Research data management policy
Department of Biomechancial Engineering

Authors: M. Vlutters, S.S. Fricke, E.H.F. van Asseldonk
Version 220615

Table of contents

List of abbreviations and definitions

Abbreviations	
BE	Biomechanical Engineering
DMP	Data management plan
ET	Engineering Technology
IP	Intellectual property
PI	Principal investigator. The person in charge of the research, such as a research chair or a grant-holder.
RDM	Research data management
SOP	Standard operating procedure
UT	University of Twente

Definitions	
Anonymization	Irreversible de-identification of data, stripping data of all personal info
P-drive	UT network drive for shared group storage
Personal data	Any data related to an identified or identifiable person
Pseudonymization	Reversible de-identification of data, for example with subject identifiers

1. Introduction

Regulations with respect to research data management (RDM) are specified at various levels: UT level, ET level, department level, research chair level, and individual researcher level. This document is intended as an addendum to the ET faculty RDM regulations (*ET/A-20.19061, v1.1, 27-05-2021*), with the purpose of defining department-level RDM regulations for Biomechanical Engineering (BE). All content in this document are **BE department-specific** rules and recommendations, which provide the boundaries within which a data management plan (DMP) has to be formulated as part of research (see section 3). A distinction is made between **rules** and **recommendations**. Rules must be adhered to by group members. Recommendations are a set of suggestions to promote uniformity in data management across group members. In addition, a distinction is made between **non-personal data** and **personal data** (see section 4).
All content laid out in higher-level regulations apply to this document, unless indicated otherwise. Further specifications might be made on research chair or individual researcher level. Such additional specifications need to comply with the higher-level regulations. Any (proposals for) deviations must be communicated to, and approved by, representatives up the chain.
BE-level: see authors of this document. ET-level: see **here**. UT-level: see **here**.

2. Department-specific roles and responsibilities

The roles and responsibilities are described in appendix A.

3. Data management plan

For every research project a DMP should be formulated, which describes how data that is generated in the research will be handled. Most research funding bodies also require a DMP to be provided as part of the funding application, or shortly after the start of the project.

*Data management plan – BE working **rules***
Roles and responsibilities with regard to DMPs as described in appendix 3 of the general UT-policy are leading.

1. A DMP must be reviewed and approved by the principal investigator (PI). Review feedback must be considered by the DMP authors.
2. A DMP must refer to the latest BE department regulation document, or lower-level regulations (e.g. chair level) with which is has to comply.
3. A DMP, as well as the actual data management resulting from it, must be updated if the DMP would no longer comply with regulations due to policy changes on any level.
4. A DMP itself should not contain personal or sensitive data. If it does, it should be treated as any other personal data and be stored accordingly. A DMP stored in the UT's DMP tool is not private, and has a backdoor to ICT staff and data stewards at the UT.

4. Privacy regulations

BE may handle personal data. Personal data are any data related to an identified or identifiable person, including name, birth date, address, contact information, social security numbers, and/or photographs. Anonymization of personal data means *irreversible* de-identification: once personal data has been stripped of identifying data, it is no longer possible to trace back the person(s) from which the data originated. Pseudonymization means *reversible* de-identification: personal data are replaced by a key, making it possible to trace back the person(s) of which the data originated.

*Privacy regulations – BE working **rules***

1. Personal data are handled according to the latest UT and ET policies. In general, collection of personal data should be in proportion to the intended purpose of the research. Any personal data must be anonymized or pseudonimyzed as soon as possible. Please refer to corresponding UT and ET policies for details. For UT-level recommendations, see **here**.
2. For any storage medium, only the researcher(s) and principle investigator(s) may have access to personal data.
3. For personal storage media such as Microsoft OneDrive, the individual user is responsible for access management of personal data. For archiving, the inclusion of personal data should be prevented. If this is not possible, the archived data must be encrypted.
4. Physical material that contains personal data has material-dependent regulations. Also see section 8.

5. Data storage and transfer

Data storage concerns all storage *during* research, including both personal and non-personal data. Data should be stored such that the risk of data loss is minimized, and data integrity is maintained. BE takes measures to avoid loss of research data during the course of a research project.

