

Data Management Plan (DMP) Checklist

Prepared by the Qualitative Data Repository (www.qdr.org)

Center for Qualitative and Multi-Method Inquiry | Maxwell School | Syracuse University

QDR has assembled a checklist outlining the issues you should address in your Data Management Plan (DMP). The checklist, which is thorough but non-exhaustive, contains questions asked in a general way so that it can accommodate various types of data projects, and so the answers are likely to satisfy the DMP requirements of any funding organization.

You may not need to answer every question for your particular project. Nonetheless, addressing all of the questions relevant for a given project in a detailed way helps you to make clear to other researchers, to funders, to scholars who review your work, and to repositories what procedures you will follow throughout the data lifecycle.

The main bulleted questions concern fundamental issues that you should begin to consider at the earliest stages of planning your work. The more detailed questions and tips included beneath each main question are designed to help you think further about each aspect of data management and should also be considered and answered as early as possible in the life of a research project.

- What **personnel** will work on the project?
 - *Qualitative data example:* in addition to the researcher, a team of three RAs will collect newspaper articles.
 - *General example:* in addition to the three-person core research team, a survey firm will be contracted to administer the survey.
 - *Details*
 - What specific responsibilities and levels of access to the data (if applicable) will each individual / organization have?
- What **practices and procedures** will you use to collect data sources/data?
 - *Qualitative data examples:* ethnographic fieldwork; participant observation; archival research; elite interviews; working with a commercial focus-group facilitator firm.
 - *General examples:* drawing from a researcher's prior project or an existing database; lab experiment; commissioned survey.
- In what **media/formats** will you collect and use data sources/data?
 - *Qualitative data examples:* WAV audio files (e.g., of recorded interviews); scans (e.g., of archival documents) in PDF format; photographed images as JPEG files; typed transcripts of focus group conversations in .docx format; paper-based handwritten notes; NVivo10 for Windows projects in .nvp.
 - *General examples:* MS Excel spreadsheets; STATA14 .dat datasets and .do files; other proprietary or open-source formats; tables with comma-separated values; computer code.
 - *QDR tip:* try to use open-source or prevalent formats from the beginning as proprietary formats will require conversion. A list of recommended formats for

common forms of qualitative data can be found here. If you must use a proprietary format, document the exact version number of the software used.

- *Details*
 - Are the media and format types used easily accessible to others (on your team or the wider scholarly community) using existing technology? If not, how can usable forms of the data products be obtained?
- What **volume** of data sources/data will you collect?
 - *Qualitative data examples*: ~ 2000 video clips, <3 minutes each; 200 cassette tapes; 1.2 cubic feet of organizational records housed in a standard archival box (15"x12"x10").
 - *General examples*: 200 MB of document scans; >1 TB geospatial images.
 - *QDR tip*: it is important to plan for the volume of data you are collecting, but do not worry about coming up with precise estimates. As long as the actual volume is no more than two or three times your original estimate, any technology you chose will scale.
- How will you **organize** your data sources/data?
 - *Qualitative data example*: paper copies of archival documents will be filed alphabetically in a standard archival box; scans of newspaper articles will be stored in folders labeled by month/year.
 - *General examples*: geospatial images will be stored in folders labeled with particular geographic coordinates.
 - *Details*
 - What file naming convention will you follow?
 - What versioning system will you use to clearly indicate updates to files?
- How will you **process and transform** your data sources/data?
 - *Qualitative data example*: anonymization of interview transcripts; scanning of newspaper articles collected in paper form; auto-coding in qualitative data analysis software.
 - *General examples*: aggregation of quantitative data from different sources; cleaning of datasets after initial assembly.
 - *Details*
 - What migrations/conversions will be needed between different media/formats?
 - How will you handle the pre-processed data sources (original versions) after you have transformed them?
- How will you guarantee the **security and integrity** of the data sources/data?
 - *Qualitative data example*: notes from interviews will be kept in a locked location, with no information identifying the respondent; typed notes are stored in encrypted files.
 - *General example*: all files will be encrypted.
 - *Details*
 - Who will have access to the data sources during the administration of the project?

- How will unauthorized access and intentional or unintentional alteration of the data be prevented?
- What **documentation** will you generate?
 - *Qualitative data examples:* interview protocols; detailed notes on the location of archive boxes from which documents were drawn; a list with the naming convention for each type of file.
 - *General examples:* descriptive metadata explaining procedures used to collect data and organize files; codebooks; metadata associated with digital files.
- How will you **store and backup** data sources/data *during* project execution?
 - *General examples:* on an external hard drive, in the cloud
 - Details
 - What is your back-up plan?
 - Who is responsible for carrying out / checking specific aspects of back ups?
 - What security measures will you take, given the sensitivity of the data? (Cloud services like Dropbox may not be suitable for sensitive data.)
- How will you **store and preserve** data sources/data *after* project completion?
 - *Qualitative data example:* after anonymization, the raw data sources will be discarded, and the processed data sources will be archived at repository X in perpetuity.
 - *General example:* dataset will be archived at repository X in perpetuity.
 - Details
 - Who will be responsible for preserving data sources?
 - Where will data sources and data be preserved and why was that venue chosen?
 - For how long will data sources be preserved?
- What are your plans, if any, for sharing your data?
 - *Qualitative data example:* Data will be deposited without access restrictions with the Qualitative Data Repository;
 - *General example:* Self-archiving at Dataverse; experimental results dataset will be shared with the NASA open-access Prognostics Data Repository; data will be shared first with the organization being researched and will not be made public for six months (while the organization's executive management decides whether to allow further dissemination).
 - *Details*
 - Will you share data sources/data publicly?
 - When will you share data sources/data?
 - Who will administer the shared data sources/data?
 - What metadata standard will be used to ensure interoperability and discoverability of the data sources/data?
 - *QDR tip:* it is best to use the practical standards common to the relevant research community. While writing your DMP, contact the venue, where you plan to deposit your data for guidance.

- Will you convert any data files to non-proprietary formats for the purposes of wider sharing?
- In what ways will the data sources/data be citable?
 - *QDR tip:* the gold standard are citations including persistent identifiers, such as DOIs, which reliably link back to the online location of the data. A data repository can usually issue such an identifier for you.
- Do the data sources/data present specific challenges for privacy of individuals or intellectual property?
 - *QDR note:* if your research will involve interaction with human participants, make sure that your DMP and your IRB application align. QDR staff can assist you with devising a consent process that thoughtfully enables eventual sharing.
 - How will you ensure confidentiality for sharing?
 - How will you address copyright and patent limitations?
 - *QDR tip:* possible approaches include securing a license, fair-use exceptions, no-fee authorization to use for research, etc.
- Will you impose special conditions for access and re-use?
 - If yes, what types of conditions?
 - *QDR tip:* remember that different access conditions can be set at the file level; you can make some files freely accessible while restricting others.
- Will any audiences be prevented from accessing the data sources/data?
 - If yes, why?
- If data sources will not be shared, what is the justification?
 - *QDR tip:* evaluate each data source for shareability individually; even if some of your data cannot be fully shared, you may be able to share other parts.