Collecting data individual work

Horizon Europe
Data Management Plan

26 January 2023

History of changes

There are no named versions.

Contributors

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

- Ziyi Wang
  Z.Wang.75@student.rug.nl
  Roles: Researcher

Projects

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

The relationship between the year the work was acquired and the age of the artist

Acronym
N/A

Start date
2023-01-18

End date
2023-01-26

Funding

This project tries to find out the relationship between the year in which an artist's work was acquired and their age. Whose works in this museum were acquired in the artist's twilight years or even after their death? Whose works were acquired at a young age?

Do artists become more experienced as they age and their work is more likely to be appreciated and acquired?

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

- **Collecting data individual work** (published)
  The dataset has the following identifiers:
  - URL:

  We will distribute the dataset using:
  - *Domain-specific repository*. We don't need to contact the repository because it is a routine for us.

  There won't be different versions of this data over time.
  We will be adding a reference to the published data to at least one data catalogue.

We will use an electronic lab notebook to make sure that there is good provenance of the data analysis.

We will be keeping the relationships between data clear in the file names. All the metadata in the file names also will be available in the proper metadata.

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 without instructions how to get access to the data.
Metadata will not be available in a form that can be harvested and indexed.

For our produced data, conditions are as follows:

- **Collecting data individual work** (published)
  The distributions will be accessible through:
  - *Domain-specific repository*. We don't need to contact the repository because it is a routine for us. The distribution will be available under the following license:
    - Freely available for any use (public domain or CC0).

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

2.3. Making data interoperable

We will be using the following data formats and types:

- **CSV Dialect Description Format**

It is a standardized format.

2.4. Increase data re-use

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

- **Collecting data individual work** (published)

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

We will be archiving data (using so-called *cold storage*) for long term preservation already during the project. The data are expected to be still understandable and reusable after a long time.

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

- We will run a subset of our jobs several times across the different compute infrastructures.
- 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.

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

Ziyi Wang is responsible for implementing the DMP, and ensuring it is reviewed and revised.

To execute the DMP, no additional specialist expertise is required.

We do not require any hardware or software in addition to what is usually available in the institute.

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 project web services are addressed via secure HTTP (https://...).

The risk of information loss in the project or organization is acceptably low. The possible impact to the project or organization if information is leaked is small. The risk of information vandalism in the project or organization is acceptably low.

6. Ethics

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

- **Collecting data individual work**
  - It does not contain personal data.

7. Other issues

We use the Data Stewardship Wizard 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/76b7dcbc-50a4-4705-a5af-f3130b810688.

We will be using the following policies and procedures for data management:

- **tate gallery open data to public**
  https://github.com/tategallery/collection/blob/master/LICENCE
  The tate gallery has a very large collection of artworks, most of which have author details, which is very helpful for data research.