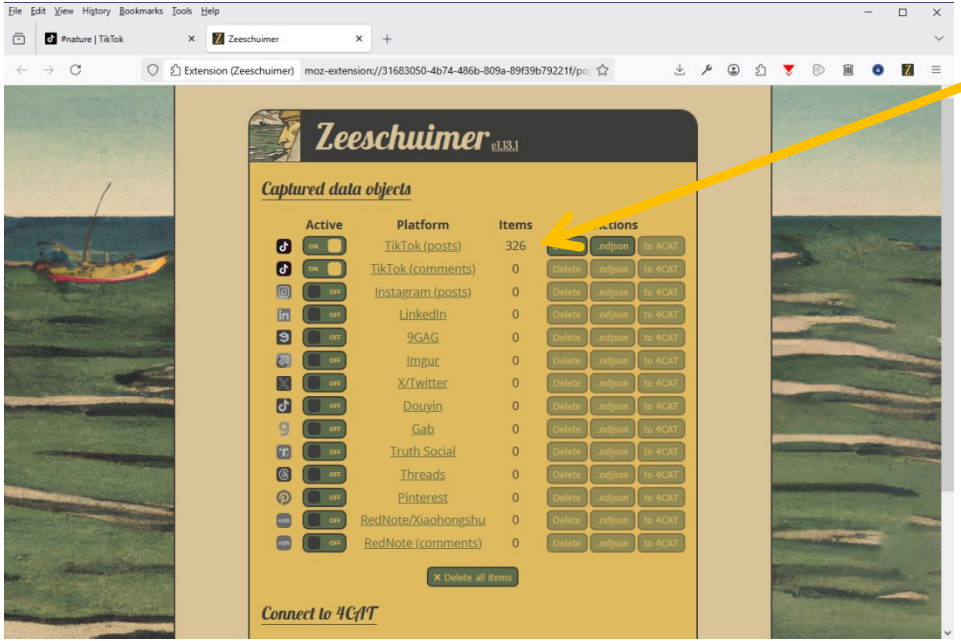
Harvesting TikTok Data



Start FoxScroller

Harvesting TikTok Data

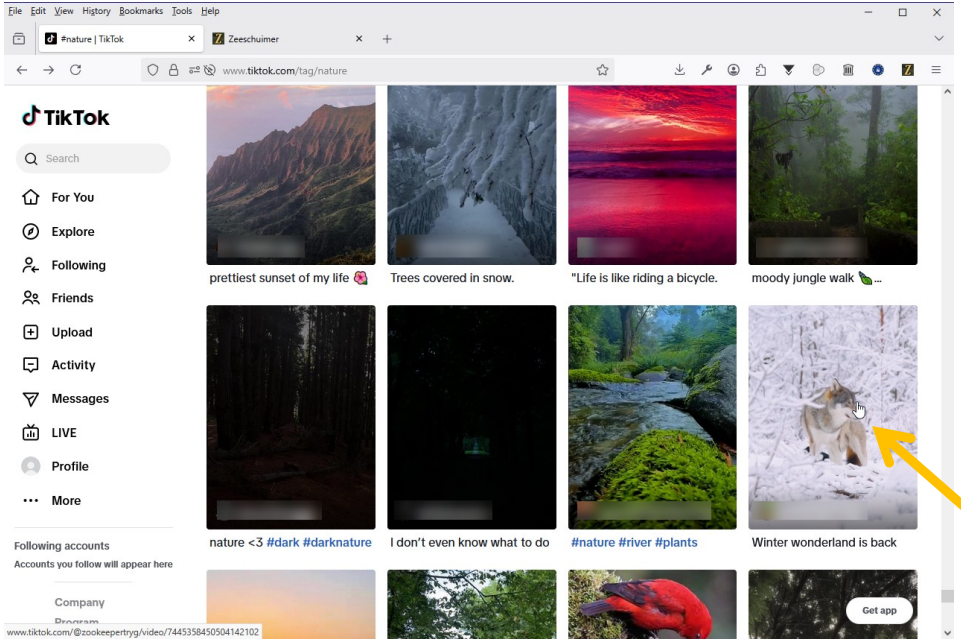
After FoxScroller has finished scrolling down the page with search results…



