# DATA MANAGEMENT PLAN

| PROJECT | |
|---|---|
| **Project number:** | [1] |
| **Project acronym:** | [MOSD-210362] |
| **Project name:** | [Hatte die wirtschaftliche Stärke eines EU-Landes Einfluss auf ihren Covid-19 Pandemie Verlauf?] |

| DATA MANAGEMENT PLAN | |
|---|---|
| **Date**: | [23/07/2024] |
| **Version:** | [1.1] |

| HISTORY OF CHANGES | | |
|---|---|---|
| VERSION | PUBLICATION DATE | CHANGE |
| 1.0 | 21.07.2024 | Initial version. |
| 1.1 | 23.07.2024 | Add newly found economic datasets. |
| | | |
| | | |

## 1. Data Summary

This project uses existing data provided by the european union. The first dataset contains weekly data on covid-19 cases and deaths for each eu-country from 2020 to 2023. The second dataset contains yearly data on the pps value (local price adjusted gdp) of each eu-country from 2012 to 2023. These datasets will be used to generate preprocessed csv files. One csv file for the cleaned covid death data, one csv file for the cleaned covid case data and one csv file for the cleaned ppp values per countriy. In addition two csv files will be generated that combine these data points. Visualizing those datasets should provide an general insight to help answering the research question. Since the original dataset require less then 1.5 Megabytes, the total amount of generated data should not exceed 10 Megabytes.

## 2. FAIR data

### 2.1. Making data findable, including provisions for metadata

Due to the small data amount and project size, there will be no persistent identifier for the data. All new data will be described with machine-readable metadata in the xml-fileformat. To create this metadata, the simple dublincore generator will be used. In Addition an extensive ReadMe.md file will provide additional information.

### 2.2. Making data accessible

***Repository:***

All project related data and resources are publicly accessible via github.com. This well established platform provides a long data storage solution and a persistent url as an project identifier to access the data. Github will deposit the data in datacenters based in the USA.

***Data:***

All used data will be publicly avialable with CC0 license via github. This allows easy access with standardized protocols or website access. Since all data can be shared without restictions no further access-management is needed.

***Metadata:***

Metadata is publicly available with CC0 license next to the corresponding data. The metadata will be available as long as the data is available. Relevent software or standardize will be described in a human readable README.md file

### 2.3. Making data interoperable

Data will be in a standardized .csv format and image data will be provided in .png and .pdf format. External data from previous research projects will contain there own metadata.

Sourcecode will be made available as jupyter-notebooks.

All Metadata will be either provided in a machine-readable .xml file way or in a human readable way with the .md format. This resembles commonly used formats to ensure easy data exchange and re-use.

### 2.4. Increase data re-use

To facilitate reusability documentation is provided with extensive human readable readme files, codebooks in form of jupyter notebooks and clean preprocessed data files. All data will be made publicly available with an CC0 open source license. It’s currently not planed to further reuse this data, but reusability should still be a priority. The provenance of data should be documented in a seperate file that links to the original published data. All relevant data should complie with the FAIR principles and data quality should always be assured. Therefore all data should be complete, unique, timely, valid, accurate, consistent. Those attributes should be checked either manually or automatic with a jupyter notebook.

## 3. Other research outputs

Any additional research output should be handled similar to the generated datasets. Therefore all additional output should be in line with the FAIR principles.

## 4. Allocation of resources

Data Storage and archiving require no addition costs with github. The github project author (Fabian Hupfeld) will be responsible for managing the data and ensuring long term preservation (10 years) on github.com

## 5. Data security

Github provides data security solutions and data backups/recovery. Therefore it will be used as a secure and trusted repository for long term data preservation.

## 6. Ethics

No personal or ethic relevant data is used in this data, only large accumulated data is used.

## 7. Other issues

To ensure good data quality, a fair assessment is required after each project milestone. If no unique identifier is present the assessment should be done using the ARDC-Website, otherwise the fair assessment from the Helmholtz Association should be used.