Overview - Extract Retrieval Data

Moritz Makowski, moritz.makowski@tum.de

Agenda

1. What is the goal of this "extraction"?
2. What happens during the filter process?
3. Where is the novelty?
4. How to access this data?
5. Background knowledge for software tools like this
6. How to use this tool
7. ~~What `calibrationDays` are & the complication with them~~
8. ~~How the automation works~~

1. What is the goal of this "extraction"?

https://retrieval.esm.ei.tum.de/hamburg/2021-08-31

```
## SENSOR LOCATIONS:
##    me17: GEO
##    mb86: HAW
##    mc15: JOR
##    md16: ROS
## LOCATION COORDINATES:
##             lng,     lat,     alt
##    GEO:   9.974,  53.568,  25.000
##    HAW:  10.200,  53.495,  41.000
##    JOR:   9.677,  53.536,   8.000
##    ROS:   9.892,  53.421,  98.000
##########################################
year_day_hour,      mc15_xco2_ppm_sc,   md16_xco2_ppm_sc, mc15_xch4_ppm_sc,   md16_xch4_ppm_sc
2021-08-1206:00:59, 412.02071425886874, NaN,              1.9041824076792346, NaN
2021-08-1206:02:59, 412.0075591602994,  NaN,              1.9042058983160552, NaN
2021-08-1206:04:59, 411.99274749780085, NaN,              1.9040060240929118, NaN
2021-08-1206:06:59, 411.9823147582188,  NaN,              1.9038608421148986, NaN
2021-08-1206:08:59, 411.95438482122177, NaN,              1.9035472063243717, NaN
2021-08-1206:10:59, 411.930338382756,   NaN,              1.9031644223403748, NaN
2021-08-1206:12:59, 411.9066550616497,  NaN,              1.903019380189071,  NaN
```

https://github.com/tum-esm/extract-retrieval-data/blob/main/docs/example-out.csv

2. What happens during the filter process?

1. **Calibrate** the raw measurement data
2. Filter out data where **GFIT flagged** some anomaly
3. Filter out any data according to specific **filter cases**
4. Compute a **rolling mean** over the remaining data
5. **Resample** the smooth curves at a given rate

Tweakable filter settings:

- `cases`
- `movingWindowSizeMinutes`
- `outputStepSizeMinutes`

An explanation of the filter cases can be found in the master thesis of Nico Nachtigall (NAS: `/tuei/esm/Thesis/Masterarbeiten/2020 MA Nico Nachtigall/Nachtigall_MasterThesis_final.pdf`)

3. Where is the novelty?

How it used to be: https://gitlab.lrz.de/esm/columnmeasurementautomation

(You should probably not try to understand this code)

Problem: One big pile of code that does everything: Triggering GFIT, loading data into the database, generating plots, generating CSV files, uploading data to the website.

Goal: Split this pipeline into two independent processes!

1. Fill the database with retrieval data
2. Use that database to generate certain output files

This project should implement a convenient way for you to **generate output files** from that database.