…data for 326 posts have been captured

Harvesting TikTok Data

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

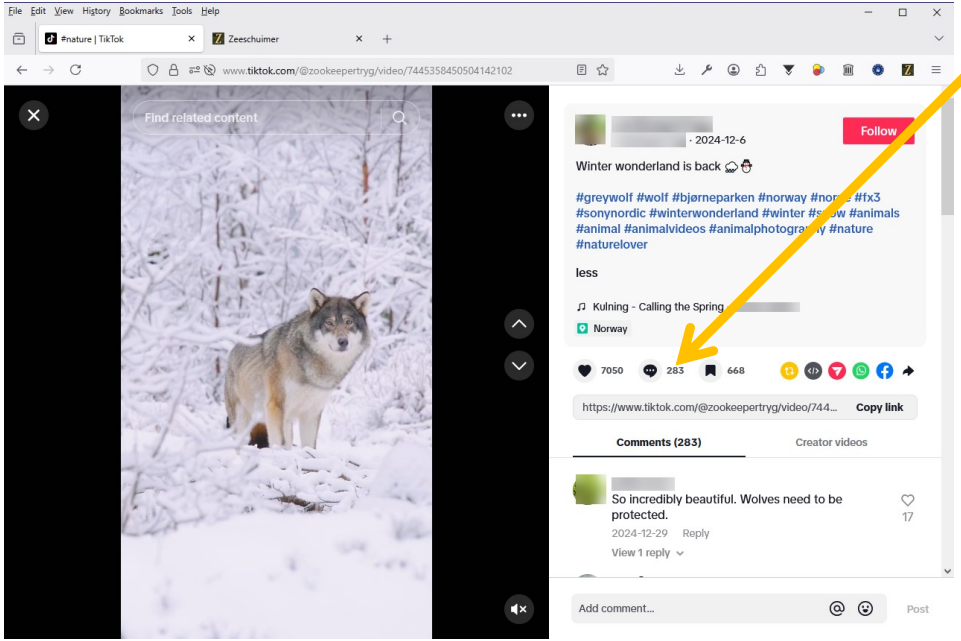


Unfortunately, post harvesting needs to be done manually, by entering the relevant posts one wishes to get the data from

