Data Management Plan
Feminist Interventions: Designing Descriptive Markup for the Marginalized Women Directors Project

Background

This Data Management Plan (DMP) is integral to the Feminist Interventions: Designing Descriptive Markup for the Marginalized Women Directors Project and should be read in conjunction with the project plan.

Key Words

Feminist, director, markup, films, women, data, metadata, encoding, database, dataset

New Information

This DMP illustrates the management of the data collection, the metadata creation, and the encoding process associated with the project including source code, data repository, framework replicability, data collaborative notes, training, and/or any other additional information materials that contributed to the outcome.

Research product

The project aims to produce innovative methods to encode information on marginalized women directors' lives and works through feminist perspectives contributing to the existing production of the scholarly history of studies on women as creators in the film industry.

Data Information

Four levels of data are managed for this project:

- The new TEI/XML tags and associated definitions created by the project team (text files)
- The documentation of the encoding process (XML files)
- The text files and video files of encoding tutorials recorded by the team
- The design (software code) of the searchable database model

The project's primary commitment is to gather, sort, organise, and present information on the lives and works of a selected subset of marginalized directors (10 to 15 individuals selected). The expected data will be produced through the following steps:

➢ Step 1: The search and collection of textual information, images, and audio/video materials, related to the lives and works of selected marginalized women directors and their transcription and recording in a textual format.

➢ Step 2: The generation of customized TEI/XML tags to describe, summarise, interpret, and explain the information collected in Step 1. The newly created textual data will be digitally encoded through the creation of tags, or defined elements. This step will constitute the foundation of a searchable database model that, in the future, could be accessed via the targeted keywords (i.e., specific text search) but also through tags. The encoding process will be recorded in textual and video formats, which will eventually be made into openly accessible tutorials.

➢ Step 3: The project will develop and run a potential searchable database model using the data generated in Steps 1 and 2.

Documentation

The data will be accompanied by an exhaustive documentation defining each data point, field, or tag used to encode and/or catalogue the original data. The data and metadata dictionary will be provided in the form of a separate discursive file which will also be embedded in the primary dataset. The data-creation process and related metadata decisions will be documented and made available on the project website through tools such as Word, Google Docs, Oxygen, or Visual Studio Code. The metadata and software code will be also provided with sufficient technical and contextual details to ensure these can be migrated to other coding languages for building the profile pages on the website and the searchable database model. The accompanying documentation will be uploaded to both the GitHub repository and the Commons website.

Data Storage and Dissemination

The project team fully appreciates the value of open-source development praxes in the Digital Humanities community and is committed to ensuring free accessibility and availability of the encoded data repositories on the project website under the Creative Commons CC BY 4.0 License. The team will also use GitHub as a collaborative coding infrastructure to simplify the exchange of information within the team and encourage future contributions to the project. Additionally, the team will keep a shared repository of the preexisting codes and materials on TEI/XML on a communal Google Drive, which can be made available upon request (the project team reserves the right to decline such a request should the latter cause harm or danger). The team believes this approach will improve the project's greater outlook and sustainability over a longer time horizon.

Accessibility and Framework Replicability

The project's outcome and materials will be available primarily to CUNY students as the website will live on the CUNY Commons platform. However, the GitHub repository will be open to anyone who wishes to learn more about the project or replicate the project's steps to achieve such encoding. In addition, the documentation discussed in the previous section will allow users to explore and download the data, metadata, and any additional materials.

Roles and Responsibilities

The data management will be implemented by the developer Gemma S., supported by the project manager Miaoling Xue the other team members, Maria Buitrago and Nadim Essey, will assist with producing and transferring data between the project website and the GitHub repository. The long-term and permanent storage needed for the data produced will be further determined after the release of the first trial of the research product.