*Data storage – BE working **rules***

The following storage rules apply to **non-personal data**.

1. Data is made public under an appropriate license unless specific exceptions apply. There is no “ownership” to public data: it is not the property of the original author or the modifying authors.
2. Data must be in a storage medium that supports sharing. An overview of allowed storage media can be found **here**. If alternatives to these storage options are desired, they need to be communicated to ET, and screened by an ISO certification auditor at LISA.
3. To counter potential data loss in case of account termination, certain data may not be owned by only one individual (see appendix C). Access rights to such data must be granted to the group’s secretary office, and if applicable, to the head of the research chair, or the daily supervisor. This is required for at least the following data (also see section 6.8):
   |—***dissemination***: papers, posters, workshops, reviews, …
   |—***education***: lecture slides, assignment descriptions, exams, ...
   |—***experimental***: experiments, protocols, ethics, raw and processed data, subject info, …
   |—***group***: BE specific data, templates, SOPs, group meeting notes, decision records, …
   |—***hardware***: developed hardware specification, requirements, CAD drawings, …
   |—***software***: developed code, data processing, simulation, modeling, …
4. For every version of data, only a single representation of the data may function as (master) source files. Data may not exist as a duplicate on one or multiple storage services, unless the duplicate is used as an active work copy.
5. BE-related data on computers of students must be stored on- and automatically synchronized with a Microsoft Sharepoint (also see Appendix C). The shared folder must be under administration of the group. The folder should be provided to the student through the student supervisor at the beginning of a bachelor or master assignment. The drive folder must be shared with the daily supervisor and the BE secretary.
   Never use personal (non-UT) Microsoft accounts. Agreements between the UT and Microsoft (e.g. GDPR) do not apply when personal accounts are used. When students leave, data remains in the OneDrive of the supervisor, or should be moved to Areda (section 8) when archiving is desired. This is the responsibility of the administrators of the shared OneDrive.

*Data storage – BE **recommendations***
BE recommends the use of the below guidelines for **non-personal data**.

1. UT-related data on employees' personal computers is stored on- and automatically synchronized with Microsoft OneDrive and Microsoft Sharepoint, available to UT staff and students. Work-related data is distributed over 10 root folders (see section 6.8).
2. Code is stored on the group's Bitbucket repository, and is version-controlled with Git.

*Personal data storage – BE **rules***
The following storage rules apply to **personal data**, **in addition** to the rules for non-personal data.

1. Personal data must be stored on OneDrive, with the correct access restrictions (section 4).
2. Storage of personal data on non-encrypted and non-password protected devices, including personal computers, measurement devices, USB drives, and external hard drives, is forbidden. Exceptions apply to cases where data can only be generated on an unsafe storage medium (e.g. the sd-card of a video camera). Personal data should be removed from such storage immediately after collection. Users have to check if the material has been removed, and complete a form declaring that they have done so upon completing measurements and/or returning equipment to the corresponding equipment manager.
3. Storage of personal data on non-personal devices, such as computers that are part of a measurement setup, must be prevented as much as possible.
   Example: a motion capture system might ask for the subjects name, height, and weight. However, these data are not required for the system to function, and should therefore not be entered.
   Example: a body-weight support system might not function without entering the subject's body weight. This is allowed, as long as the weight cannot be linked to other personal information of the subject, such as a name. Subject identifiers should be used instead.

6. Data documentation and metadata

Documentation is 'information about the research data'. Data must be documented in line with the FAIR principles (see appendix B of the ET policy). Metadata is 'data about data'. It is standardized, structured information that may describe the purpose, origin, time frame, geographic location, creator, access conditions and terms of use of a data set.

