Checklist for Quality Assurance and Quality Control

An essential part of Research Data Management (RDM)

prior to, during and *after* **a research project.**

1. Why RDM?
The Erasmus University Rotterdam provides research data management services to its research staff for three reasons:

1. Understanding the data and the context. It results in documentation of key aspects during the whole research process, thus providing, in a practical sense, the context to understand the data, and the relations between them, in the course of time.
2. Transparancy. Documenting key aspects during the whole research process provides transparency, for instance for reproduction and validation purposes and even in cases where questions have arisen with regards to scientific integrity.
3. Research funding. Research grant suppliers and funding agencies increasingly demand research management plans and data paragraphs from researchers and check the submitted grant proposals on aspects of planned and embedded strategies to ensure that research data is archived in a sustainable manner and made available for further research.

Data management is identified as a crucial aspect of fostering academic professionalism and integrity in research.

[Taskforce Scientific Integrity, 2013]

Responsible data management and data storage is part of a professional approach, it is not considered as 'additional'.

[Letter by the RM (dated December 6, 2013) addressed to the academic staff stating that all recommendations formulated by the *Taskforce Scientific Integrity* in its Report have been adopted by the Executive Board and shall be implemented.]

RDM basics: doing what, why, for whom?
Why is data stored? For checks by yourself? Is access to the data required for specific software? Should others have online access to the data? During which defined period?

See the figure below for an example of different scenarios of storing data related to different content, which will specify the specific required RDM services:

Content	Purpose of storage		Access by whom?		
Digital data	**Access**	**Archive**	You	The research group	Everyone
Digital Data and Software	**Access**	**Archive**	You	The research group	Everyone
Digital Data, Software and Codebook or Labjournal	**Access**	**Archive**	You	The research group	Everyone

ERASMUS UNIVERSITEIT ROTTERDAM

2. RDM – the 5 basic questions

When preparing a research data management plan the following 5 questions should be addressed:

1. What data will you produce?
2. How will you organise the data?
3. Can you/others understand the data
4. What data will be deposited and where?
5. Who will be interested in re-using the data?

The answers to these questions will specify the specific required RDM services. See the checklist[1].

3. RDM - Quality Assurance[2]
Quality Assurance and Quality Control should occur throughout the data lifecycle. Budget time and funds for QA/QC. General approach to QA/QC: (1) Develop a QA/QC plan.
(2) Document QA/QC procedures and follow them
(3) Automate QA/QC where possible.

3.1. Before you start Collecting Data:

3.1.1. Which standards apply to the data?

1. **Format** - Decide the format of how the data will be collected. Will the data be collected by hand on paper? Electronically via an instrument? If the data are digital, what format will be used? Is the format supported by the software you intend to use or is conversion required? Is the format supported by the desired digital archiving service, or is conversion required?

2. **Codes** - Define what each code word means and document these codes and corresponding meanings in a digital codebook.

3. **Specify units of measurement**. Furthermore, define and document the rules for calculation (eg. rounding of numerical values; when and when not and how).

4. **Metadata** - Be sure to create metadata in unison with the data to be collected. Which metadata can already be generated automatically and which metadata are generated manually? Will the metadata be used by third parties to access and / or analyse the data and if so, are the metadata descriptions made understandable for third parties in general? Have a review of the metadata, performed by someone not familiar with the data and format.

5. **Software** – Is the data easily accessible by the intended software and is it easily exported in standard formats or is the data locked in specific software? Document the used version(s) of the software for

3.1.2. Assign responsibility to a person over quality assurance

---

[1] See also Appendix 1: M. Domingus, *Checklist for a Data Management Plan*, v.0.2. Research Support Office EUR.
[2] Based on the best practices of RDM Quality Assurance: http://www.usgs.gov/datamanagement/qaqc.php

ERASMUS UNIVERSITEIT ROTTERDAM

3.2. QA/QC During Data Entry

3.2.1. Data entry.

1. Have more than one person independently enter and / or check the entered data.
2. Use a (documented) procedure for data entry. Examples: (1) use consistent terminology (documented) to enter data, (2) minimize the number of times the data entry. (3) reduce data to one piece of information per cell.

3.2.2. Scope of data entered. Store as much stages in the data cycle.

1. Example: record a reading of the data and transcribe from the recording. OR: record interviews (video / audio), store the video- and / or audio files of interviews, store the transcriptions, store the translations of the transcriptions and describe the whole in relevant metadata.

3.2.3. Use a (documented) procedure for QC during data entry. If possible use software to support the QC tasks.

1. Example: use software for checks (consistency, duplicates) during data entry. Example: use a text-to-speech program to read the data back.
2. Always document any modifications to the dataset. This avoids duplicate error checking.

3.2.4. Use an efficient storage system for the data, fit for the intended use.

1. Examples: (1) safe storage of raw data on the EUR research data server (files in folders with a readme file which describes the files in the folder and the relation(s) between them),(2) safe storage in an online storage service of snapshots of live datasets (examples: (1) files stored in a co-creation environment such as a virtual research environment or (2) files stored in a dataset repository such as the Dutch DataVerse Network by DANS)
2. Access rights to the storage system should be defined and controlled, preferably based on the institution accounts of the researchers. Access rights should be defined based on roles (read / edit) and to specific folders and files. Access rights should be defined for an agreed period of time (usually related to the role of the researcher or the lifecycle of a research project).
3. In case of sensitive data, contact research support for advice on security and legal matters.

Checklist QA / QC RDM
M. Domingus, v.0.2. Research Support Office EUR.

ERASMUS UNIVERSITEIT ROTTERDAM

3.3. QC After Data Entry

3.3.1. Make sure data columns and rows line up properly.

3.3.2. Look for missing or irregular data entries.

3.3.3. Perform statistical summaries.

3.3.4. Check for outliers. This is important because if outliers are found, their presence may be due to a mistake from some level of "contamination" from the data collection or data entry. The outlier may not be a mistake at all but it is still important to check outliers for quality control and assurance. Use the following methods to look for outliers.

- Graphical methods - Normal probability plots, regression, scatterplots.
- Maps.
- Subtract values from the mean.