A post is selected here

Harvesting TikTok Data

The selected post has been entered



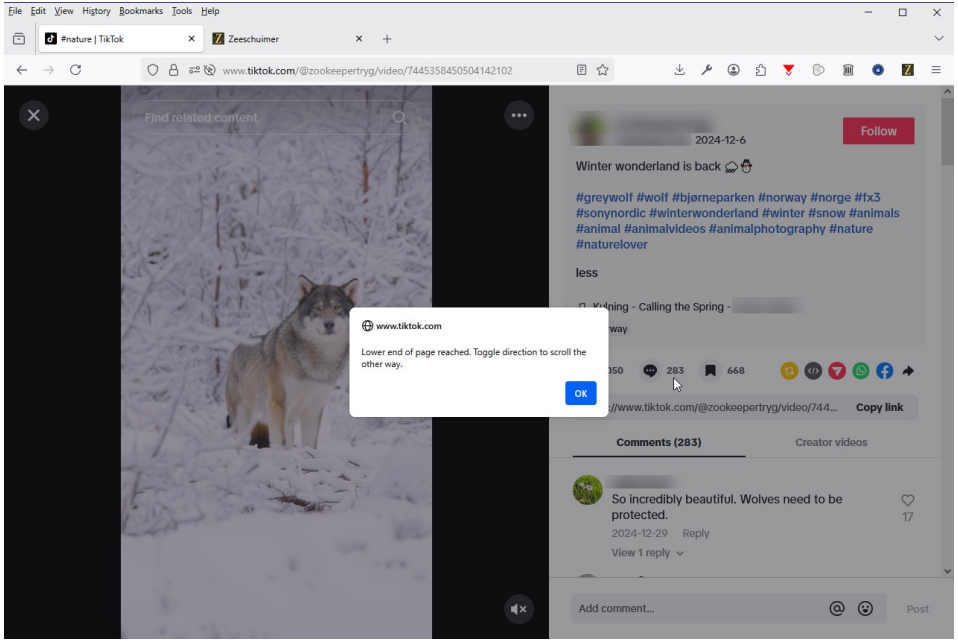
283 comments are listed

The count includes comments and replies

It also includes comments that have been removed again

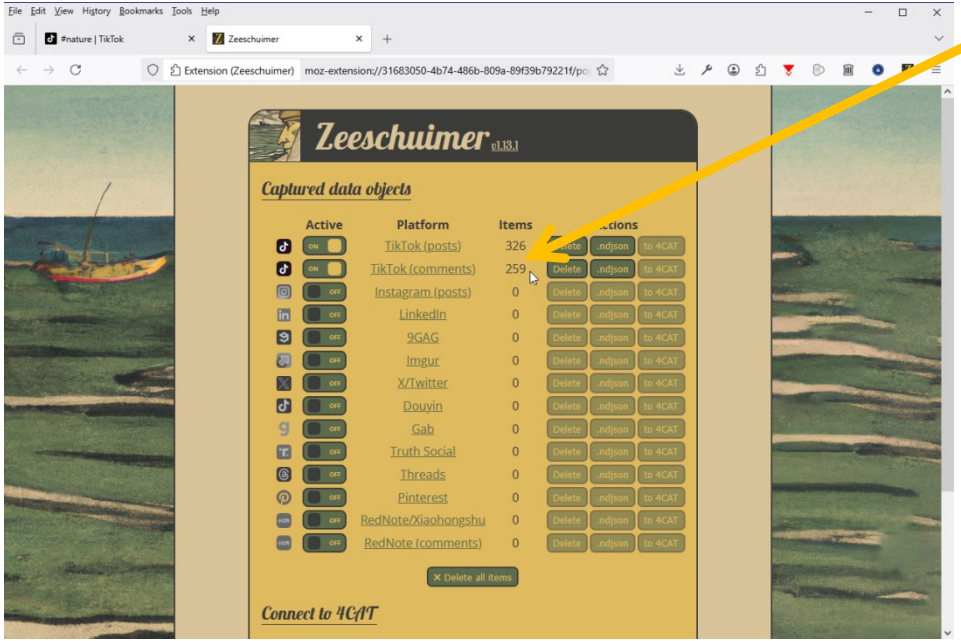
Harvesting TikTok Data

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

