ERIM

Erasmus
Research Institute
of Management

Basic principles for responsible research data management

Created by: Finn Wynstra
Updated by: Monique van Donzel, Miriam Braskova

Erasmus University Rotterdam

ERiM

Erasmus
Research Institute
of Management

Basic principles for responsible research data management

Researchers, project leaders or principal investigators are primarily responsible for storing, managing, sharing and archiving research data according to ERIM's principles for research data management, which are outlined below.

1. Original data

Keep a copy of your original, raw research data in a safe place. Although the nature and form of raw data may vary, the same basic principle applies: researchers should be able to demonstrate that an original version of raw research data has not undergone any selection, purification or transformation.

- If you use **pen and paper** questionnaires, this means storing the paper responses.
- For **electronic data**, it means storing the original completed electronic forms.
- For **qualitative research**, it means storing the original audio files, transcripts of interviews, or field notes.
- For **secondary data** or data collected by others, it means storing the data you originally obtained (if permitted in the rules for data ownership).
- If you **harvest data from databases**, then store the search queries that you use; also make a note of when you accessed the data – the time and date.

2. Collecting data

Describe the process of **collecting data** clearly.

- Include the names and roles of researchers involved and/or organisations providing data, such as research agencies.
- Descriptions should be detailed enough to trace the process back.
- Best practice is to develop a data manual or code book in which you record in real time your decisions about data collection and the steps you take in your analyses. If you write your methods section long after you collect and analyse your data, your recollection of the process may be unreliable.

3. Data input and analysis

Document your **data input** and **analysis procedure** in detail, so that the analysis can be replicated exactly.

- Document each data input and step of analysis as soon as possible after performing it. This includes major analysis steps that you may never report or publish, but which are instrumental in steering the analysis process.
- Store all substantial files, including specific software syntax, diagrams, and graphical presentations.
- Record the names and roles of the researchers involved too.

4. Data compilation, purification and transformation

Store a clearly described and identified data set for every crucial **data compilation**, **purification** or **transformation** step. This is because such steps make it impossible to revert from transformed data to a more raw or previous version. This particularly applies to steps that lead to potential attrition of samples such as removing outliers (e.g., Winsorizing) or altering scores for variables.

5. Keep your data for 10 years

Store all original, raw data, plus documentation of the process of data collection and analysis (your log files) for a minimum of 10 years after the most recent publication that uses this data. This applies unless a longer storage period is required by the professional organisation, funding organisation or journal.

6. Store your contribution at the time of publication

- Store your complete **data set**, including raw data, metadata and analysis log files securely at the time of publication of the article or contribution it is based on, or preferably before publication.
- For PhD dissertations, store underlying datasets securely before your defence – if it has not already been stored as a requirement of a particular journal or funding organisation.
- For research master theses, store your complete data sets before your graduation.

7. Co-authored papers

- For **co-authored papers** for which someone else is collecting the data, doing the input and/or analysis, we recommend you store your own copy of the raw data, if confidentiality and data ownership rules permit it.
- You should also store documentation of data collection, data input and data analysis procedures.

8. Datasets with personal data

- **Anonymise** datasets containing **personal data**, if possible.
- If personal data is needed for your research, then **pseudonymize** the data and store the key file in a secure location.
- You must ensure privacy for the people whose personal data is in your datasets. Take appropriate mitigating measures; ask the ICT department for more information.
- If your dataset contains **sensitive personal information**, work with a privacy officer to conduct a Privacy Impact Assessment to ensure compliance.

9. Sharing datasets

Make sure you have the correct contracts in place if you share datasets that include personal data outside of the university, or with other researchers, companies, or tools. Always contact a privacy officer or legal counsel to ensure compliance.

A note on personal data

Personal data is defined by the EU as any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data. Fully anonymized data is not considered personal information.

Sensitive personal information relates to racial or ethnic origin (including nationality), political opinions, religious or philosophical beliefs, trade union memberships, genetic data, biometric data or data concerning health, sex life, or sexual orientation. Such information includes photographic or video material.

For the EU definition of Personal Data, see here: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en