

# Preparing for Analysis

*It is very likely that there are several steps between the data you collect and the data you ultimately examine, analyze, and publish. Properly preparing data involves both ensuring that your data exists in a form ready for examination or analysis, and ensuring that you have documented how and why you prepared your data in the manner that you have. This is where you need to think about what you planned and address the reality about what you can or need to do.*

---

## What does it mean to prepare data?

Preparing data means cleaning, coding, processing, or otherwise transforming it in some way. While doing this, it is important to document what you've done so that your steps can be re-traced – by yourself or by others – in the future. Remember, documentation about your data is part of your data.

## Requirements and how to meet them

Your research community, institution, or research group (e.g. lab) may have specific standards and requirements about how you should prepare your data and document your activities.

If you are unsure about what procedures apply to your data, check against your data management plan, your research group's existing protocols and practices, and any requirements set forth by the places you want to use to share or publish your work.

## Things to think about

- Whenever possible, maintain a copy of your data in its original form. The link between the original and prepared data should be clear. If you generate new things, they should fit into your existing schemes for organizing and saving.
- Whenever possible, save any intermediate steps as you prepare your data. This will make it easier to trace back to what you did last. Doing this can be as simple as assigning different file names to different steps or as advanced as incorporating a version control system.
- Preparing data may affect risks related to sensitive data or personally identifying data. You need to be aware of this, but it should not affect the degree to which you document your procedures.
- Don't make assumptions. Even if you automate your preparation, you may still want to do manual quality assurance checks. Even if your decisions seem obvious, you should still document what you did and why.