Five Poems of Dickinson

Horizon Europe
Data Management Plan

18 January 2024

Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»
using Common DSW Knowledge Model v2.6.3 (dsw:root:2.6.3).

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
There are no named versions		

Contributors

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

- **Zixuan Wu**
  wzx001210@outlook.com
  Roles: *Contact Person*, *Data Collector*, *Data Curator*, *Data Manager*

Projects

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

Five Poems of Dickinson

Acronym:

N/A

Start date:

N/A

End date:

N/A

Funding:

: grant number not yet given

This corpus is some of Dickinson's poems.

This corpus is for lovers of Dickinson's poetry, help them gain a deeper understanding of the textual structure of Dickinson's poetry.

1. Data Summary

Re-used datasets

We have found the following non-reference datasets that we have considered for re-use:

- Five poems of Dickinson
  We are owners of the dataset.

  We will first need to convert the format before using it.

  We already have a copy of this dataset.

  The dataset may get updated in the future; therefore, we need to make a snapshot.

  We will make sure the selected subset will be available together with our results.

  We will use the dataset as follows: Just for the final individual work of my course.

We will need to harmonize different sources of existing data.

Data formats and types

We will be using the following data formats and types:

- **Comma-separated Values** (CSV) type model and format

  A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format. A CSV file typically stores tabular data (numbers and text) in plain text, in which case each line will have the same number of fields.

  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

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 over time.

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 available in a form that can be harvested and indexed (managed by the used repository / repositories).

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

For the reference and non-reference data sets that we reuse, conditions are as follows:

- Five poems of Dickinson
  It is freely available for any use (public domain or CC0).

2.3. Making data interoperable

We will be using the following data formats and types:

- **Comma-separated Values** (CSV) type model and format

  A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format. A CSV file typically stores tabular data (numbers and text) in plain text, in which case each line will have the same number of fields.

  It is a standardized format.

2.4. Increase data re-use

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

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 after the project but also already during the project. The used data archiving service is budgeted by one or more of the participating institutes. Data formats of data in cold storage will not be upgraded over time. Archived data will not be migrated to other storage media over time.

None of the used repositories charge for their services.

Zixuan Wu 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.

Zixuan Wu is responsible for finding, gathering, and collecting data.

Zixuan Wu is responsible for maintaining the finished resource.

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

We are not running the project in a collaboration between different groups nor institutes. Therefore, no collaboration agreement related to data access is needed.

6. Ethics

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 with its *Common DSW Knowledge Model* (ID: dsw:root:2.6.3) knowledge model to make our DMP. More specifically, we use the https://researchers.ds-wizard.org/wizard DSW instance where the project has

direct URL: https://researchers.ds-wizard.org/wizard/projects/1077e86c-2624-4059-bfd4-bb641cfad86e.

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