

DATA MANAGEMENT PLAN [in English]

1. Data description and collection or re-use of existing data

How will new data be collected or produced and/or how will existing data be re-used?

The displacements and forces obtained from tester machine will be recorded in .txt file format. The course of the experiments will be also recorded in the form of videos, photos. During the experiments, the test procedure proposed by the manufacturer of the tester machine will be used. Load cell calibration will be performed before each test session to ensure the quality and consistency of experimental data. The results from the experimental tests will be used in validation of the numerical models.

MicroCT scans of 3D prints and tissue samples will be processed using 3D Slicer and HyperMesh software to obtain meniscus geometry.

Data from numerical simulations will be obtained in Finite Element Method (FEM) programs - Abaqus, FEBio. The results of computations will be processed using WORD, EXCEL and Grapher programs.

The geometry of the knee joints from <https://simtk.org/projects/openknee> (open, freely available project) and the principal investigator's previous models will be used.

What data (for example the types, formats, and volumes) will be collected or produced?

Raw data from experimental tests will be stored in .txt file format. Processed data from experimental tests will be stored in the form of excel sheets (xlsx), with the maximum size of 200 MB. The course of the experiments will be stored in digital video formats (avi, mp4), photos (bmp, jpg, png) and the experiment parameters in text format (pdf, docx, txt). The total size of the experimental data is estimated at 200 GB.

MicroCT scans of 3D prints and tissue samples will be saved in standard .dcm file format of MicroCT scanner. The geometry developed in 3D Slicer and Hypermesh from microCT scans will be saved in .stl file format. The total size of the scans and geometry is estimated at 300 GB.

The data related to the numerical models will be kept in formats characteristic to Abaqus software (cae, inp, odb) and FEBio (feb, log, xplt). The size of the results from a single simulation should not exceed 20 GB. The size of the simulation data is therefore estimated to be up to 1200 GB.

2. Documentation and data quality

What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?

Details of data organization will be described in the documentation prepared by the research team. The naming convention for directories and files will be standarized. The rules of data cleaning [1] will be used to obtain tidy data with appropriate structure of metadata.

Metadata descriptions will be stored in JSON-LD format. In the case of experimental datasets, the method of conducting experimental tests, parameters and the column content in the .txt file with raw data will be included in the README.txt file attached to the dataset. In the case of datasets derived from numerical simulations, all information necessary to repeat the computations, date of calculation and a description of the results will also be included in the README.txt file attached to the dataset.

The authors will be identified and authorized by ORCID number.

[1] Wickham, H. . (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>

What data quality control measures will be used?

Load cell calibration will be performed before each test session to ensure the data quality from experimental tests. The parameters of experimental tests will be controlled by two members of team. The results from tests will be compared with results from literature and the course of the experiment will be checked on video to eliminate measurement errors. Manufacturer's recommendations and test procedure instructions will be used during testing. All data will be produced by trained and qualified team members.

The numerical models will be checked by two members of the research team before calculations to eliminate potential errors. The results of simulations will be validated with data from experiments and literature.

A version control system will be used to assure continuity and quality of the gathered data.

The principal investigator will be responsible for overall data quality control.

Access to research data will be limited to the project team to avoid unauthorized modification of data.

3. Storage and backup during the research process

How will data and metadata be stored and backed up during the research process?

Data from experimental tests and the accompanying metadata will be kept on the computer of tester machine and on the department's data cloud (<https://cloud.wilis.pg.edu.pl/>). Data from CT scans, models of geometry and FEM models will be kept on personal computers and on the department's data cloud. Results of numerical simulations will be stored in the Centre of Informatics Tricity Academic Supercomputer and network (CI TASK), where backups are conducted on monthly basis, and on the department's data cloud. Project documentation, manuscripts of the papers, and partial project reports will be stored on the Microsoft OneDrive cloud service (Office 365 A1 for faculty license) and on personal computers.

Additionally, two backup copies of all data and the accompanying metadata will be created at least every month on two additional password-protected external discs which will be purchased as part of the project. The Principal Investigator will be responsible for backup copies and discs.

