

# Basic Principles for Responsible Research Data Management



© ERIM  
January 2018

Author: Monique van Donzel  
Updated by: Miriam Braskova, August 2018

## Where Are You with Your Data Management?

ERIM adheres to the highest standard in personal research integrity and quality of research. Responsible data management practice is becoming an important research skill. ERIM focus on providing advice and support on all aspects of research data management to ERIM researchers. It is expected from all ERIM researchers that they are knowledgeable of all applicable laws, codes of conduct, and standards (including ethical) regarding their work.

## ERIM Principles for Responsible Data Management

The researcher / project leader / principal investigator is primarily responsible for storing, managing, sharing and archiving research data in line with the ERIM principles for research data management outlined below.

1. Always maintain copies of the original, "raw" research data. In case of **paper and pencil** questionnaires, this means storing the actual forms. In case of **electronic data**, it means the original completed electronic forms. In case of **qualitative research**, it means the original audio files or transcripts of interviews, or field notes. In case of **secondary data** or data collected by others, it means the originally obtained data (data ownership issues permitting). Also store the specific search queries used when harvesting data from databases, in connection with the time of access. Thus, while the nature and form of the actual raw data may vary, the basic principle applies that the researcher should be able to convincingly demonstrate that this original version of the raw research data has not yet undergone any selection, purification or transformation steps.
2. The **data collection process** should be clearly described. This includes the names and roles of the researchers involved and/or the organisations providing the data (such as research agencies). The descriptions should be detailed to the extent that the process can be fully traced back. Methods sections in articles are often written long after the data is collected and analysed. Best practice recommendation is to develop a data manual or code book in which data collection decisions and analysis steps are recorded in real time.
3. The **data input and analysis procedure** should be documented in detail, so that the analysis can be replicated exactly. The documentation concerning each data input and analysis step should be documented as soon as possible after performing the step. This includes major analysis steps that may in the end not be reported in further publications, but which have been instrumental in steering the analysis process. All substantial files should be stored, including for instance specific software syntax, diagrams, graphical presentations etcetera. Again, the names and roles of the researchers involved should be provided.
4. For each crucial **data compilation, purification or transformation** step, it is recommended that clearly identifiable and described data sets are stored. (Crucial steps transform data such that it is impossible to revert to the rawer data when only the transformed data is available.) This particularly applies to steps leading to potential sample attrition (e.g., outlier removal) or the alteration of scores on variables (e.g., Winsorizing).
5. All original, "raw" data and the documentation of the data collection and analysis process (log files) should be stored for a minimum of ten years after publication of the most recent publication using this data. This applies as long as specific professional or journal policies or funding organisations do not require a longer storage period.

6. The complete **data set**, including the “raw” data, the metadata and analysis log files should at the latest be stored at the time of publication of the article or contribution it is based on. This also applies to the underlying dataset(s) for dissertations by the time of the PhD defence, in as far as it is not yet stored and required by specific or journal policies or funding organisations. For Research Master’s theses, the complete data set(s) underlying the needs to be stored at the time of the MSc graduation.
7. In the case of **co-authored papers** where another person is executing data collection, input and/or analysis, we recommend storing a copy of the raw data yourself as well (confidentiality and data ownership issues permitting), and storing the documentation of data collection, input and data analysis procedures.
8. If the dataset contains **personal data**, it should be anonymized if possible. If the personal data is needed for the research, pseudonymize the data and store the key file in a secure location. If your datasets contain any personal data, make sure mitigating measures have been taken to ensure the privacy of the persons in the dataset. If your dataset contains **sensitive personal information**, conduct a Privacy Impact Assessment together with a Privacy Officer to ensure compliancy.
9. If you share datasets which include personal data outside of the university, including other researchers, companies and/or tools, please ensure you have the correct contracts in place. Always contact a Privacy Officer or Legal Counsel to ensure compliance.

## Short Remark on Personal Data

**Personal information** is defined as any information relating to an identified or identifiable natural person. 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 membership, genetic data, biometric data or data concerning health sex life or sexual orientation. This includes photographic or video material.