*Data documentation and metadata – BE working **rules***

1. Any data that is not immediately self-explanatory to another person *must* be documented. This holds for anything that is developed, including all terminology, produced hardware, experimental setups, executed procedures, experimental outcomes, simulations, and code.
2. Data, metadata, and documentation must be version controlled to track the creation and modification dates, the changes that were made, and who the modifying authors are. If no version control system (e.g. Git) is available, changes must be manually tracked in a separate change-log file that describes a version number and the changes at every "release".
3. Metadata and documentation should be such that another person in the same field of work is able to **understand**, **use**, **maintain**, **update**, and **depend** on the data to which the metadata and documentation refers, **without prior knowledge on the data content**.
4. Metadata and documentation should at least specify the **name**, the **purpose** ("what is it for? / what does it do?"), and the **description** ("how does it do that / how do I use it?"). Additionally, the **original author**, the **modifications**, and the **modifying authors** should be reported, if they are not already tracked by a version-control system or by the file system.
5. Documentation must be adjusted and updated if they no longer accurately describe the underlying content. Previous documentation versions may not be discarded, considering the

references to it that may exist. Incorrect metadata, documentation, references, and broken links must be reported to the first author, or otherwise one of the modifying authors, such that they can be updated.

6. Units must always be reported. Units must be **SI units** whenever possible.
7. Handwritten documents may not function as main source for documentation and metadata. Only indelible ink (e.g. no pencil) should be used for handwritten files intended for preservation.
8. **Folder- and file-structure** must comply with the BE root folder structure, within which data may be further subdivided in files and folders as desired. The root level may consist of at most the following folders:
   |—***dissemination***: papers, posters, workshops, reviews, …
   |—***education***: lecture slides, assignment descriptions, exams, ...
   |—***experimental***: experiments, protocols, ethics, raw and processed data, subject info, …
   |—***group***: BE specific data, templates, SOPs, group meeting notes, decision records, …
   |—***hardware***: developed hardware specification, requirements, CAD drawings, …
   |—***literature***: books, papers, reference lists. … Preferably link to external material.
   |—***software***: developed code, data processing, simulation, modeling, …
   |—***supervision***: information linked to a specific person, such as bachelor / master students
   |—***self***: any data related to you that does not fit any of the above
   |—***project***: any project-specific data that does not fit any of the above
   |—***(secretary***): secretary only
   These root folders may exist distributed over various storage media (e.g. hardware on OneDrive, software on Bitbucket), or exist on multiple storage media if their content is unique (non-duplicate content). Each of these folders has their own accompanying metadata and documentation, for each component in it (sub-folders, set of files, and/or individual files).
9. It is the responsibility of the person(s) producing data to further structure the data into sub-folders. When doing so, it must be considered that:
   1. A flat organization structure is preferred, with limited nested folders.
   2. A sub-folder relates to the context of its parent folder, and not to another folder in the root structure when possible. *Example*: the hardware folder may contain an "*example_device*" sub-folder, but the "*example_device*" folder may not contain another "*control_code*" sub-folder that relates to software.
   3. Sub-folders should not be moved out of their parent folder context.
   4. If data would relate to multiple of the root folders provided above, it should be placed in the one where it is most likely to be found, and cross-referenced to from the other(s). *Example*: code used for education purposes is stored in software, and is referred to from other material stored in the education folder.
   5. A sub-folder name is descriptive, relates to its content, and has a date when appropriate.

No:	***Suggestion:***
Experiment Mark	*201200_lateral_perturbed_gait*
Data subject 1	*201231_s001*

   6. Folder and file names do not contain special characters (~!@# , etc).
   7. Folder and file names do not contain leading or trailing spaces (" my_folder ")
   8. Files must have a file extension (.doc, .pdf, etc).

Data documentation and metadata – BE ***recommendations***

BE recommends the use of the below guidelines:

1. A metadata standard is selected to determine which aspects of the data need to be stored. An example is the Dublin Core, which is a domain-agnostic standard. Various metadata standards can be found **here**.
2. Folder- and file names that are expected to be used in automated (batch) processing may not contain spaces (e.g. experimental data files).
3. Folder, file- and variable names are written using consistent typography.
   ***No**: snake_case, CamelCase, Mixed_Case used interchangeably for folders, files, code, etc*
4. Date specifications are year-first. When year, month, or day is unknown or omitted, use "00" as replacement.
   ***No**: ddmmyy, mmddyy*
   ***Yes**: yymmdd, yyyymmdd, yymm00*
5. Time specifications are 24-hour format.
   ***No**: 11:45 PM, 2:30 min*
   ***Yes**: 23:45, 2 minutes and 30 seconds*
6. Ordered numbering has an appropriate number of leading zeros.
   ***No**: 1, 2, …, 10, 11, …100, 101*
   ***Yes**: 001, 002, …, 010, 011, … 100, 101*
7. Metadata and documentation is **DRY**: *"every piece of knowledge must have a single, unambiguous, authoritative representation within a system."*. That is, duplication of data, metadata, and documentation should be prevented as much as possible. If dependencies on other sources exist, refer to the correct version of the source's documentation, rather than duplicating it.

