Project Management Plan

Individual Analysis of the British Library Drama Dataset

Xinran Liu
Project Duration: 8 December 2025 – 4 January 2026

1.Project Overview and Research Goals

Based on the group's master dataset, the individual task shifts the focus to analysis and interpretation. I retain the group plan's principles (protocol-first processing, quality assurance, and documentation), but also adopt methods of extracting variables for further analysis (decade aggregation, concentration metrics). The final outputs will be a written report and tableau dashboard.

Individual research goals (what I want to find out):

➢ Descriptive Analysis of the dramatic data, mainly on the publication volume, authorship, and publishing patterns
 ✓ Describe changes in **publication volume** across decades and identify shifts in **genre composition**
 ✓ Examine **authorship** patterns, especially "Unknown" attribution and author concentration.
 ✓ Examine **publisher** patterns, including missingness and publisher concentration (CR3).
➢ Analysis of Titles and Socio-cultural Reflections
 Track title-based thematic signals (geographic discourse and socio-cultural framing)

2.Work Breakdown

Phase 1 — Research Scoping & Methodological Setup (8–14 Dec 2025)
Conduct a targeted review of 3–5 key texts on British drama history and quantitative literary studies. Formulate 2–3 specific research questions based on these readings. Perform exploratory analysis to identify outliers or limitations in the group-cleaned dataset.
Outputs: Research questions and required variables, GitHub repository for version control.

Phase 2 — Structural Analysis of Metadata (15–21 Dec 2025)
Analyze publication volumes and genre proportions by decade, authorship evolution (identified and unknown) and concentration over time, and publisher concentration (using metrics like CR3; HHI is not used because publisher missingness is high (42.33%) and would make HHI estimates unstable.)
Outputs: Jupyter notebook file for code, Export tables (CSV) used for visuals/dashboard: drama_master_cleaned_final.csv, viz_top10_publishers.csv, viz_cr3_market_analysis.csv,

Phase 3 — Thematic Text Mining & Interpretation (22–28 Dec 2025)
Build semantic dictionaries for Geographic and Socio-cultural Terms derived from the literature review; Run initial matching tests; Generate frequency counts.
Outputs: Versioned dictionaries and notes, geo_trends_by_decade.csv, tableau_public_private_refined.csv
Research questions were revised at this stage, with analysis of title length removed, as the preliminary results don't show much interpretability.

Phase 4 — Synthesis, Visualization & Critical Reflection (29 Dec – 4 Jan)
Draft the report, combining quantitative trends with historical interpretation. Finalize the Dashboard/Figures ensuring they accurately represent the underlying data logic. Write the self-evaluation section, and document the peer review experience.
Outputs: final **report** including a Reflection subsection, brief peer-review note;Tableau dashboard (link), submission package (code + charts + documentation)

3.Timeline and Milestone:

Milestone 1 (14 Dec 2025): Research Questions + Dataset Checked

Final research questions selected based on literature, dataset reviewed and confirmed.

Decade/genre/Unknown rules confirmed and documented.

Milestone 2 (21 Dec 2025): Core Quantitative Results Exported

Publication/genre/authorship/publisher metrics completed.

Export tables produced and validated (counts, denominators, missingness notes).

Milestone 3 (28 Dec 2025): Title-Thematic Method Finalised and Validated

Dictionaries (v1) with a validation memo (false positives/negatives + ambiguity log).

Theme indicators exported with both counts and rates.

Milestone 4 (4 Jan 2026): Final Deliverables, Self-Evaluation and Peer Review;Word report final, dashboard final.

Report finalised, Tableau dashboard published, Reflection/self-evaluation section completed.

4.Tools and Resource Management

Tools (and how they are used)

Python (Jupyter): processing, feature engineering, exporting charts and tables.

OpenRefine: targeted checks/standardisation if new inconsistencies are detected.

Tableau: dashboard development from exported tables.

GitHub (Git): version control and change tracking.

Resource management strategies (project management discipline)

Folder structure: data_raw/, data_clean/, exports/, figures/, notebooks/, docs/.

Versioning: semantic naming (e.g., clean_v1.0, geo_dict_v1.0).

5.Quality Assurance and Risk Mitigation

· **Missingness checks:** report missingness by decade for Author and Publisher; annotate in outputs.

· **Reproducibility:** full pipeline run from a clean environment before final submission

Risk	Impact	Mitigation
High missingness (Publisher/Author)	weak inference	cautious interpretation
Small sample sizes	unstable percentages	add small-N warnings
Dictionary subjectivity	biased themes	versioned dictionaries; manual vetting memo; sensitivity note

6.Deliverables

Final Report

dataset/methods; results; interpretation; limitations; references.

Dashboard

Tableau dashboard (published link) or final set of charts.

Reproducible Analysis Package (GitHub repo)

Notebooks; exported tables used by Tableau; README.

Documentation Bundle

Data dictionary; changelog; dictionary files (geo/social + stopwords).