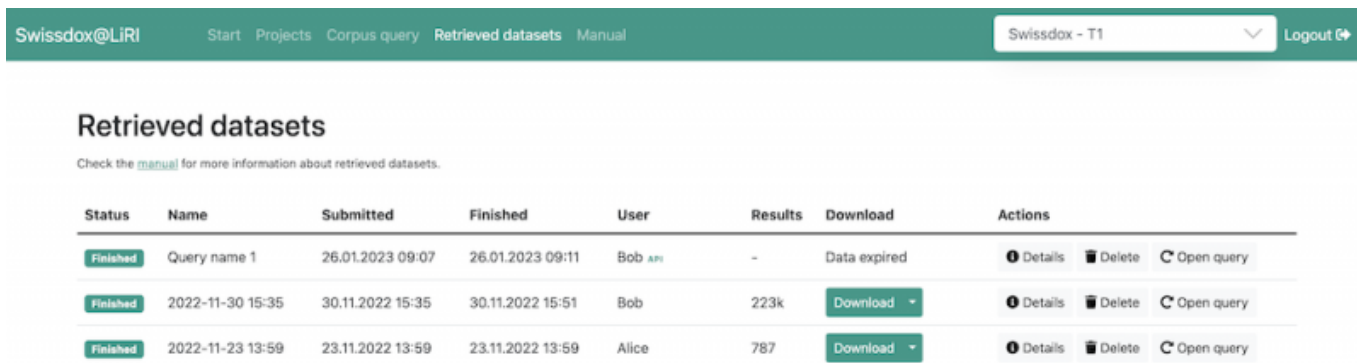


Compiled datasets

On the **Retrieved datasets** page, you will find a list of the queries that you and other project members have submitted to Swissdox@LiRI, along with additional information such as the number of retrieved articles per query.



Status	Name	Submitted	Finished	User	Results	Download	Actions
Finished	Query name 1	26.01.2023 09:07	26.01.2023 09:11	Bob <small>API</small>	-	Data expired	Details Delete Open query
Finished	2022-11-30 15:35	30.11.2022 15:35	30.11.2022 15:51	Bob	223k	Download	Details Delete Open query
Finished	2022-11-23 13:59	23.11.2022 13:59	23.11.2022 13:59	Alice	787	Download	Details Delete Open query

In general, the time it takes for a query to complete is directly proportional to the amount of data requested. As a result, queries with fewer results will typically complete more quickly.

Clicking on **Details** will display more information about the corresponding query. By selecting **Open query**, you can load the filters of a completed query into the query interface.

To download a dataset in a compressed [TSV \(tab-separated values\) format](#), use the **Download** icon on this page. The same result will be accomplished by clicking on the link sent in notification email when the dataset has been compiled and is ready for use.

Uncompressing an XZ file

To uncompress an XZ file (ending in `.xz`), Windows users can use programs like [7-Zip](#) or [Winzip](#). Mac users may prefer to use [The Unarchiver](#). In the Mac or Linux terminal, you can use the `tar` command to unpack a file. For example:

```
tar xvf filename.tsv.xz
```

It is usually not necessary to uncompress the files before processing them. The contents of the files can be extracted on the fly using commands such as `xzcat`:

```
xzcat filename.tsv.xz
```

