**Research Data Management paragraph - Radboudumc Theses**

**Radboudumc Template** **v1.0 May 2024**

**Need help?**

For support in drafting your data management paragraph or feedback, please get in touch with RTC Data Stewardship (datastewardship.im@radboudumc.nl)

In this document we provide you with guidelines and examples on what to mention in the Research Data Management paragraph of your thesis.

The following sections should be included in the paragraph:

1. **Ethics and privacy**

Provide specifics on ethical approval, funding and on how privacy of the participants has been

warranted. This includes information on:

* Medical and ethical approval of the study(-ies), if applicable.

***Example texts:***

* + *This thesis is based on the results of research involving human participants (or existing data from published papers), which were conducted in accordance with relevant national and international legislation and regulations, guidelines, codes of conduct and Radboudumc policy.*
	+ *The recognized Medical Ethics Review Committee ‘METC Oost-Nederland’ has given approval to conduct these studies (file numbers: NL00000.000.00; NL00000.000.00).*
	+ *The institutional ethical review committee CMO Radboudumc, Nijmegen, the Netherlands has given approval to conduct these studies ( CMO Radboudumc dossier number: 2019-xxxx, 2020-yyyy, 2022-zzzz).*
	+ *A statement that the study was not subject to the Dutch Medical Research Involving Human Subjects Act (WMO), was obtained from the recognized Medical Ethics Review Committee ‘METC Oost-Nederland' (YYYY-0000).*
* How privacy of the participants was safeguarded, for instance by pseudonymization/anonymization, or other measures:

***Example texts:***

* + *According to Dutch legislation, data collection from electronic patient files was performed by personnel with a treatment relationship with the patient AND/OR by the researcher(s) upon consent by the study participant.*
	+ *The privacy of the participants in these studies was warranted by the use of pseudonymization.*
	+ *The pseudonymization key was stored on a secured network drive that was only accessible to members of the project who needed access to it because of their role within the project. The pseudonymization key was stored separately from the research data.*
	+ *The privacy of the participants in these studies was warranted by the use of the pseudonymization tool PIMS (*[*PIMS (radboudumc.nl)*](https://pims.radboudumc.nl/)*).*
	+ *The privacy of the participants in these studies was warranted by the use of fully anonymous data.*
* If informed consent was obtained for collecting and processing the data from participants, and if yes, if this consent also covers sharing data after research for re-use.

***Example texts:***

* + *Informed consent was obtained from participants to collect and process their data for this research project.*
	+ *Consent was also obtained for sharing the (pseudonymized) data after research.*
	+ *The sensitivity and confidentiality of the raw qualitative data (i.e. interviews, forum groups) makes sharing of the data without compromising confidentiality and privacy impossible, therefore consent for sharing of the raw data was not asked from the participants. Optional: Where possible, the raw qualitative data will be anonymized by data aggregation to enable sharing for reuse.*
	+ *For chapter 4 data was used that was previously collected in the context of healthcare. To ensure responsible reuse of healthcare data, specific informed consent procedures were followed that are aligned with applicable laws, regulations and the national Code of Conduct for Health Research.*
1. **Data collection and storage**

Describe as detailed as possible how the data were collected or created, and where these data were stored. This included information about:

* Whether you created/collected your own data or that you re-used data from other sources
* The methods used, for example care data from the Electronic Health Records (EHR; e.g. Epic, Dentium), data from existing biomaterial, external data sources (e.g. CBS, PALGA, IKNL, EGA, etc), data from lab experiments, devices, questionnaires, interviews, literary review, etc.
* Which systems or tools were used to collect and store these data, and where these data were stored during research.

***Example texts:***

* + *Data for chapter 2 and 3 was obtained through laboratory experiments involving anonymous or non-human materials and/or from experiments involving animals.*
	+ *Data for chapter 2 and 3 was extracted, by using the Cliniquest application (*[*CliniQuest (radboudumc.nl)*](https://cliniquest.radboudumc.nl/)*), from (electronic) health records (EPIC).*
	+ *Data for chapter 2 and 3 was obtained by using Castor EDC for secured online questionnaires.*
	+ *Data for chapter 4 was collected through electronic Case Report Forms (eCRF) of a prospective data collection in Castor EDC. Data were converged from (electronic) health records or Castor EDC to SPSS (SPSS Inc., Chicago, Illinois, USA).*
	+ *Raw data from patient devices / recording devices / laboratory systems are archived in their original form in a Data Acquisition Collection (DAC) in the Radboud Data Repository*\*
	+ *Processed data and documentation (research protocol, experimental setup, codebook, software versions and a readme file) are archived in a Research Documentation Collection (RDC) in the Radboud Data Repository*\*
	+ *Data from chapter XX were stored and analyzed in workspace (WORKSPXXX) the Azure DRE (*[*DRE Portal (mydre.org)*](https://mydre.org/)*).*
	+ *Data from chapter XX were stored and analyzed on the department server and in Castor EDC and are only accessible by project members working at the Radboudumc.*
	+ *These secure storage options safeguard the availability, integrity and confidentiality of the data.*
	+ *Paper (hardcopy) data is stored in cabinets on the department.*
1. **Data sharing according to the FAIR principles**

Corresponds to: *Findability*, *Accessibility*, *Interoperability*, and *Reusability of data/research outputs*

In this section you should mention in which (trusted) repository you will store your data after your research and how this data complies to the FAIR principles. These principles most clearly apply to your dataset when you publish your data open access (i.e., open to the public), however there may be valid reasons not to make all your data available for reuse, or at least not right away. Perhaps you do not have the ownership, or your data might be privacy- or competition-sensitive or confidential. In most research projects data can be made available for reuse, but if this is not the case, please explain clearly why. Also, this does not mean that the FAIR principles do not apply to these types of datasets. For example, if you use a repository with restricted access, the dataset can still receive a persistent identifier and the metadata can be visible and indexed by search engines, being “as open as possible, as restricted as needed”.

