Research on journal articles

Horizon Europe
Data Management Plan

16 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:

- **Li Mengqi, Wu Zixuan, Wang Shuwei, Lu Yaran, Huang Yuxin, Chen Jiaming**
  S.Wang.68@student.rug.nl
  Roles: *Contact Person*, *Data Collector*, *Data Curator*, *Data Manager*
  Affiliation:
  University of Groningen type Education

Projects

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

Research on journal articles within the scope of Literature and Literary theory from the Scopus database

Acronym:

N/A

Start date:

2023-12-21

End date:

2024-01-19

Funding:

Did not apply for any funding yet.

When we talk about humanities, literature and literary theories are always the first to be mentioned as an important carrier of humanistic ideas. And in the current period of rapid development of the Internet, what topics are generally being studied in academic journals about literature? In order to explore this question, we decided to conduct a textual analysis of academic journals on literature for the whole year just past 2023. Due to the more stringent copyright restrictions on journals, we chose to download open-source journals based on the Scopus database. Also, since English is the most widely spoken language and has the largest number of journal articles written in English, we chose to limit the language used in the journals to English in order to find a more generalised answer.

Therefore, in this project, we focus on journal articles within the scope of Literature and Literary theory from the Scopus database. There is our research question: Analyze the trends and preferences in topic selection of English literature journals on Scopus in the year 2023.

1. Data Summary

Re-used datasets

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

- **Scopus** (Scopus) type repository

  Scopus is an abstract and citation database of peer-reviewed literature across all research areas. It provides tools for visualization, tracking and analysis of its data. Non-subscribed users access Scopus through Scopus Preview, which allows the searching of authors and source material (e.g. journal) titles.

  Owner of this dataset: Scopus.

  The dataset can be used in the provided format without any conversion needed.

  We will use version "the newest version" of this dataset. If a new version becomes available during the project, new analyses will be done with the new version.

  The original dataset will be available both from the provider and from us together with our results for the reproducibility.

  We will use the dataset as follows: We will use it for our research.

There is no need to harmonize different sources of existing data in our case.

We will be using data that needs to be (re-)made computer readable first. We will provide machine readable, standardised metadata to others. The data itself will be made available in computer readable form to others through a standard repository.

Data formats and types

We will be using the following data formats and types:

- PDF, TXT
  It is a standardized format. This is a suitable format for long-term archiving. We expect to have 1072 files of average size 0.000062 GB (i.e. approximately 0.07 GB in total).

2. FAIR Data

2.1. Making data findable, including provisions for metadata

There are no 'Minimal Metadata About ...' (MIA...) standards for our experiments. However, we have a good idea of what metadata is needed to make it possible for others to read and interpret our data in the future.

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

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 including 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).

All data will be owned by the institute.

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

- **Scopus** (Scopus) type repository

  Scopus is an abstract and citation database of peer-reviewed literature across all research areas. It provides tools for visualization, tracking and analysis of its data. Non-subscribed users access Scopus through Scopus Preview, which allows the searching of authors and source material (e.g. journal) titles.

  It is available under specific restrictions, which we will follow in our project: We will only use the journal articles which are in open access.

2.3. Making data interoperable

We will be using the following data formats and types:

- PDF, TXT
  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.

We do not plan to be archiving data (using so-called *cold storage*) for long term preservation already during the project.

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.

Li Mengqi, Wu Zixuan, Wang Shuwei, Lu Yaran, Huang Yuxin, Chen Jiaming 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.

Li Mengqi, Wu Zixuan, Wang Shuwei, Lu Yaran, Huang Yuxin, Chen Jiaming is responsible for finding, gathering, and collecting data.

Li Mengqi, Wu Zixuan, Wang Shuwei, Lu Yaran, Huang Yuxin, Chen Jiaming 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

Project members can carry data with them on password-protected laptops. 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 possible impact to the project or organization if information is leaked is small. The possible impact to the project or organization if information is vandalised is small.

We are not using any personal information.

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".

The data collection is not subject to ethical legislation.

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/dc63185c-2141-47ca-b2fd-ca05608eba19.

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