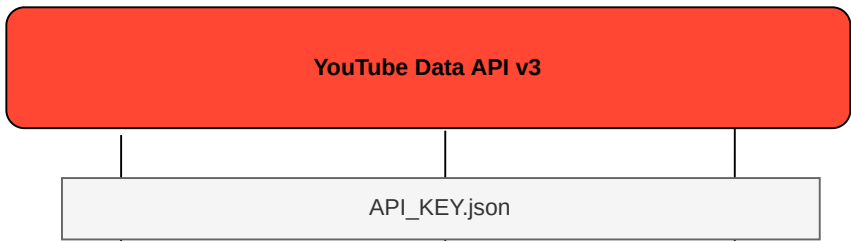


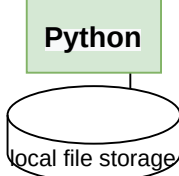
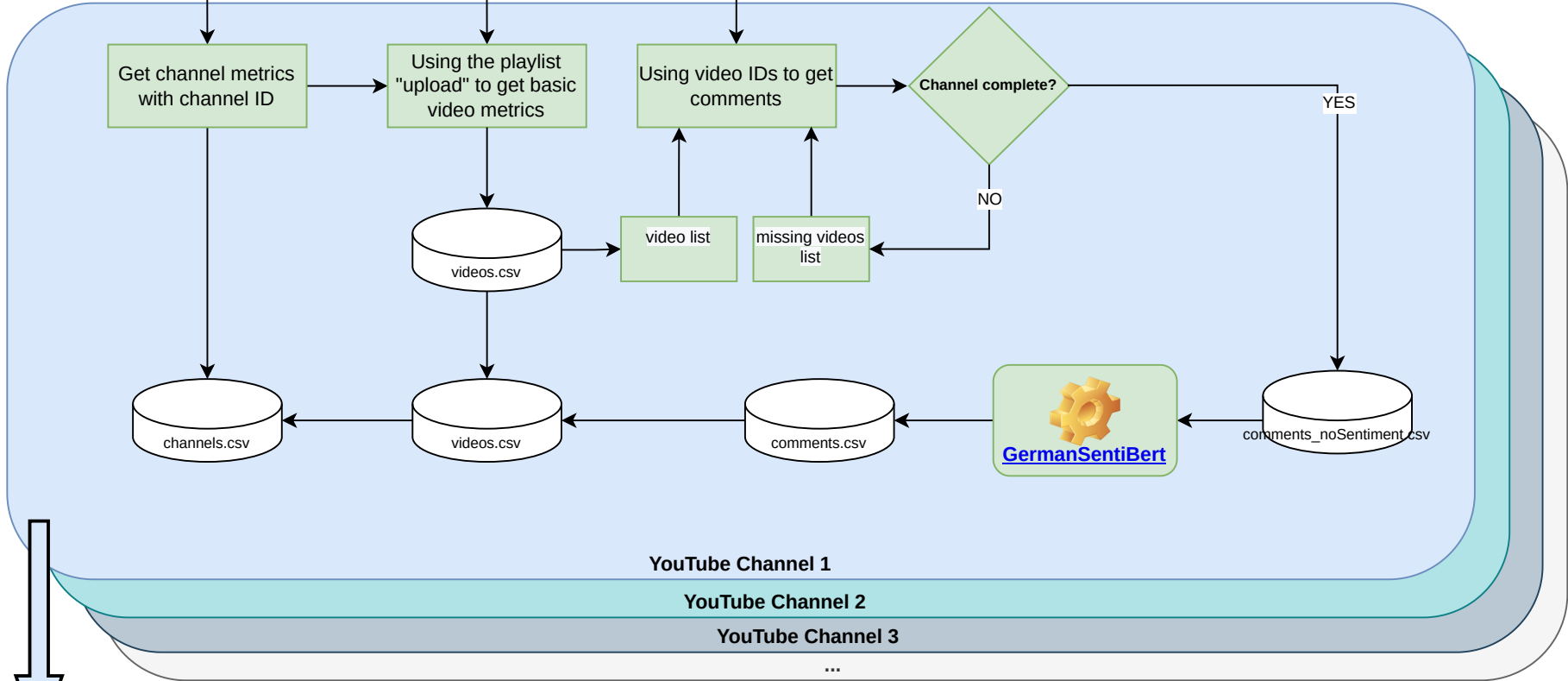
Data source

**API Authentication und limit**  
50 000 Quotas per key. Each requests usually costs 1 quota. Some request fetch packages of 50. For the 10 ZDF channels, this translates roughly to about ~1 million comments. The actual daily ZDF comment volume is about 10 000.