   ***No**:*

   perturbation_experiment
   - *201231_s001*
     - *experiment_description.md*
     - *201231_s001_info.md*
     - *201231_s001_trial0001.mat*
     - *201331_s001_trial0002.mat*
   - *210101_s002*
     - *experiment_description.md*
     - *201231_s002_info.md*
     - *201231_s002_trial0001.mat*
     - *201331_s002_trial0002.mat*

   ***Suggestion**:*

   201200_perturbed_gait
   - *experiment_description.md*
   - *201231_s001*
     - *subject_info.md*
     - *0001.mat*
     - *0002.mat*
   - *210101_s002*
     - *subject_info.md*
     - *0001.mat*
     - *0002.mat*

   If the experiment_description.md contains a general description of the experiment, it is better to not repeat it in each subject folder. This prevents changes to one but not the other file. The latter form furthermore does not repeat the experiment date and the subject number in its file names. By doing this, one is forced to keep the files in their respective parent folder hierarchy. These files may not be taken out of their parent folder's context.
8. In versioning, the version number itself has a meaning. Also see the rationale of **SemVer**. Preferably, it is clear from the version number and/or corresponding version change-log *what* has changed with the version increment. Exceptions can be e.g. word documents, though still a meaningful version number can be provided.

   ***No**:*

   thesis_v1.docx
   thesis_v2.docx
   thesis_vfinal.docx

   ***Suggestion**:*

   210101_thesis.docx
   210301_thesis.docx
   210312_thesis.docx
9. Page numbering is used whenever possible.

7. Data sharing

Data sharing focuses on sharing *during* the research. To guarantee that data can be accessed and checked during the research, digital and non-digital research data and related materials must be shared. Implementation is dependent on intellectual property and responsibilities regarding research data, and the terms of use of data suppliers.

*Data sharing – BE working **rules***

1. Data sharing within BE is an integral part of data storage. Also refer to section 5.
2. The options mentioned in section 5, specifically Microsoft OneDrive, is also used to share data with other groups within the University of Twente, or with external parties.
3. If data is shared, and the receiving party is not already part of the DMP, the DMP must be updated accordingly. For personal data, an update to the GDPR registration is required as well.
4. When sharing sensitive data (e.g. personal data, IP sensitive, commercial interests), approval of all persons having access to that data is required before further sharing. The shared data must be encrypted, with a key sent in a separate (email) message. An expiration date must be set at which the access by an external party will be revoked (automatically, or otherwise manually).

8. Data archiving

Data archiving concerns storage *after* a research project. Sustainable archiving of data and its access regulation is required to support open science and scientific integrity. At the end of each research project, research data and metadata are stored in a sustainable administered repository. At a minimum, this applies to all data used for published results, including dissertations, that cannot in its entirety be found in the publication itself or in the 'supplementary information' accompanying the publication. This refers to both experimental data and numerical models, where e.g. result files can be stored, or even all data required for reproduction.

*Data archiving – BE working **rules***

1. Areda is used for (internal) archiving data. When research is finished, as defined by the PI and/or individual researcher, data should be moved to the Areda archive, including any source data that is not associated with a publication.
2. Within Areda, the same top-level folder structure is used as specified in section 6.8. When possible, data should be archived following this structure. *Example*: if hardware is developed in a project, the hardware-related information and documentation should be archived separately from the project. For a rationale see appendix B.
3. The inclusion of personal data should be prevented. If this is not possible, the archived data must be encrypted.
4. Data must have corresponding documentation and metadata. Presence of documentation and metadata is checked by a data steward during the archiving process. A metadata form can be filled out through Pure, and linked to Areda.
5. Data must be compressed (.zip) before uploading to Areda
6. Externally available data and information should not be archived, such as third-party data or software installation files. Instead, the documentation must refer to such external data. An exception may apply when the usability of archived information is highly dependent on the external data, *and* there is an indication that the external data will become unavailable in the (near) future, *and* the licensing of the external data allows archiving in Areda.