4. How to access this data?

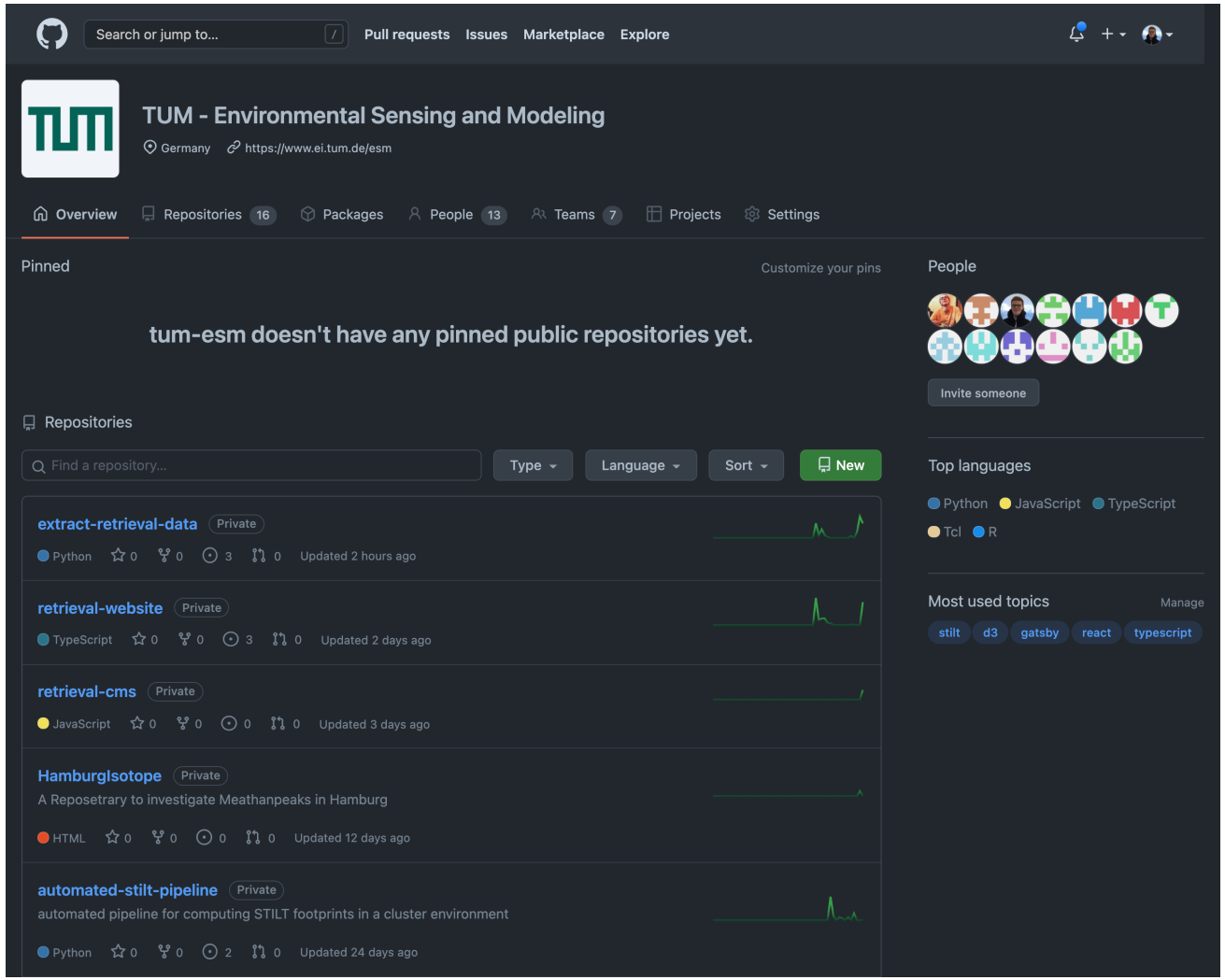
https://wiki.tum.de/display/esm/EM27+Retrievals

https://github.com/tum-esm/extract-retrieval-data

5. Background knowledge for software tools like this

- Git (https://wiki.tum.de/display/esm/Version+Control+with+Git)

- GitHub/GitLab (https://wiki.tum.de/display/esm/GitHub+and+GitLab)



```
12  scipy = "^1.7.1"
13  colorcet = "^2.0.6"
14  geographiclib = "^1.52"
15  sklearn = "^0.0"
16  matplotlib = "^3.4.2"
17  seaborn = "^0.11.1"
18  bokeh = "^2.3.3"
19  pyproj = "^3.1.0"
20  mysql-connector-python = "^8.0.26"
21  httpx = "^0.19.0"
22
23  [tool.poetry.dev-dependencies]
24  autopep8 = "^1.5.4"
25  pytest = "^6.2.4"
26  importlib-metadata = "^4.0.1"
27  typing-extensions = "^3.10.0"
28  black = "^21.6b0"
```

- Virtual Environments (https://wiki.tum.de/display/esm/Python+Development)

```
which python
# output: /usr/bin/python

source /Users/moritz/Documents/research/extract-retrieval-data/.venv/bin/activate

which python
```

6. How to use this tool

👉 Why this is not a workshop.

See setup instructions in the `README.md` at https://github.com/tum-esm/extract-retrieval-data.

`config.example.json` : https://github.com/tum-esm/extract-retrieval-data/blob/main/config.example.json

`example-out.csv` : https://github.com/tum-esm/extract-retrieval-data/blob/main/docs/example-out.csv

Overview - Extract Retrieval Data

Moritz Makowski, moritz.makowski@tum.de

7. What `calibrationDays` are & the complication with them

8. How the automation works