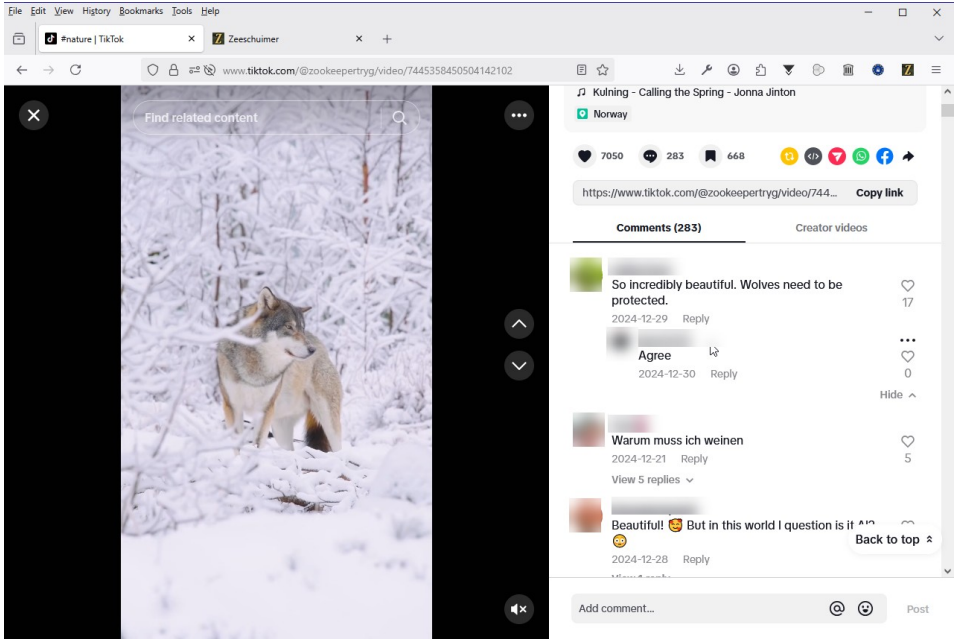
Harvesting TikTok Data



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

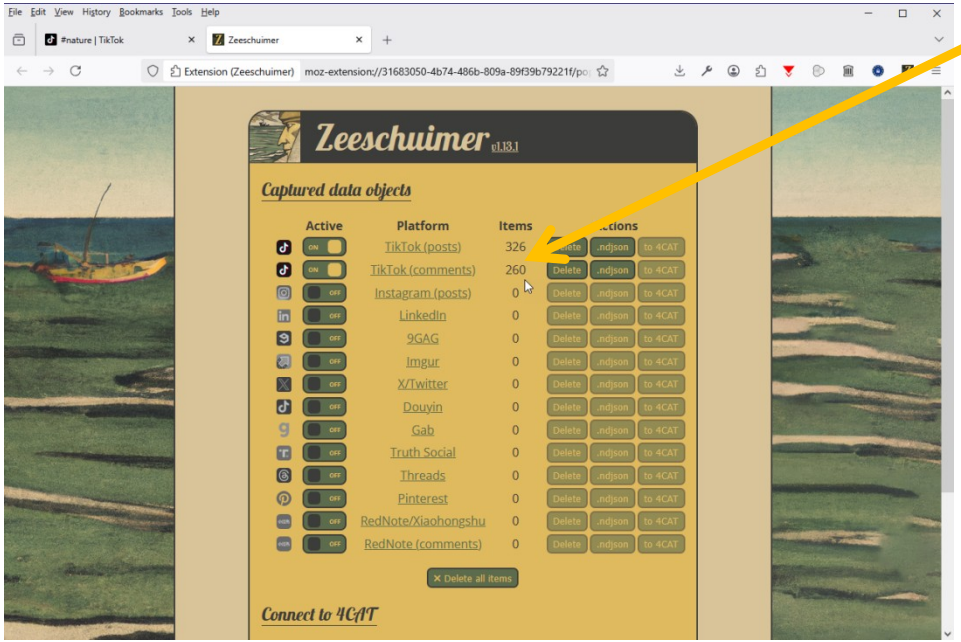
If replies need to be included in the data collection, they have to be opened manually

Harvesting TikTok Data



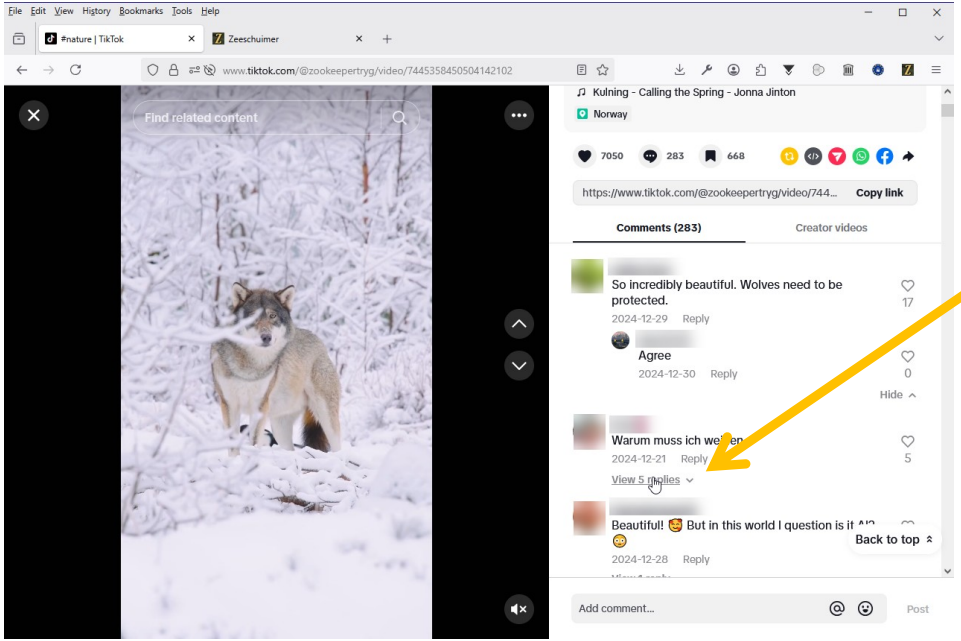
After going back to the top, a single reply has been opened

