



Managing Evaluation Data

WHY IS DATA MANAGEMENT IMPORTANT?

Good data management is essential to a successful evaluation because without it, your data may not be accurate. Evaluation results depend on accurate data, so the methods you use to enter, store, and prepare data for analysis will directly affect your ability to:

- Identify what's working in the project (and what's not)
- Complete required reports
- Fulfill funding requirements
- Demonstrate outcomes

WHY DO I NEED A DATA MANAGEMENT PLAN?

Data management is complicated and includes many steps. Planning ahead with a data management plan that details each of these steps will help keep you on track throughout your evaluation and ensure that you don't miss any steps along the way. Your plan will help you create an accurate dataset for analysis, maintain consistency in data treatment over time, and run your analyses more smoothly. In addition, a clear plan will make it easier to answer questions that arise and conduct additional analyses later. Your data management plan should detail the steps associated with:

- Tracking data in the collection phase
- Entering data into a database
- Cleaning data
- Storing data
- Eventually disposing of data

On the next page, we have outlined a detailed list of steps associated with data management.

HOW DO I "CLEAN" MY DATA?

One of the most important parts of data management is cleaning the data. Data cleaning refers to all steps you take to detect and eliminate errors in the dataset; these steps are a key part of your overall plan. Here are some of the key components of data cleaning for which you should plan:

Confirm data format and ranges. You'll want to ensure that all responses are in the expected format (for example, text or numeric). Next, confirm that data values fall within the range of possible responses. For example, for a scale that goes from 1 to 5, all responses should be within that range. Finally, search for and eliminate duplicate data (such as multiple forms completed in error by the same respondent or data accidentally entered twice).

Check for data entry errors. Data entry errors happen despite precautions, so you'll want to develop a plan to systematically check for these errors. For example, you may want to randomly select a sample of the data in your dataset and check for accuracy by comparing the dataset to the original data. Correct any errors you find and, if you discover a pattern of errors, check your entire dataset for this pattern and consider revising your data entry procedures.

Check for other inconsistencies. Make sure that your dataset contains only data from valid participants. For example, if the topic of a survey is a particular set of services and the

respondent did not receive those services, you should remove that respondent's survey from the dataset. You should also check for consistency within a respondent's data. For example, it would be inconsistent if a teacher indicated five years of teaching experience but seven years of teaching in her current school. If you discover an inconsistency, return to the original forms to check for data entry error. If the inconsistency is on the original form and you can't determine the source, you may have to eliminate all of the inconsistent data from that form.

Investigate missing data. Check back against original data to ensure that missing data isn't missing due to a data entry error. Then, review each case and consider whether the amount of missing data is acceptable. For example, if more than half of the responses are missing from a survey, you may need to remove that survey from the analysis dataset. With survey data, you may be able to follow up and get responses from participants who skipped several questions, if you contact them right away.

Always remember to save your original data file and create a new "clean" version before cleaning your data!

STEPS FOR SUCCESSFUL DATA MANAGEMENT

- **Create a data management plan.** Your plan should include steps for tracking data, entering data into a database, cleaning data, storing data, and eventually disposing of data.
- **Document all processes and decisions.** Throughout the data management process, it's critical to document all rules, processes, and decisions—especially those that deviate from the data management plan or new decision rules related to cleaning your data. Documentation will let you answer questions related to analyses, easily duplicate or update analyses if needed, and ensure consistency throughout the evaluation period.
- **Develop a data tracking system.** Before you begin to receive data, develop a data tracking system that allows you to track all of the data you'll receive, identify respondents who may need additional follow-up, and document missing data. Your tracking system might stand alone or be part of a larger database that will eventually contain all data for analysis.
- **Track data as they arrive.** Using your data tracking system, track the status of your data. Create codes for the various statuses (for example, complete, refusal, ineligible, non-locatable) and, as you receive forms or learn information about participants, enter and update the codes for each participant in the tracking system.
- **Choose a program for data entry.** You can enter your data directly into a program designed for analysis (such as SAS for quantitative data or NVivo for qualitative data), or you might choose to enter the data using software such as Excel or Word and then import it into an analysis program.
- **Organize your data file.** You'll want to set up your data file to facilitate later analysis and to reduce the likelihood of data entry errors. The best organization will depend on the type of analysis you plan to conduct. However, regardless of the type of analysis you plan to conduct, you can reduce data entry errors with steps such as specifying cell format (for example, text, numbers, or dates), specifying the range of possible responses (for example, only birthdates within a specific range), or including drop-down boxes to allow only pre-selected responses.
- **Enter data.** You have many options for data entry, including manual entry, digital scanning, and respondent entry (as with online surveys). If you enter data by hand, train data entry staff using standardized and documented data entry procedures. Also, consider double-keying (entering data twice). Double-keying reduces data entry errors (any discrepancy between the two sets of entered data can be corrected using the original response).
- **Save your original dataset.** You may need to refer to your original dataset later, especially to investigate extreme responses or if you revise any coding or cleaning decisions. Do your cleaning in a new dataset with a new name.
- **Clean your data.** Cleaning data refers to checking data for accuracy and completeness. You should confirm the format and ranges of your data, check for data entry errors, check for other inconsistencies, and investigate missing data. Each of the steps associated with cleaning your data is briefly described on the previous page.
- **Recode data for analysis.** You might need to convert scaled responses (for example, never, rarely, sometimes, often) to numbers or recode individual responses into categories (for example, convert years of teaching experience into categories like 0-1, 1-5, 5 or more).
- **Secure your data.** Take all necessary steps to protect your data. For example, update antivirus software, install firewalls, encrypt data, use password authentication to limit and track user access to datasets, and keep removable storage devices and hard copies in secure areas.
- **Protect confidentiality.** Implement procedures to protect personally identifiable information. These procedures might include using identification codes (IDs) in place of identifying information (for example, a random number instead of a name), storing data and identifiable information separately (use the ID to link the files), and using secure data transfer protocols.
- **Store data.** Store all electronic and hard copies of data securely. Store electronic data on secure servers using password authentication. Use locked data filing for hard copies of data collection instruments and forms.
- **Determine the final disposition of data.** Determine how long you'll store data and whether and how you'll eventually archive or destroy it. The client or funder may require certain timelines or treatment after the project ends. When you dispose of the data, it's important to use secure methods of data destruction, including destroying backup copies, shredding paper copies, and performing electronic "wiping" of computer disks.