



Collecting data DMP



Horizon Europe Data Management Plan

28 November 2023

*Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»
using Common DSW Knowledge Model v2.4.4 (dsw:root:2.4.4).*

HISTORY OF CHANGES		
Version	Publication date	Changes
<i>There are no named versions</i>		

Contributors

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

- **Haozhe Bai**

h.bai.3@student.rug.nl

Roles: *Contact Person, Data Collector, Data Manager, Project Manager, Project Member, Researcher*

Affiliation: [*University of Groningen*](#)

Projects

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

Social network and characters analysis on dramas from different cultures

Acronym: N/A
Start date: 2022-11-24
End date: 2024-01-01
Funding: [Rijksuniversiteit Groningen](#): *grant number not yet given (planned)*

This study aims to deeply investigate the evolutionary process of traditional theater in different countries, its influencing factors and its position in the global cultural network through social network analysis and graph theory methods, in order to reveal the dynamic interaction and change of traditional theater in the digital era.

Reveal the topological structure of international traditional theatre network: With the help of social network analysis and graph theory, analyze the relationship network between traditional theatres in different countries, reveal its topological structure, and explore the connection and interaction between various nodes (representing traditional theatres).

Analyzing the evolutionary path of traditional drama: Through studying the historical evolution of traditional drama, trace the developmental path of traditional drama in different countries, and discover its evolutionary trends in different historical periods, as well as the impact of these trends on international cultural exchanges.

Examine the impact of digitalization on traditional theatre: analyze the impact of the digital era on traditional theatre, including the dissemination and interaction on social media and online platforms, as well as the impact of digital technology on traditional theatre performance forms and narrative structures.

Assessing the position of traditional theater in global cultural networks: To measure the influence and position of traditional theater in global cultural networks in different countries, and to gain insights into its performance on the international stage and its interaction with other cultural forms.

Propose strategies for sustainable development: Based on the findings of the study, propose strategies for promoting the sustainable development of traditional theatre, including digital promotion and international cooperation, in order to ensure the vitality and sustainability of traditional theatre in contemporary culture.

1. Data Summary

Re-used datasets

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

- **Dracor** (<https://dracor.org/shake>) ✓

Owner of this dataset: Prof. Dr. Peer Trilcke Phone: +49 331 977 0 Fax: +49 331 97 21 63
E-mail: buero.praesident@uni-potsdam.de.

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

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: This repository contains a collection of Shakespearean dramatic texts encoded in XML TEI (Text Encoding Initiative) format. The dataset is specifically curated for researchers studying Shakespeare's works, leveraging the power of structured and standardized markup.

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

- **Dracor**

We decided not use this non-reference dataset because it misses the data we need.

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.

Data formats and types

We will be using the following data formats and types:

- **ipynb**

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

- **News dataset** (not published)

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.

The provenance will be captured using W3C PROV.

We made a SOP (Standard Operating Procedure) for file naming. Name the file by time and content of the file. 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 including instructions how to get access to the data. Metadata will not be available in a form that can be harvested and indexed.

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

- **Dracor** – freely available for any use (public domain or CC0).

For our produced data, conditions are as follows:

- **News dataset** (not published)

2.3. Making data interoperable

We will be using the following data formats and types:

- **ipynb**

It is a standardized format.

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

- **python**

2.4. Increase data re-use

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

- **News dataset** (not 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 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.

None of the used repositories charge for their services.

Haozhe Bai is responsible for implementing the DMP, and ensuring it is reviewed and revised.

Haozhe Bai is responsible for finding, gathering, and collecting data.

Haozhe Bai 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 will not store data or software on computers in the lab or external hard drives connected to those computers. They can carry data with them on encrypted data carriers and 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 possible impact to the project or organization if information is lost is small. 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.

All personal data will be anonymized as early as possible.

The archive will be stored in a remote location to protect the data against disasters. The archive need to be protected against loss or theft. It is clear who has physical access to the archives.

6. Ethics

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

- **News dataset**
 - 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/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/wizard> DSW instance where the project has direct URL: <https://researchers.ds-wizard.org/wizard/projects/00e33f6f-46d6-4bfc-939a-d1dc8dd6f4b9>.

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

- **RUG Policy**

<https://www.rug.nl/research/research-data-management/policy/ug-rdm/>

Its purpose is to ensure innovative research and research integrity.