Individual Project Management Plan for *All the Metadata's a Stage*:

Exploratory Data Analysis of British Comedy Dramas Across 17th-19th Centuries

Course: Introduction to Digital Humanities
Academic Year: 2025–2026
Instructor: Roberta Pireddu

Student: Chahna Ahuja (r1019982, chahnaparesh.ahuja@student.kuleuven.be)

Submission Date: January 5, 2026

1 Project Scope and Objectives

Following the team's data wrangling in *OpenRefine* using the British Library's bibliographic dataset of 17th-19th century British dramas, my personal contribution to the project focuses on the **exploratory analysis of the comedy genre**. Specifically, I will examine temporal trends in publications, spatial networks of publishing locations, authorship distribution, and lexical patterns in comedy titles. The scope of my work encompasses:

1. **Data cleaning and enrichment** tailored to the requirements of the individual project.
2. **Exploratory data analysis and visualization** in *Python*.
3. **Text network analysis** using *Gephi*, with findings disseminated via interactive HTML Jupyter notebook and a *GitHub* repository.

The primary goal of this project is to explore the comedy dataset through an exploratory, distant-reading approach, situating the findings within the context of literary history scholarship on British stage comedies in 17th-19th centuries. The secondary goal is to produce a reproducible, digital-humanities-oriented workflow thorough documentation and accessible dissemination.

2 Project Lifecycle

The project lifecycle is divided into eight iterative phases to ensure a feasible and structured workflow:

Phase 1: Dataset Exploration and Preliminary Research Proposal: Exploring the dataset in *OpenRefine*, reading literature, compiling an annotated bibliography on the literary history of British dramas (17th–19th centuries) in *Zotero*, and formulating preliminary research questions based on the literature and dataset.
Phase 2: Defining Methods: Reviewing methodological approaches that support exploration of the research questions.
Phase 3: Data Cleaning, and Descriptive Statistics: Loading and inspecting the dataset in Python, rectifying inconsistencies in *OpenRefine*, and generating overview visualizations.
Phase 4: Data Enrichment: Enriching linked authors with *WikiData* by adding more information such as gender, images, etc.
Phase 5: Exploratory Data Analysis: Conducting temporal, spatial, and lexical analyses and visualization using *Python (Pandas, Plotly, spaCy, NLTK)*.
Phase 6: Text Network Analysis: Preprocessing text, programming networks in *Python*, and visualizing them in *Gephi*.
Phase 7: Interpretation for Report: Analyzing visualizations alongside reviewed literature to prepare findings.
Phase 8: Deliverables: Formatting the report, peer review, producing an interactive HTML Jupyter notebook, and uploading all materials to the *GitHub* repository.

3 Project Timeline and Milestones

The project timeline is structured with clearly defined milestones to ensure feasibility and timely delivery (Fig 1). A detailed Gantt chart accompanies this plan.

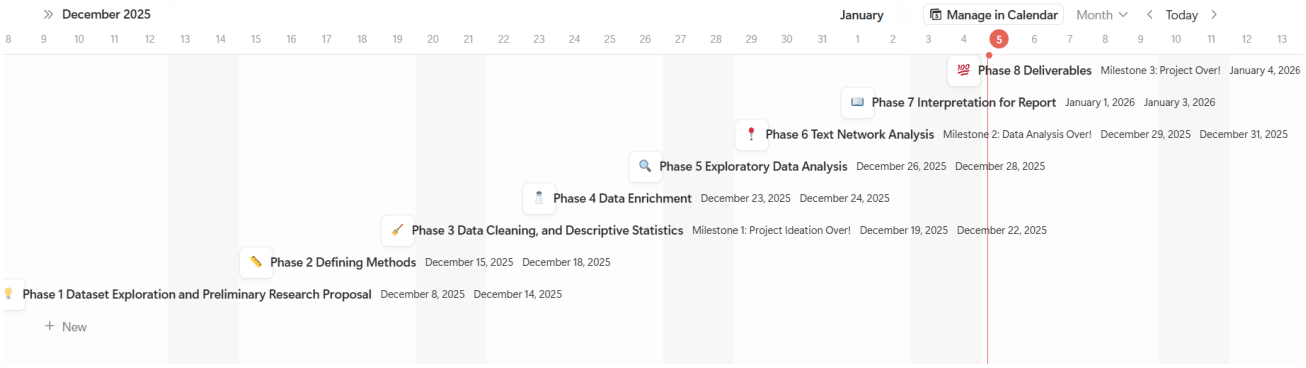

Figure 1: Project timeline and Milestones

4 Technical Requirements and Resource Management

The project workflow relies exclusively on open-source and freely available resources and platforms to ensure efficient resource use and long-term accessibility.

Table 1: Tools used across project stages

Category	Tool(s)	Purpose
Organization	Zotero, Notion	Literature collection, annotation, note-taking, and research planning
Data Cleaning	OpenRefine	Resolving inconsistencies, normalizing values, and ensuring data quality
Analysis	Python (Pandas, Plotly, NLTK)	Data processing, NLP-based analysis, and visualization
Network Analysis	Gephi	Network visualization
Metadata Enrichment	Wikidata API in Python	Enrichment of author metadata (e.g. gender and biographical portraits)
Dissemination	Jupyter Notebook (HTML export)	Interactive presentation and dissemination of analytical results
Version Control	GitHub	Version control and public sharing of datasets, scripts, and notebooks
Documentation	GitHub README, in-notebook comments	Reproducibility, transparency, and methodological clarity

5 Risk Assessment and Mitigation Strategies

The following risks are anticipated during project planning and the folloiwing are mitigation strategies to be executed during the project:

What if there are data quality issues? This risk will be **addressed** early in the ideation phase through systematic null-value checks, prioritizing the subset with most data completeness, and resolving inconsistencies in *OpenRefine*.

What if the hypotheses are not supported by the data? This risk will be **retained**, and an iterative exploratory approach will be adopted, allowing the analytical focus to shift if initial methods (such as co-occurrence networks) do not yield meaningful patterns because research is allow about revising initial questions.

What if there is data loss? This risk will be **avoided** by maintaining regular cloud backups and using GitHub version control with frequent commits to ensure data security and recoverability.

What if technical difficulties are encountered? This risk will be **reduced** by working across multiple tools (Python and Gephi) and relying on incremental testing, dry coding, and package documentation for troubleshooting.

What if I run out of time? This risk will be **insured** through a detailed milestone plan with regular self-checks to monitor progress, particularly in light of the additional time commitment required for peer review responsibilities.

6 Project Deliverables

The key deliverables will include:

- Cleaned and enriched British Library drama dataset in line with this project.
- Reproducible Python scripts for exploratory data analysis and visualization.
- Network files and visualizations for Gephi.
- Interactive HTML Jupyter notebook presenting visualizations and results.
- Written analytical report contextualizing findings within literary history scholarship.
- GitHub repository containing datasets, code, and documentation as workflow.