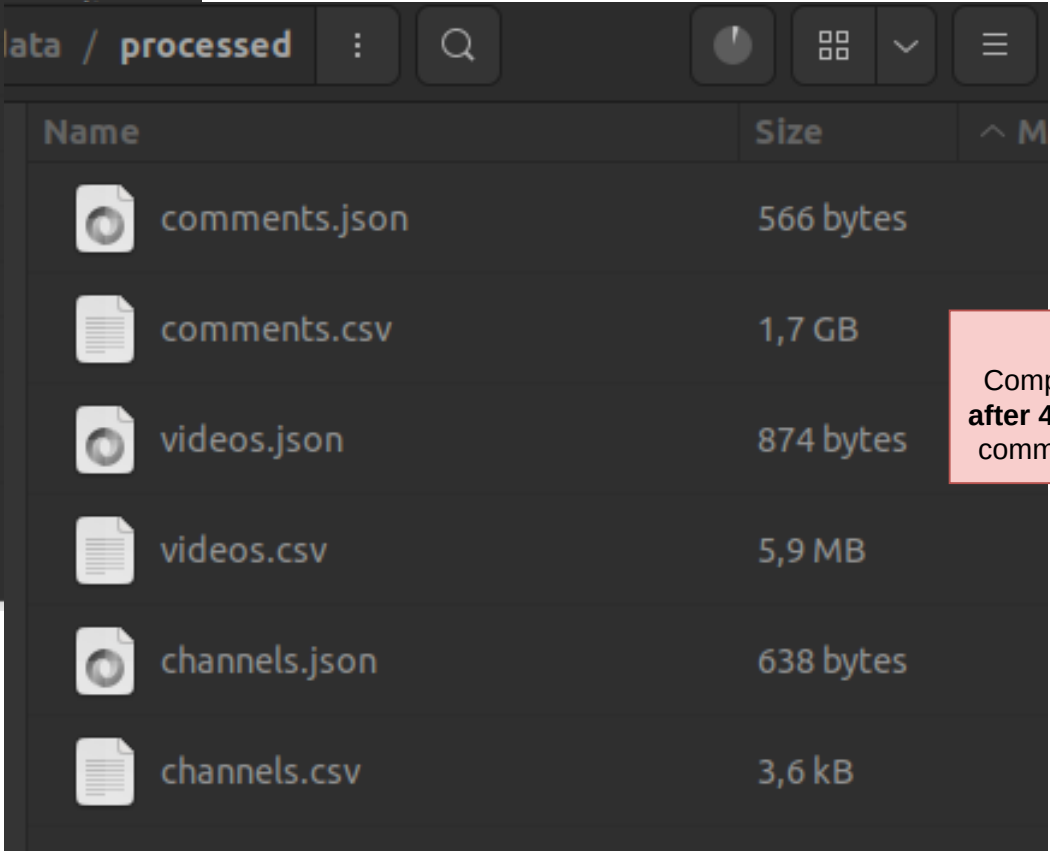
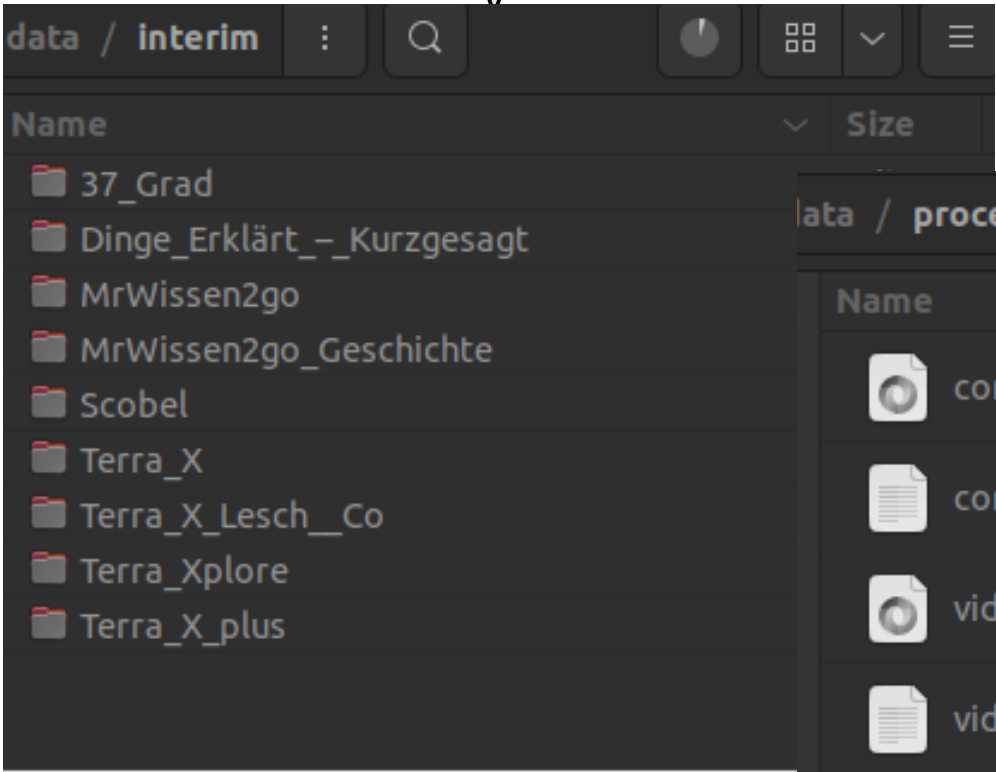


