



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

Commented [DMG1]: Personal data are data that can directly or indirectly identify a person. Personal data can be divided into:

- Non-sensitive personal data, such as CV, address, date of birth, tax records.
- Information about criminal matters, such as convictions or an address in prison.
- Sensitive personal data, such as race, political views, health information, sexual orientation.

Confidential data are data other than personal data to which only a limited number of people should have access, and where accidental or deliberate exposure of the data can have considerable consequences. Examples are:

- Company data (trade secrets)
- Data that have commercial potential
- Classified government data
- Sensitive biological data

Commented [DMG2]: For example, 4GB of digital data, a corpus of 5 text documents, 25 interviews, 65 biological samples, etc.

Commented [DMG3]: For example, plain text, XML, MP3, JPEG, CSV, etc.

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

Commented [SDB4]: Examples that could be relevant for students:

- *The UCPH policy for research data management.* All students at UCPH who start their Bachelor or Master project should adhere to this policy.
- *The General Data Protection Regulation*, if you work with personal data
- *Collection permits or export permits*, for example when you collect and/or export biological samples in/from other countries.
- *Guidelines at the place where you carry out your project.*

Perhaps the research group/company/... has some specific rules for how to manage data.

- *Policies or guidelines determining terms of data use by institutions that provide data.* For example, Statistics Denmark, the Danish National Archives, online databases.

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.

Commented [SDB5]: Does the person/institution/database etc that provides the data pose any rules for data management?

Here you can for example consider:

- Who can have access to the data and work with the data during the project?
- In which timeframe (for how long) can they have access to the data?
- Can the material / data (or a copy of the data) be kept after the project?
- Who can keep (a copy of) the data after the project?
- Can the data be shared with others who are not part of the project, during and after the project?
- Can the data be used for research publications?

Commented [SDB6]: Make sure you agree with your supervisor and other project members (if any) on this.

Consider, for example:

- Who can have access to the data in your project, both during and after the project?
- Can the data be kept after the end of the project or do they need to be deleted?
- Who can keep (a copy of) the data after the project?
- Can the data be used for research publications? And if so, who can be author on these publications?
- Can the data be reused in other (new) projects?
- Who can reuse the data in other projects?

Commented [DMG7]: Relevant to ensure that measures are in place to safeguard the rights, well-being and dignity of human participants involved in the project, or to ensure adherence to (international) guidelines for animal welfare.

Examples of research where ethical approvals may be required are:

- *Clinical trials*, for example, studies where human subjects are involved to test medical treatments.
- *Psychological research*, for example, studies where human behaviour is explored in response to interventions.

Commented [DMG8]: Relevant to ensure compliance with laws and regulations. Examples of projects where legal approvals are required are:

- *Clinical studies testing new pharmaceuticals or medical devices*. Here, approvals are required from the Danish Medicines Agency (Lægemiddelstyrelsen) for compliance with Danish regulations and the EU Clinical Trials Regulation.
- *Research involving environmental samples or studies in protected areas*. These projects may require legal approval.

Commented [SDB9]: For example, here you can describe your folder and file structure (the file path).

A file naming convention is a systematic and standardized approach to name files consistently, for example "date_keywords_author initials_version.format".

To keep track of different versions, you could for example:

- 1) Always include a version number in your file names
- 2) Add a description of the different versions made in the document (e.g. a table).

Commented [SDB10]: How will you document your project?

Will you use a (laboratory) notebook, a codebook, digital text files, databases, specific software, or other ways, to record information about your project and your data?

What will you as a minimum write down?

Think for example about:

- The research question you are addressing
- The project plan / methods / protocol
- The equipment or software used to collect or process the data

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.

Commented [SDB11]: Please note that the two questions in this block are connected to your data type descriptions under block 2; if you work with personal or confidential data, you will have to have more focus on the security compared to when you work with normal data.

Please note that you are being asked about the storage solution for

- 1) Digital files (what IT tool)
- 2) Physical materials (what storage location)

Commented [DMG12]: Consider:

1. Technological measures.

Besides the IT storage tool you have mentioned under 6a, you could e.g. consider back-up strategy, password protection, encryption.

2. Physical measures

What physical measures do you have in place to protect the storage location of physical material and digital files? For example, locks on doors and storage cupboards, privacy filters on laptop screens, climate control in rooms storing vulnerable artifacts, shredding paper files containing personal data when they are no longer needed etc.

3. Procedural/operational measures

How can you set up your project and design your methods to improve security? For example, only allowing specific people access to the data instead of the whole team when you collaborate with multiple people, anonymising data asap when working with personal information, scrutinizing the security of any software tools or services you employ to manage data.

Commented [SDB13]: You will have to assess what data to keep and what data to discard when the project has ended. Also make the same assessment for other research materials, such as protocols, project plans, interview guides, code and physical samples.

Perhaps not all material can or should be kept, e.g. when material quality deteriorates over time, or there is no space to store the material indefinitely. There may be no need to keep all the digital data in your project, e.g. the data that you collected when testing a method. On the other hand, it may be important to keep some other data, e.g. the data that underlie the results described in your thesis, or the data that could be relevant for future students.

Describe your conclusions here and make sure that your supervisor agrees.

Commented [SDB14]: According to the UCPH Policy for Research Data Management, a copy of digital data sets needs to be made available to supervisors at project end.

Describe how you will do that, and where you intend to store any other copies of the data that should be preserved after project end.

Preserving data only makes sense, if the data can be found and understood at a later date. Therefore also make sure to describe what types of documentation that explain the data you will store along with the data.

Please note that your answers to the questions under 7 and 8 can be somewhat identical if

- you intend to share all the data you will preserve.
- the solution you share your data with also ensures stable data storage for at least 5 years.

Your answers to 7 and 8 will be different, if

- you do not share all the data you will preserve, or you share different data sets than the data sets you will preserve.
- you use an archiving solution to store data that is not suitable for sharing.