7. Research data associated with publications is made public, unless exceptions apply (see UT policy). Apart from archiving in Areda, data associated with a publication must also be archived in an appropriate online public repository (e.g. 4TU.ResearchData, Areda is not public). Such repositories may have additional requirements (e.g. specific file formats) that should be taken into account from the beginning of the research.
8. Archived data may never be altered. It may only be appended to. Access management should be set such that files cannot be modified once placed in archive. If modifications to archived data are required, a new version should be created instead.
9. Data must be deleted after expiration of its storage term (e.g. 10 years), and possible other conditions. Deleting data from archives requires written permission from either 1) all members having access to that data, or 2) an authoritative person, such as an appointed administrator or project PI.
10. It is forbidden to hold (personal) copies of (deleted) archived data, unless copies are used for active research.
11. Physical material that needs to be archived has material-dependent regulations. No department-level regulations may exist yet for certain materials.
    1. Informed consent forms must be delivered to the BE secretary accompanied by an archiving form, to store the forms in the UT register. Informed consent forms may not be digitized.

9. Data registration

In addition to archiving, all digital and/or non-digital research data and related materials must be registered and described by metadata, including a link or reference to the location of the digital or non-digital objects. Because research data is becoming a valuable asset and in the near future will be formally recognized as scientific output, it is important to know what digital and/or non-digital research data and related materials have been created or used and where these are located.

Data registration – BE working rules

1. No additional registration rules apply. Please refer to the ET policy.

Appendix A – Faculty-specific roles and responsibilities
The roles and responsibilities for different stakeholders within BE are described below. They are an addendum to Appendix A in the ET policy.

The department staff

1. Is responsible for this policy and its implementation.
2. Ensures that the BE policy is reviewed annually.
3. Arranges RDM support and expertise in the faculty: ICT account manager, RDM contact person.

Appendix B – Rationale topic-based folder structure within the group

Section 6 proposes a folder structure that is topic-based, as opposed to a folder structure that is project-based or publication-based. There are several reasons for why this topic-based structure is preferred:

1. Material exists that cannot be attributed to any project or publication (e.g. lecture material, unpublished data), but still needs to be stored, archived, and/or re-used.
2. It promotes **DRY** storage. For example, if multiple projects require the same code to read data from a measurement device, it is desirable to store this version-controlled code in a central place, instead of shattering code duplicates over multiple project folders.
3. It still allows gathering material for a specific project or publication. Multiple publications may exist in the dissemination folder, that all *refer to* the same underlying data in the experimental folder, *without replicating* the actual data in multiple publication folders. If a journal requires a person to upload publication-specific data to a public external repository, it is still possible to gather this data from our storage through the references for upload to the external storage.
4. It promotes re-use of material by making it easier to find. Following section 5, data is “public unless...”. There is no ownership to public data: everyone should be able to take it and re-use it, including everyone outside a project. When searching for material, someone should not have to reason about the project in which any material was produced.

Appendix C – About Microsoft Storage

Please refer to the BE confluence page for more information on Microsoft cloud storage, and how various Microsoft applications interact with that storage.

Appendix D: Suggested future updates

1. Forms for signing data removal (e.g. from camera sd) need to be created
2. Guidelines for anonymizing video data.
3. Overview of all active DMPs, which must be reviewed annually, for example to ensure that data gets (manually) deleted after their expiration period.
    1. *Via LISA, perhaps possible to obtain a list of all DMPs in the group*
    2. *DMP > to pdf. How to manage DMP, and register which data relates to it. DMP must be updated regularly.*
    3. *Alternatively, on archiving, it must be written down in a sheet when the data must be deleted (!).*
4. Procedures on data recovery after unintended loss (though mostly covered by cloud storage)
5. Rules with regard to handling of IP generated by bachelor / master students. These are most likely on a case-by-case basis.
6. Naming conventions for Areda zip data