Eight Tips for Handling Secondary Data

Have you identified any pre-existing data that could be relevant to your project? When reusing someone else's data, it's crucial to follow key steps to ensure proper documentation and its provenance. This includes detailing its origin, context, and lineage, which helps maintain transparency and traceability throughout your work.

VERIFY THE DATA SOURCE'S CREDIBILITY

Check the reputation of the data provider and authors, including their qualifications, affiliations, and any potential biases. If available, check peer reviews, citations, or previous uses of the data to gauge its credibility.

CHECK LICENSES & TERMS

Verify that your intended use is permitted to avoid unnecessary complications and legal issues. Carefully review the terms and licensing conditions, and if any restrictions apply, make sure to address them before proceeding.

REVIEW DOCUMENTATION

Thoroughly read the accompanying documentation, including README files and metadata, to understand the data's structure, limitations, and context. Examine the data and identify any unclear aspects. To avoid making assumptions, reach out to the data provider for clarification, if possible.
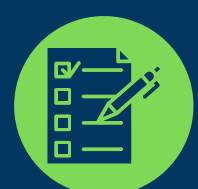
ASSESS ITS FIT FOR PURPOSE

Evaluate whether the data aligns with your research objectives by assessing its relevance, scope, variables, methods for obtention, and granularity to ensure it meets the specific needs of your project.

KEEP A COPY OF THE ORIGINAL DATASET

Store the data safely, separating the original dataset from working copies, and locking it to prevent accidental alterations and allow for traceability.

RECORD DATA PREPARATION STEPS

Document all data preparation steps before analysis, including cleaning (e.g., handling missing values, outliers), correcting errors, transforming variables (e.g., normalization, encoding), and removing unnecessary data. This ensures transparency, enables reproducibility, and provides a clear record of modifications to the original dataset.

CITE PROPERLY

Give appropriate credit to the original data source, following citation guidelines or requirements from the provider. Include a formal citation in the reference list of your associated publications.

CONSIDER RESHARING THE ORIGINAL DATA

Evaluate redistributing the original dataset, if permitted and not already stored in a secure repository. This will ensure continued access, even if the original data is moved or becomes inaccessible.

Want to learn more?
rds@library.ucsb.edu

UC **SANTA BARBARA**

Library

www.library.ucsb.edu