How will data security and protection of sensitive data be taken care of during the research?

Since no sensitive data will be processed in the project, no special protections measures are needed. However, the safety of other data will be ensured by regularly changing the computer password at least once every six months and the security procedures of backup services providers. Access to computers at the university from outside the university' network is only possible using an authorized VPN connection, assigned by the University's IT Department.

4. Legal requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

The personal data of the patients from whom the human menisci will be harvested during total knee arthroplasty will not be processed.

Samples taken from the menisci by an orthopedist will be sent for testing with an anonymous identifier. The written consent for participation in a research study describes in detail the necessary procedures to ensure the protection of the personal data.

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

The ownership and management of any intellectual property developed in collaboration relating to the Project remains in the rights of the Gdańsk University of Technology accordingly to the Polish law and institutional regulations (Resolution of the Senate of the Gdańsk University of Technology No. 117/2021/XXV of 19 May 2021 https://link.pg.edu.pl/GdańskTech_intprop).

The data and results will be published in open-access model accordingly to Plan S requirements under the one of the Creative Commons licenses. The data shared in open research data repository will have CC-BY or CC-0 license assigned. Metadata descriptions for these datasets will be always available without any restrictions (CC-0).

5. Data sharing and long-term preservation

How and when will data be shared ? Are there possible restrictions to data sharing or embargo reasons?

The part of the data will be shared through the open research data repository – MOST Wiedzy Open Research Data Catalogue (<https://mostwiedzy.pl/en/open-research-data/catalog>) and the open-access repositories of scientific journals during the grant period.

No embargo or other restrictions will be applied once the research results are published.

If the publisher requests it, raw data will be shared on the day of publication.

How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

All data used in published papers and necessary to support conclusions formulated during the project will be preserved long-term. Additionally, data that could be reused in the future will also be selected for preservation, ensuring its completeness.

These data will be prepared according to FAIR (Findable, Accessible, Interoperable, Re-usable) principles and stored for at least 10 years in Centre of Informatics Tricity Academic Supercomputer and network (CI TASK) and on additional external discs which will be purchased as part of the project. At CI TASK backups are performed regularly once a month.

Data deposited in the MOST Wiedzy Open Research Data Catalogue repository will be automatically categorized for long-term preservation, without expiration date. The repository is the only service in Poland CoreTrustSeal certified, which means that it has established good preservation and dissemination practices.

What methods or software tools will be needed to access and use the data?

Data stored in the MOST Wiedzy Open Research Data Catalogue repository will be made available in open formats, including: txt, png, stl and csv. The MOST Wiedzy Open Research Data Catalogue repository will offer advanced data search tools.

The experimental data will be stored in txt file format. The geometry of 3D prints and menisci will be stored in stl file format. The input files for the computations will be stored in inp and feb file formats. Access to all this type files (txt, stl, inp, feb) will be possible through a text editor. The results of numerical simulations will be stored in odb format and xplt format, for which Abaqus and Febio are required, respectively. Currently, Abaqus is commercial software and FEBio software is freely available.

It should be possible to open older results files (odb, xplt) with newer versions of the Abaqus and Febio programs, so conversion is not necessary.

How will the application of a unique and persistent identifier (such us a Digital Object Identifier (DOI)) to each data set be ensured?

Datasets deposited in the MOST Wiedzy Open Research Data Catalogue repository will receive unique DOI identifiers.

6. Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data mangement (i.e the data steward)?

The Principal Investigator will be responsible for data management at all stages of project implementation, as well as after its completion.

As part of the task performed, a team member is responsible for quality control, sharing, preparing and archiving data. The University Open Science Competence Center - offers support in creating and versioning Data Management Plans (DMPs), support in ensuring the quality and completeness of dataset metadata in the repository, and substantive support for repository users.

Centre of Informatics Tricity Academic Supercomputer and network (CI TASK) is responsible for maintaining archiving infrastructure.

University data steward provides DMP support.

What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

The data management will be carried out by statutory employees of the university according to FAIR principles. Therefore, ensuring FAIR principles in the project will not involve any additional costs.