

The suggested checklist contains a number of questions asked in the most general form possible so that it can accommodate answers for various types of data projects. Not every single question needs to be answered in every case, but the researchers planning for the data management component of their work should be able – by addressing all aspects relevant for a given project – provide sufficient and sufficiently detailed information that makes it clear to other researchers, to funder-based or publication-based reviewers and to repositories what procedures will be followed throughout the data lifecycle.

I. General information – this is the most concise (bare minimum) list a researcher should address even if at the earliest planning stages. The more detailed breakdowns of the project into phases below will ideally also be considered in advance and decisions will be made on all relevant points to the degree possible.

- ✓ What are the expected research data products?
- ✓ What is the source of the data collected?
- ✓ What are the formats of the expected research data products?
- ✓ What is the expected size of the research data products?
- ✓ What are the plans for preservation of the research data products?
- ✓ How will the research data products be shared? Will special access conditions be imposed?
- ✓ If no provisions for data sharing are being made, what is the justification?

II. Data management during the course of the project

- ✓ Storage: what is the back-up plan? Who has responsibility for carrying out specific portions?
- ✓ Media/formats used: will any migrations/conversions be needed between different media/formats? Are the media and format types used easily accessible with mainstream current-day technology? If not, what are the best ways for researchers (on the team or potential secondary ones) to obtain usable forms of the data products? (If there are open-source or prevalent formats, try to use those from the beginning rather than proprietary ones which will require conversion.)
- ✓ Security and integrity of the data: who will have access to the raw data during the administration of the project? How will unauthorized access and intentional or unintentional alteration of the data be prevented?
- ✓ File names: what naming convention and organization will be followed during the project?
- ✓ Versioning: what versioning system will clearly indicate different updates to the same files?
- ✓ Collaboration and teamwork: can you provide a full list of all personnel expected to work on the project, with specific responsibilities and levels of access to the data (if applicable)?

III. Data retention after the end of the project:

Data Management Checklist
Teaching Materials Prepared by the Qualitative Data Repository (www.qdr.org)
Center for Qualitative and Multi-Method Inquiry | Maxwell School | Syracuse University

- ✓ Preservation venue: where will the data be preserved? Can you justify explicitly the reasons why this choice was made? (Related to Administration below.)
- ✓ Length of commitment to preserve: how long is preservation planned for?
- ✓ Administration of the preserved data products: who will be responsible for the preservation?
- ✓ Raw and processed data handling: will raw data be collected in a way that will necessitate substantive transformation of any sort (e.g., aggregation, anonymization)? If so, can you provide specifics about how both the raw and the processed data products will be handled after the end of collection?

IV. Plans for data sharing

- ✓ Will data products be shared in a publicly accessible way?
- ✓ When will data products be shared?
- ✓ Who will administer the shared data products? What audiences will be allowed/not allowed to access the data products?
- ✓ What metadata standard will be used to ensure interoperability and discoverability of the data? (It is best to use the practical standards common to the relevant research community.)
- ✓ What additional documentation (project- and file-level descriptors of various sorts: descriptive metadata explaining content/file organization/selection procedures used; codebooks; interview scripts; metadata associated with digital files; etc.) will the researcher provide to make the data products sufficient for re-use?
- ✓ Will any files be converted to non-proprietary formats for the purposes of wider sharing?
- ✓ In what ways will the data be citable? (Gold standard is to have persistent identifiers, such as DOIs, which can then be used to hyperlink back to the online location of the data.)

V. Access conditions for sharing

- ✓ Will any special conditions be imposed for access and re-use?
 - ✓ If yes, why?
- ✓ Do the data present specific challenges for privacy of individuals or intellectual property?
 - ✓ If yes, how will privacy and confidentiality be ensured for sharing? (Should have already been addressed with IRB process: informed consent, no collection of personal identifying information or de-identification before sharing, etc.)
 - ✓ If yes, how will copyright and patent limitations be addressed? (Licenses, fair-use exceptions, no-fee authorization to use for research, etc.)