Data Paper Template

~~~~~~

Dataset Curation Grant recipients will use this template over the course of their grant period to document their process and create a final "data paper" that reports on the collection, cleaning, and management of their publicly-accessible dataset. Finished data papers will resemble one of the following peer-reviewed reports on the making of humanities datasets, a genre increasingly published in DH journals. CDH staff can assist with preparing this final data paper for journal submission if desired.

- Miriam Posner et al., "Early African-American Film Database, 1909–1930," *Journal of Open Humanities Data*, 4.
  - dataset: https://doi.org/10.5281/zenodo.56591
- Sherif Abuelwafa et al., "Detecting Footnotes in 32 million pages of ECCO," *Journal of Cultural Analytics*, December 3, 2018.
  - dataset: https://doi.org/10.7910/DVN/FMZYFP
- Markus Neuwirth et al., "The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets," *Frontiers in Digital Humanities,* 5:16.
  - dataset: https://github.com/DCMLab/ABC

Section 0, "Charter," can be filled out now. Fill out the other sections during orientation to the best of your knowledge. Going forward, **think of this document as your control panel.** Check in regularly with your team and document your process here.

~~~~~~

0. Charter

a. **Project team:** *names, roles.*

b. **Deliverables and milestones dates.** *List the main project outcomes and significant points in project development for this grant period.*

c. **Scope statement** (2-3 sentences). *Describe the main work to be done (during dataset curation only). List what is out of scope.*

d. **Work plan.** *What are the main tasks you want to accomplish per month during the course of the grant?*

e. **Audience.** *Who do you envision using your dataset?*

f. **Risks or interdependencies.** *Is there anything that might go wrong that would derail your project? Does your success rely on anyone else participating or completing some work?*

g. **Rights and permissions.** *Do you have all of the permissions and rights you need?*

- Have you established who owns the copyright of your data? Might there be joint copyright?
- Do your data contain confidential or sensitive information? If so, have you discussed data sharing with the respondents from whom you collected the data?
- Are you gaining (written) consent from respondents to share data beyond your research?
- Do you need to anonymize data, e.g. to remove identifying information or personal data, during research or in preparation for sharing?

h. **Post-grant plans for project.**

i. **Budget**. *Use template below.*

Expense	**Cost ($)**	**Paid to**
Total		

1. Overview

a. **Dataset title.**

b. **Dataset abstract and justification** (2-3 sentences). *Brief summary of the dataset: what the data covers and a short description of its intended impact.*

c. **Context.** *Was this data produced as part of a book project, dissertation, course work, or article? If so, please list the appropriate bibliographic information here.*

2. Methods

a. **Steps.** *Describe the series of procedures followed to create and organize the dataset. This should include any source data used, as well as software, algorithms, or instrumentation involved. Describe your standardization of names, codes or abbreviations. Mention any controlled vocabularies/thesauri used, or, if you developed one, describe it.*

b. **Sampling strategy.** *If relevant, please outline the sampling strategy used to produce the data.*

c. **Quality control.** *If applicable, please list the methods used for quality control in the production of the data. Were steps taken to normalize spellings or dates? What scholarly or historical definitions of key terms were used to decide what to include in the dataset? Was any accuracy verification done?*

3. Documentation

a. **Data structure and format.** *List the data formats you used, e.g. ASCII, CSV, Autocad, EPS, JPEG, Excel, SQL, etc. Use formats that enable sharing and long-term validity of data, such as non-proprietary software and software based on open standards. If you have multiple files, describe the relationships between them.*

b. **Date of creation & date(s) of updates.**

c. **Language(s).**

d. **Access information.** *Does a user need any special software or skills to get the data*?

e. **Rights.** *Intellectual property or licensing rights for the data. We recommend: CCBY (for reusability)*

f. **Data citation.** Preferred format for citing data.

4. Management and Reuse

a. **Repository name.** *E.g. Figshare, Zenodo, Dataverse.*

b. **Persistent identifier.** *DOIs will be generated by your repository.*

c. **Data storage, security and backup.** *Are your digital and non-digital data, and any copies, held in a safe and secure location? If data are held in various places, how will you keep track of versions? Are your files backed up sufficiently and regularly and are back-ups stored safely?*

d. **Reuse potential.** *Please describe the ways in which your data could be reused by other researchers both within and outside of your field. For example, this might include aggregation, further analysis, reference, validation, teaching or collaboration. This section should also include limitations to, or potential barriers for reuse.*