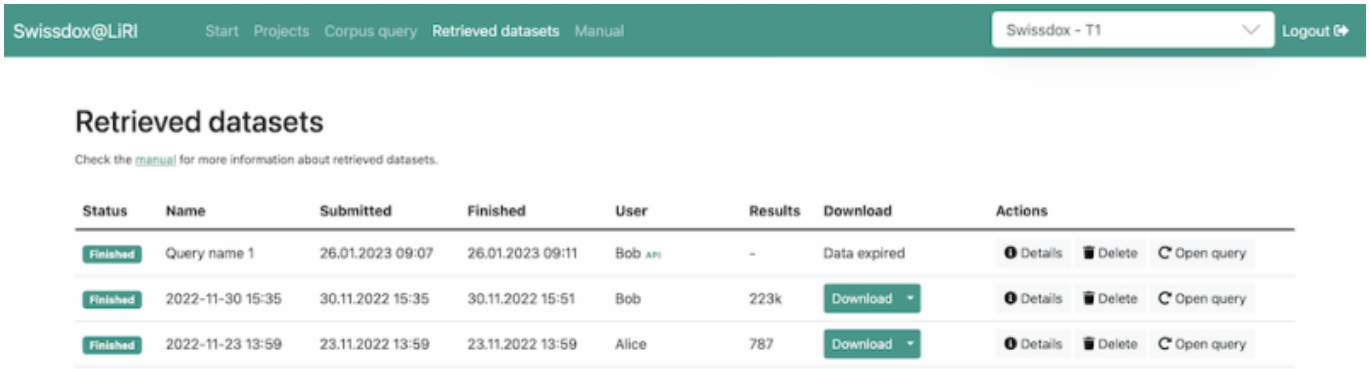
Compiled datasets

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.



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

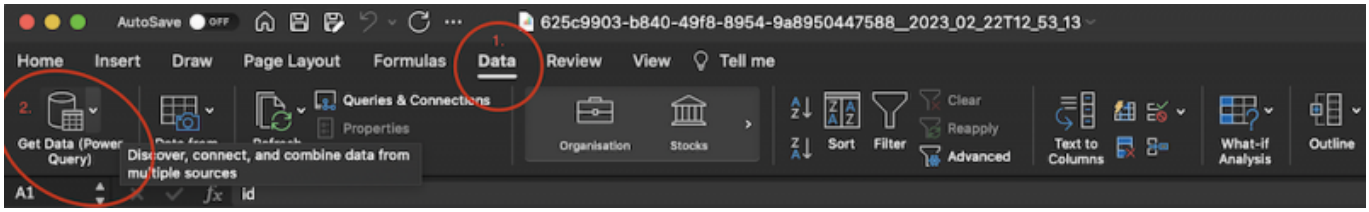
Another option is to use a programming language, such as Python, to directly read from the compressed TSV file. Here is a Python snippet for this purpose:

```
import lzma

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

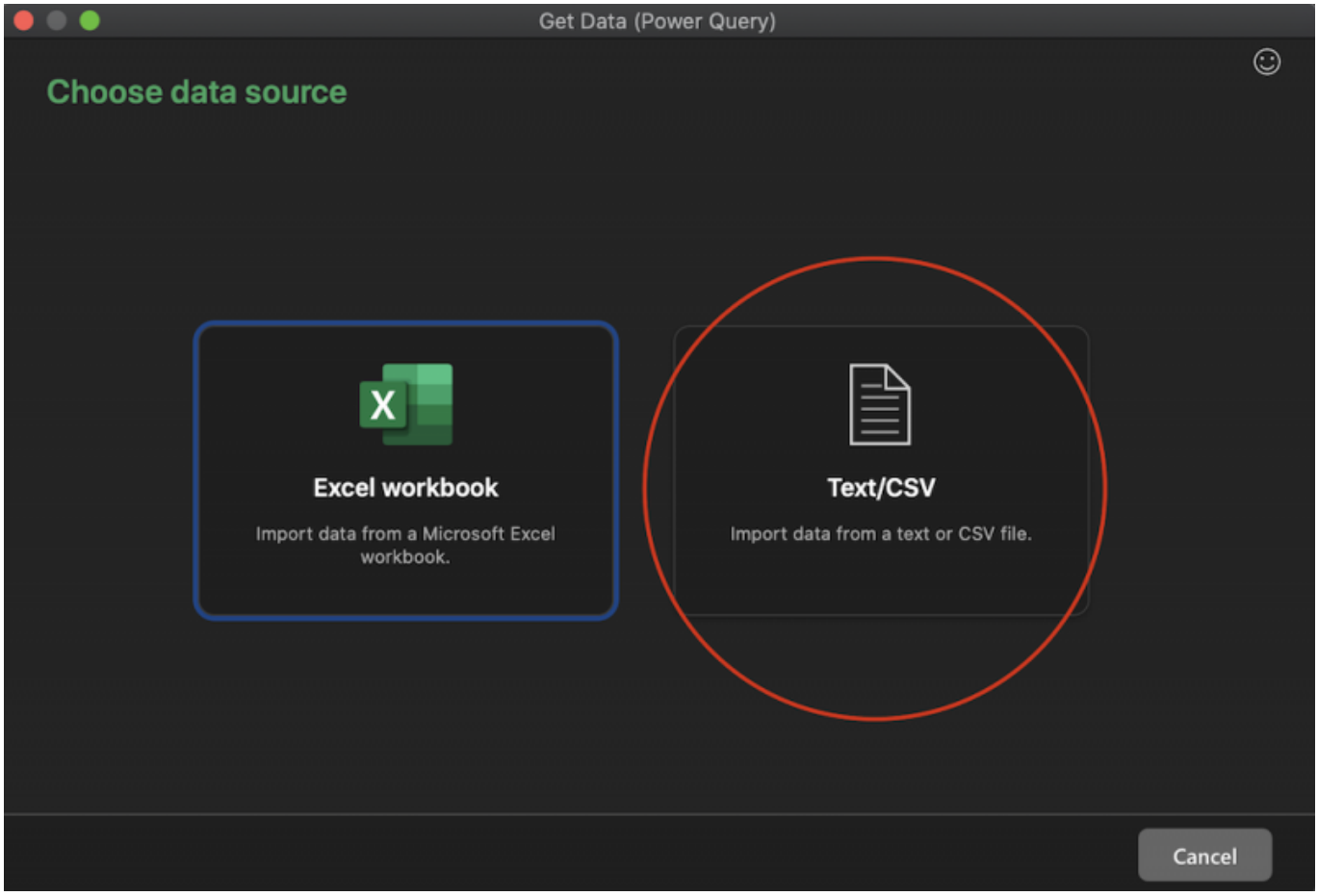
```
        if not line.strip() or line.startswith('#'):
            continue
        yield line.rstrip().split('\t')