So, whether you publish open access or not, the following points regarding the FAIR principles can be mentioned in your data management paragraph:

**Findable and Accessible**: How can others easily find and access your dataset, and how will the accessibility be regulated?

* Repository: Start out by mentioning the name of the data repository you will use for long-term archiving and publishing.
* Persistent identifier: Does the repository that you have chosen assign a persistent identifier (e.g., a DOI) to your dataset? That way your dataset is uniquely and persistently identifiable. If so, you could mention the identifier.
* Metadata: Metadata describe the data, for example by keywords, a summary, authors, codebooks etc. This can be done by the repository, but also by you. In case of the former: often, repositories with a certificate adhere to specific metadata standards. You can mention these standards. In case of the latter: you can add documentation on the content, context and structure of the dataset.
* Intellectual Property Right (IPR) considerations: In general, when you are employed at the Radboudumc, the UMC is the legal owner of the data. You can mention this. However, in some circumstances other arrangements have been made regarding the property rights. For example, some consortia have documents stating who owns the data. In that case, mention the arrangements made and who has the intellectual property rights.
* Mention any data use agreements or licences under which you will share your data.
* Timeline: When you decide to share the data, mention when they are accessible. For example, at the end of your project or when a corresponding article is published.

**Interoperable and Reusable**: Is it possible for people and computers to interpret the data and combine it with other datasets, and how will you make your data ready for (re-)use by others?

* Besides proper documentation (see also the third bullet under findability) make use of standard vocabularies, ontologies and thesauri in your (meta)data, or provide mapping of your data to these vocabularies, ontologies and thesauri. Whether these are available and which you choose to use depends on your research discipline. Examples are SNOMED for health care data or RefSeq accession numbers for gene sequences.
* Mention that you have used interoperable file formats where possible. In general, most preferred interoperable file formats are those that are widely used and free to access. For example, .odt (Open Document Text) or .pdf has a preference over .docx files, seeing that Microsoft Word is no free software and possibly not available to everybody (now or in the future).
* Mention any data use agreements (DUA) or licences under which you will share your data. For example, the CC BY 4.0 licence is a commonly used license. For sensitive data, such as pseudonymized data, the CC BY-NC license is recommended. With this license, the data may only be reused for the purpose of scientific research, by an external not-for-profit party. Like many other repositories, the Radboud Data Repository has standard licences that will accompany your dataset.
* Software: When you work with specific software, for example for your analyses, explain how and where the software is available. If the software is not commonly available, describe how you will deal with that. If you don't use special software, mention that all the data can be opened with generally available software tools.

***Example texts:***

* + *All studies are published open access.*
	+ *The data underlying the published chapters XX are available for reuse through the following data repositories: .*
	+ *The datasets from chapters 3 and 4 are published in Data Sharing Collections (DSC’s) in the Radboud Data Repository\**
	+ *The dataset from chapters 5 is published in the DANS Data Station Life, Health and Medical Sciences (DOI: 10.17326/dans-xhj-pxyt).*
	+ *Chapter 6 is based on existing data, which was obtained from [original authors] and available for reuse via DOI.XXX / upon request from the original authors.*
	+ *Gene expression data of chapter XX is available for the patients treated with short-term presurgical raloxifen in EGA (Dataset ID* [*EGAD000XXXXXXXX*](https://ega-archive.org/datasets/EGAD00000000038)*)*
	+ *Sequencing data are shared through the NCBI in the Sequencing Read Archive, Accession: SRX24727555.*
	+ *The proteomics data generated in this project are published in the PRIDE repository*, *project identifier PXD052592* *.*
	+ *Data were made reusable by adding sufficient documentation (research protocol, codebook and a readme file), by using preferred and sustainable data formats and by publishing under the CC.BY.4.*0 license.
	+ *The datasets from chapters 3, 4 and 6 were published with restricted access. Requests for access will be checked by [list the persons that are responsible for the dataset, e.g. the PI, data manager/data steward of the department, Data Access Committee, etc], against the conditions for sharing the data as described in the signed Informed Consent.*
	+ *The data underlying chapter 2 will be published without restrictions, only after an embargo period of 12 months to enable publication of new results based on the data.*
	+ *The data underlying chapter 7 is not suitable for reuse and will be archived for 15 years in DACs and RDCs of the Radboud Data Repository after termination of the study.*

  *\*The table below details where the data and research documentation for each chapter can be found on the Radboud Data Repository (RDR). All data archived as a Data Sharing Collection remain available for at least 15 years after termination of the studies.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Chapter* | *DAC* | *RDC* | ***DSC*** | ***DSC License*** |
| *1* | *DOI:* |  | *DOI:* | *CC-BY-NC* |
| *2* |  |  |  | *RUMC-RA-DUA-1.0* |
| *4* |  |  |  | *RUMC-RA-DUA-1.0* |
| *5* |  |  |  | *RUMC-RA-DUA-1.0* |

*DAC = Data Acquisition Collection, RDC = Research Documentation Collection, DSC = Data Sharing Collection*