



UCPH Data Management Plan for Students

1.

DMP Details

| | |
|------------------------------------|---|
| Title of the project | Elon Musk Through the Lens of Reddit: Stance analysis of the discourse surrounding Elon Musk on selected subreddits |
| Name of the student | Martin William Grossman, ghj600@alumni.ku.dk |
| Name of the supervisor | Anders Blok & Jacob Aarup Dalsgaard |
| Expected start date of the project | 2025.01.31 |
| Expected end date of the project | 2025.05.31 |
| Short summary of the project | <i>Stance analysis of posts and comments about Elon Musk from Reddit. Data has been collected from subreddits of different political alignment to observe differences. Time frame is the 2024.03.13. to 2024.11.13.</i> |

2. Data description

Describe what data/material you will collect, observe, generate, create or reuse in the project.

a) Indicate the type(s) of data for each data set in your project.

Comprehensive collection of Reddit data organized by subreddits, originally collected by Pushshift (until April 2023) and subsequently by user u/raiderbdev, have been processed and made available by user u/Watchful1 via Academic Torrents.

b) Are there any personal data or confidential data in your project?

No, there are none.

c) Origin/source: where do the data/material come from?

Reddit is the original source. The data has been collected by a Reddit user and made publicly available.

d) Estimated size or volume: how much data/material will you work with?

Total size of the original dataset (pre filtering) is 50 GB, after filtering it was about 5 GB or about 100,000 rows in total.

e) Expected file format(s): in what format will your digital data be saved?

CSV and XLSX

3. Requirements, rights, and responsibilities

a) Are there any policies or guidelines you will have to adhere to in your project? If so, please list them here.

| |
|------|
| None |
|------|

b) Will you reuse already existing data/material? Yes ☐ No ☐

If yes, are there any rules for what you can and cannot do with the data? Summarize the rules in brief.

No

c) Will you produce your own data/material? Yes ☐ No ☐

If yes, describe what you and others can and cannot do with the data/material.

No

4. Ethical and legal approvals

a) Does your project require **ethical** approval? Yes ☐ No ☐

If yes, explain why and what approvals need to be in place before you can start. Who has collected, or will collect, this approval?

No ethical approval needed.

b) Does your project require **legal** approval? Yes ☐ No ☐

If yes, explain why and what approvals need to be in place before you can start. Who has collected, or will collect, this approval?

No legal approval needed.

5. Data documentation

a) In short, describe your methods for collecting and processing the data/material. If a detailed methods description already exists, you can refer to it instead.

The data was downloaded in zst files from Academic Torrents. Detailed description of it is in the Methodology section of my thesis.

b) Describe how you will organize and structure your digital data/material, how you will name your files and keep track of different file versions.

The data was downloaded by subreddits in different files for posts and comments and processed and saved with a name showing the subreddit and if it contains posts or comments. Later they were merged into one file called "combined" and 900 rows were taken into another file called "sample".

c) Describe how you will document your project and the data. What information will you record about your data and material to ensure that your project/data/material is understandable to others?

The data in its original form, the processing steps taken and the results are detailed in the Methodology section of my thesis.

6. Storage and information security

a) Describe where will store your data and material.

The data was stored locally on my PC.

b) Describe what you will do to make sure your data/material are protected against loss, theft, unauthorized modification, and unauthorized access.

The data is on my PC and not available to anyone.

7. Sharing data openly

a) Will any of the data/material in the project be shared openly with others? Yes ☐ No ☐

If yes, describe which datasets.

The data will not be shared with anyone (it is already publicly available) but my thesis detailing it will be available publicly.

b) If yes, how will the data/material be shared openly? Address what repository you expect to use and what documentation will be sent along with the data/material.

Not applicable.

8. Preservation

a) Describe what data/material/project documentation should be kept once your project is over.

The data used in my thesis will not be kept. The text of my thesis and the appendices (including the code used for processing) is available on GitHub.

b) Describe where the data/material/project documentation will be stored after project end, and how a copy of the data will be made available to your supervisor(s).

The data will be made available to my supervisor upon request from my personal, local drive.