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

Opening a TSV file in Excel

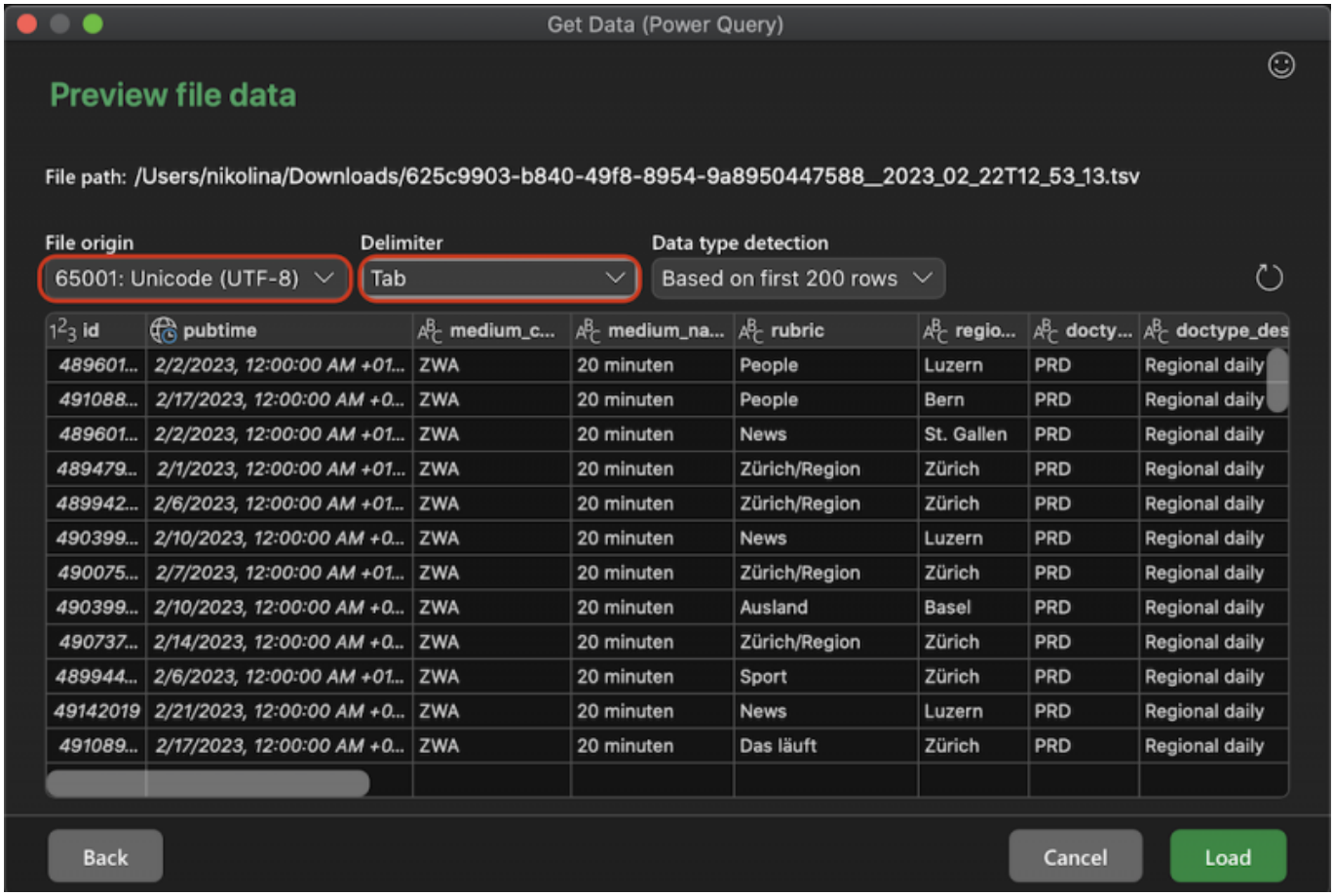
Opening a TSV file in Excel requieres the use of the *import wizard*. To do this, click on the **Data tab** in the Excel navigation menu and select the **Get Data (Power Query)** option, as shown in the screenshot below.



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.



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**