



Factors that influence people to watch fitness videos



**Horizon Europe
Data Management Plan**

02 March 2023



History of changes

There are no named versions.

Contributors

The following contributors are related to the project of this DMP:

- Yang Li

ly13922160914@163.com

Roles: Contact Person, Data Collector, Data Curator, Researcher

Projects

We will be working on the following projects and for those are the data and work described in this DMP.

Factors that influence people to watch fitness videos

Acronym

WFV

Start date

2023-02-28

End date

2023-03-31

Funding

Did not apply for any funding yet.

The 30 most viewed fitness videos posted on youtube were selected and the 50 most liked comments below the videos were collected using the youtube tool. Then analyze the high frequency words in these comments in python, and the sentiment scores of these words. Analyze the content that people commented on the most after watching these youtube fitness videos and the sentiment scores of the content. To understand the factors that influence people to watch fitness videos.

1. Data Summary

Data formats and types

We will be using the following data formats and types:

- **CSV Dialect Description Format**

It is a standardized format. This is a suitable format for long-term archiving.

We will have only a small amount of data stored in this format.

2. FAIR Data

2.1. Making data findable, including provisions for metadata

- **Factors that influence people to watch fitness videos** (published)

The dataset has the following identifiers:

- URL: https://github.com/yangli171/Collectingdata_resit.git

We will distribute the dataset using:

- *Domain-specific repository*: [GitHub](#). We are going to contact the repository.

A persistent identifier will be assigned by the repository. The repository will make sure that the persistent identifier can be resolved to a digital object. The assigned persistent identifier is specified: https://github.com/yangli171/Collectingdata_resit.git.

There won't be different versions of this data over time.

We will not be adding a reference to any data catalogue because the data will be stored in a repository that is the prime source of data for re-use in the field.

2.2. Making data accessible

We will be working with the philosophy *as open as possible* for our data.

All of our data can become completely open immediately.

Limited embargo will not be used as all data will be opened immediately.

Metadata will be openly available including instructions how to get access to the data. Metadata will be available in a form that can be harvested and indexed.

Our data is legally not copyrightable, there is no legal owner.

For our produced data, conditions are as follows:

- **Factors that influence people to watch fitness videos** (published)

The distributions will be accessible through:

- *Domain-specific repository*: [GitHub](#). We are going to contact the repository.

A user of this data can use it without any specific software.

The dataset will be published after all our processing has finished.

2.3. Making data interoperable

We will be using the following data formats and types:

- **CSV Dialect Description Format**

It is a standardized format.

We will be using the following standards (encodings, terminologies, vocabularies, ontologies):

- **CSV Dialect Description Format**

2.4. Increase data re-use

The metadata for our produced data will be kept as follows:

- **Factors that influence people to watch fitness videos** (published) — This data set will be kept available as long as technically possible. — The metadata will be available even when the data no longer exists.

As stated already in Section 2.2, all of our data can become completely open immediately.

We plan to be archiving data (using so-called *cold storage*) for long term preservation already during the project.



To validate the integrity of the results, the following will be done:

- We will be instrumenting the tools into pipelines and workflows using automated tools.
- We will run part of the data set repeatedly to catch unexpected changes in results.

3. Other research outputs

We use Data Stewardship Wizard for planning our data management and creating this DMP. The management and planning of other research outputs is done separately and is included as appendix to this DMP. Still, we benefit from data stewardship guidance (e.g. FAIR principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

4. Allocation of resources

FAIR is a central part of our data management; it is considered at every decision in our data management plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible. Making our data FAIR is therefore not a cost that can be separated from the rest of the project.

None of the used repositories charge for their services.

Yang Li is responsible for reviewing, enhancing, cleaning, or standardizing metadata and the associated data submitted for storage, use and maintenance within a data centre or repository.

Yang Li is responsible for finding, gathering, and collecting data.

To execute the DMP, additional specialist expertise is required and we have such trained support staff available.

We require the following hardware or software in addition to what is usually available in the institute: First, we will use youtube tools to collect the comments below the fitness videos posted on youtube. Then we will use python to tokenize the comments and then use library to get the wordcloud and sentiment scores of the comments.

5. Data security

Project members will not store data or software on computers in the lab or external hard drives connected to those computers. They will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centers where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (<https://...>). Project members have been instructed about both generic and specific risks to the project.

The risk of information loss in the project or organization is acceptably low. The risk of information leak in the project or organization is acceptably low. The risk of information vandalism in the project or organization is acceptably low.

All personal data will be collected anonymously.

6. Ethics

For the data we produce, the ethical aspects are as follows:

- **Factors that influence people to watch fitness videos**
 - It does not contain personal data.
 - It does not contain sensitive data.

Data we collect

We will not collect any data connected to a person, i.e. "personal data".

7. Other issues

We use the [Data Stewardship Wizard](https://researchers.ds-wizard.org) with its *Common DSW Knowledge Model* (ID: dsw:root:2.4.4) knowledge model to make our DMP. More specifically, we use the <https://researchers.ds-wizard.org> DSW instance where the project has direct URL: <https://researchers.ds-wizard.org/projects/4dab3267-3446-4a09-9b67-94c9021ab998>.

We will not be using any extra national, funder, sectorial, nor departmental policies or procedures for data management.