Using the data programatically

Another option is to use a programming language, such as Python, to directly read from the compressed TSV file. Here is a Python snippet for this purpose:

```
import lzma

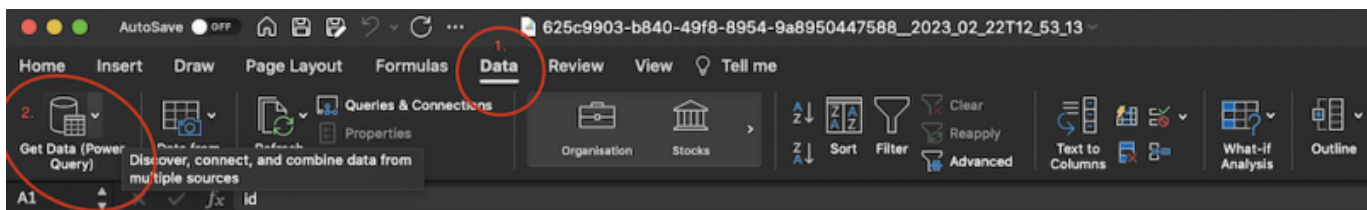
def read_xz_compressed_tsv(filepath):
    fh = lzma.open(filepath, mode='rt', encoding='utf-8')
    for line in fh:
```

```
if not line.strip() or line.startswith('#'):
    continue
yield line.rstrip().split('\t')

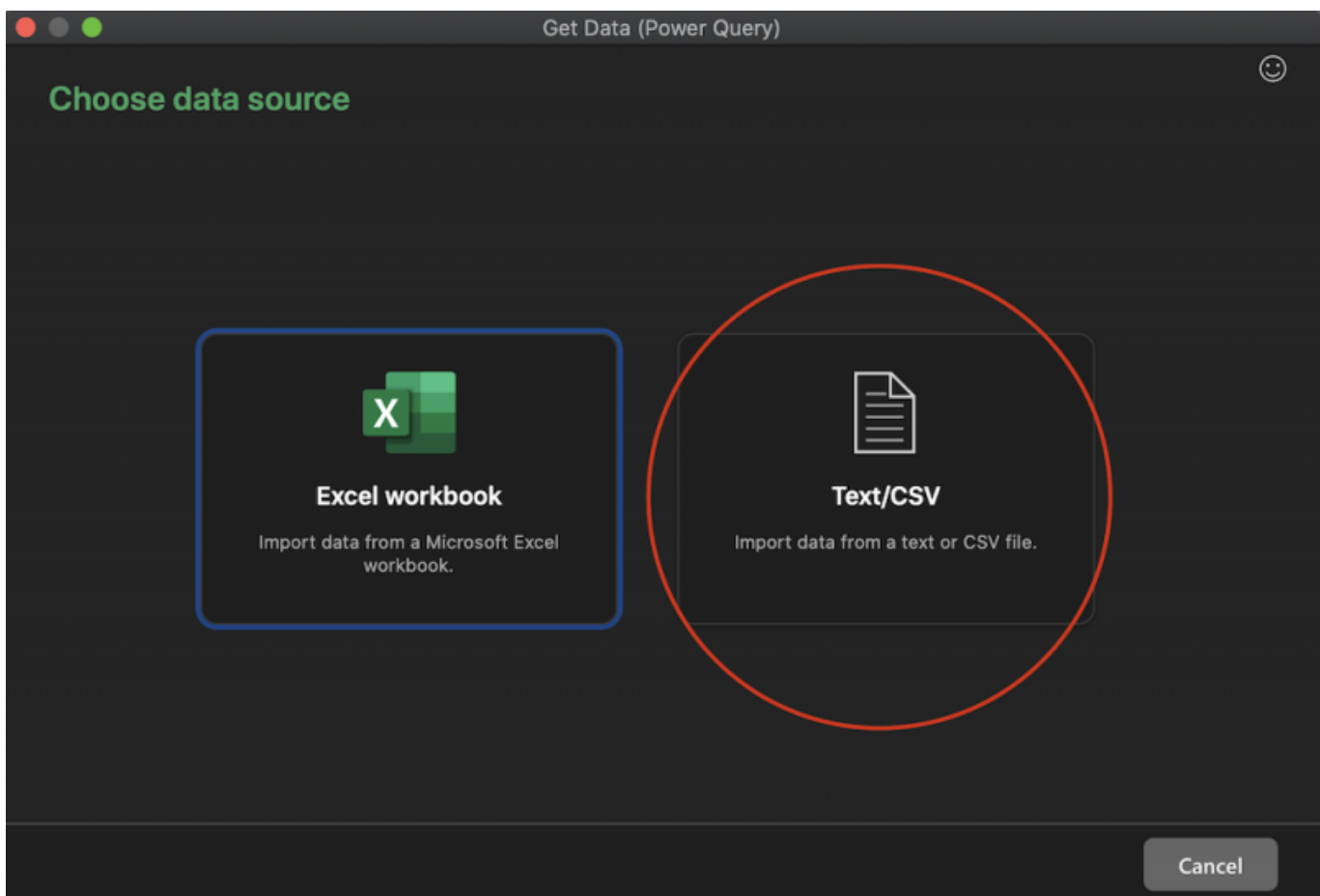
for row in read_xz_compressed_tsv('file.tsv.xz'):
    print(row)
```

Opening a TSV file in Excel

Opening a TSV file in Excel requires the use of the *import wizard*. To do this, click on the **Data** tab in the Excel navigation menu and select the **Get Data (Power Query)** option, as shown in the screenshot below.



Once the import wizard is open, select that you wish to import data from text.



After selecting the TSV file that you previously saved to your computer, you will get a screen which offers you the option to define encoding and delimiter for your data. Here, you should select UTF - 8 for encoding and Tab for the delimiter. Now you can successfully load your data into the Excel.

Get Data (Power Query)

Preview file data

File path: /Users/nikolina/Downloads/625c9903-b840-49f8-8954-9a8950447588_2023_02_22T12_53_13.tsv

File origin: 65001: Unicode (UTF-8) | Delimiter: Tab | Data type detection: Based on first 200 rows

id	pubtime	medium_c...	medium_na...	rubric	regio...	docty...	doctype_des
489601...	2/2/2023, 12:00:00 AM +01...	ZWA	20 minuten	People	Luzern	PRD	Regional daily
491088...	2/17/2023, 12:00:00 AM +0...	ZWA	20 minuten	People	Bern	PRD	Regional daily
489601...	2/2/2023, 12:00:00 AM +01...	ZWA	20 minuten	News	St. Gallen	PRD	Regional daily
489479...	2/1/2023, 12:00:00 AM +01...	ZWA	20 minuten	Zürich/Region	Zürich	PRD	Regional daily
489942...	2/6/2023, 12:00:00 AM +01...	ZWA	20 minuten	Zürich/Region	Zürich	PRD	Regional daily
490399...	2/10/2023, 12:00:00 AM +0...	ZWA	20 minuten	News	Luzern	PRD	Regional daily
490075...	2/7/2023, 12:00:00 AM +01...	ZWA	20 minuten	Zürich/Region	Zürich	PRD	Regional daily
490399...	2/10/2023, 12:00:00 AM +0...	ZWA	20 minuten	Ausland	Basel	PRD	Regional daily
490737...	2/14/2023, 12:00:00 AM +0...	ZWA	20 minuten	Zürich/Region	Zürich	PRD	Regional daily
489944...	2/6/2023, 12:00:00 AM +01...	ZWA	20 minuten	Sport	Zürich	PRD	Regional daily
49142019	2/21/2023, 12:00:00 AM +0...	ZWA	20 minuten	News	Luzern	PRD	Regional daily
491089...	2/17/2023, 12:00:00 AM +0...	ZWA	20 minuten	Das läuft	Zürich	PRD	Regional daily

Buttons: Back, Cancel, Load

From: <https://liri.linguistik.uzh.ch/wiki/> - LiRI Wiki

Permanent link: <https://liri.linguistik.uzh.ch/wiki/langtech/swissdox/datasets>

Last update: 2023/03/22 16:54