Ingestion/Transformation

- Currently **9 ZDF channels** but **more will probably come!** (other ZDF channels and/or competition)
- Current scripts are designed to fetches **all historically available** data for the desired YT-Channel (**full refresh**)
- Besides **sentiment analysis**, some data **cleaning, filtering, aggregating, simple calculations** (means, sums, medians)
- Whole pipeline executed with three scripts (fetch.py, sentiment\_analysis.py, transform.py) accessing functions in src/data.py



Storage

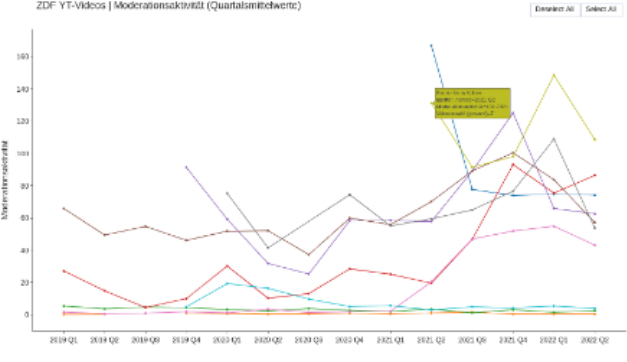


9 ZDF channels  
~3 Millionen comments  
~3000 Videos  
last data update January '23  
  
total data volume ~3 GB

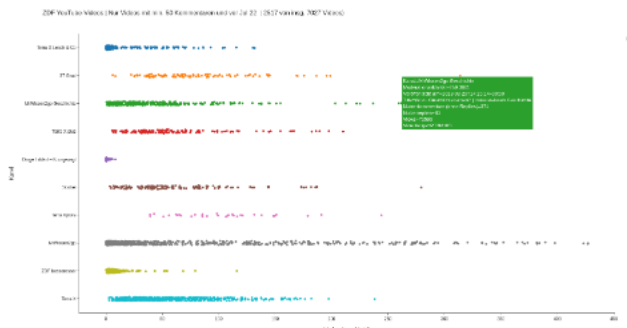
Update cycle?  
Complete data set refresh with **2 API Keys** **done** **after 4 days** ( ~3 000 000 comments / ~ 1 000 000 comments per day) + Analysis time for Sentiment

Dashboard reporting for YouTube supervisors

- Time series (quarterly resolution) of the six key metrics for each channel. For each metric an individual time series (currently done with plotly with "reports.py")



- "Strip plots" showing the distribution of (currently) six key metrics (for each channel and video). For each metric, one strip plot.



- Wordclouds using the annual word corpus for each channel (e.g. since "37°" exists since 2020, this would lead to three wordclouds (2020, 2021, everythin available for 2022)

Access for MeFo DataScience

SQL Access to raw data data warehouse for further analysis

UserStories MeFo Analyst
Als Analyst in der Medienforschung möchte ich einen (Lese-) Zugang zu den Rohdaten der YouTube Kommentardatensätze

UserStories YT-Content-Betreuer
Als YT-Content-Betreuer möchte ich die zeitliche Entwicklung von diversen Kommentarkennzahlen visuell und tabellarisch einsehen können (pro YT Kanal und min. pro Quartal)
Als YT-Content-Betreuer möchte ich die Verteilung von diversen Kommentarkennzahlen visuell und tabellarisch einsehen können (pro Video und pro Quartal)
Als YT-Content-Betreuer möchte ich Wordclouds pro Kanal und pro Jahr einsehen können