Harvesting TikTok Data



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

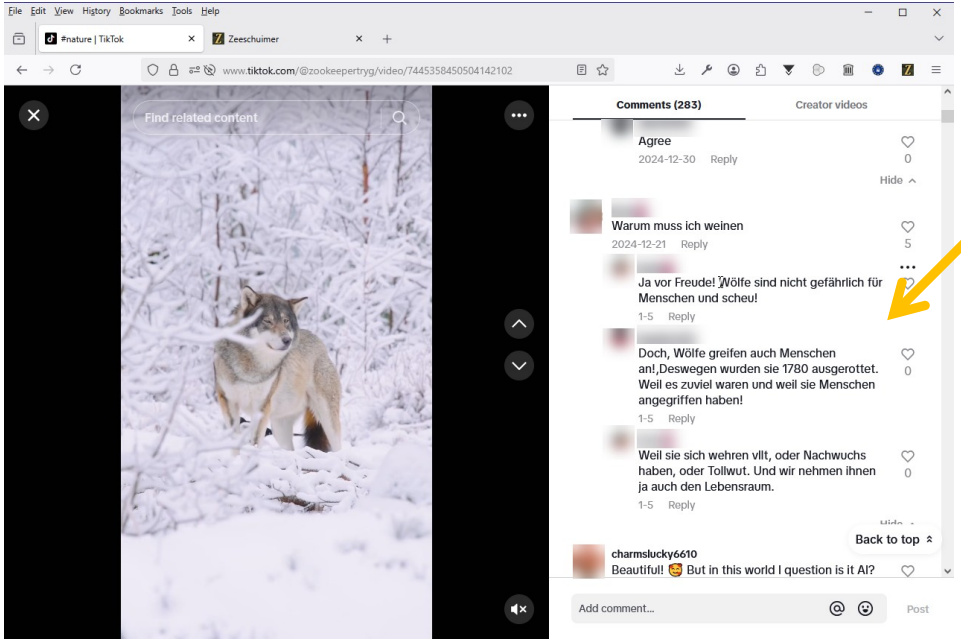
Harvesting TikTok Data



Here the replies to another comment are about to be opened

Note that five replies are listed

Harvesting TikTok Data

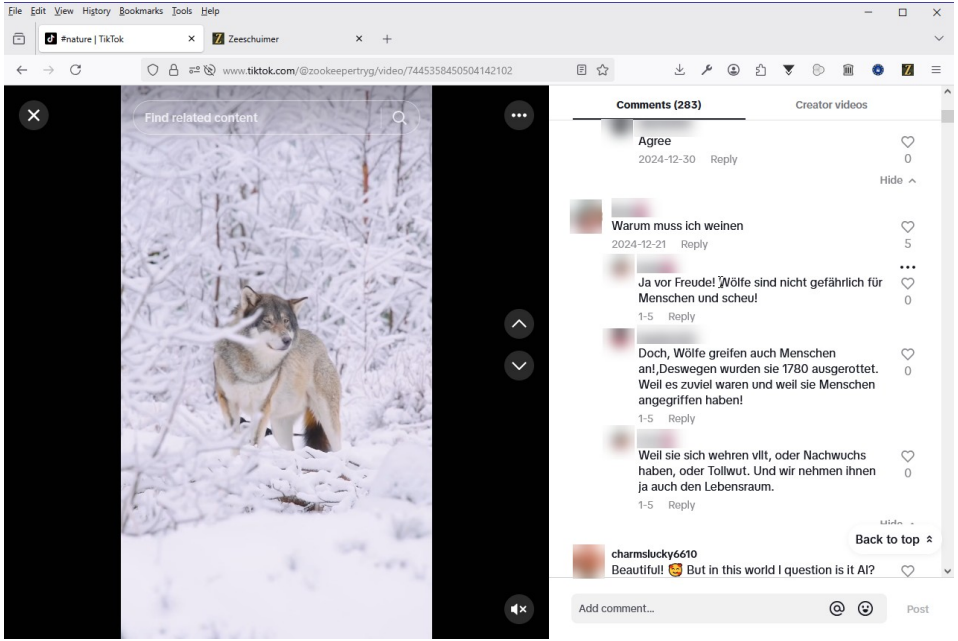


Five replies were listed, but there are only three. Two of the counted replies have thus been deleted again.

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