

From “Letter about scientific integrity” by Professor Marno Verbeek to ERIM fellows, members and doctoral students, 04/07/2012 (reference MV/tv 0012.003840), and applied by Stefan Eriksson to Swedish research. The document has continuously been revised by PhD students participating in SE’s course in Research Integrity for Natural Science & Technology at Uppsala University. It promotes re-use of data, the generation of better datasets, visibility, credibility and more.

General recommendations for handling research data (version 2.1)

1. *Raw data*: If possible, maintain the original, raw data (including laboratory notebooks, samples, specimens, photographs, scanned papers, etc.); or, if not possible, document them so that the researcher and/or data collecting facility are able to convincingly demonstrate that the original research data has not yet undergone any selection, purification or transformation steps.
2. The *data collection* process should be clearly described in the research records through meta-data. This includes aim, equipment (logs), software, simulations and corrections made, primary reference results, etc., but also the dates, names and roles of the researchers involved and/or organizations providing data (such as research agencies or collaborators). Make descriptions detailed so that it is possible to trace the process fully back, or to reproduce it.
3. *Analysis*: The data input, analysis procedure and troubleshooting should be documented in detail, so that the analysis or simulation can be replicated exactly. It is good practice to provide any code written especially for the data analysis alongside with its documentation. Use of freely available software are encouraged and version number must be provided. Store identifiable and fully described data sets for each crucial data compilation, purification or transformation step.*
4. All raw data and the documentation of the data collection, input and analysis process should be *stored* safely, electronic data on a central server with backup or in duplicate to prevent accidental deletion. No original data should be removed from the research body without explicit permission from authorized officials. Only use authorized cloud solutions.
5. Data that are e.g. sensitive or protected by secrecy should be *securely* handled (by measures such as ‘pseudonymization’, logging access, encryption or password protection).
6. Data should be preserved according to national regulations on *archiving*, and, if possible, kept in a standardized format that facilitates the aggregation and re-use of data. For examples, see <https://www.loc.gov/preservation/resources/rfs/TOC.html>
7. To ensure *public access* to data: describe the data set, and order and list data as well as the accompanying research records in such a way that a layperson can understand what is archived.
8. *Open data*: It is good practice to register or publish data sets in open repositories (such as [Dryad](#), [GDC](#) or [dbGap](#)) when or after reporting results, or after filing a patent application.
9. Original data and research records should only ever be *deleted* for justified reasons and after results have been made public, and reasonable time has passed for verification of research results (often at least 10 years after publication is believed to be reasonable).
10. It is also good practice to use resolvable and persistent identifiers such as e.g. *digital object identifiers* (DOIs) and make use of them e.g. in citations or on a CV, in order to distinguish the data and make them traceable (www.doi.org).

* Crucial steps transform data such that it is impossible to revert to the rawer data when only the transformed data is available.