An image-based personalized musculoskeletal modeling and simulation framework for functional analysis of activities of daily living in adult spinal deformity patients

A Data Management Plan created using DMPonline.be

Creators: Birgitt Peeters, n.n. n.n., n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n., n.n. n.n.

Data Manager: Birgitt Peeters, Thomas Overbergh https://orcid.org/0000-0002-2144-0713

Grant number / **URL:** 1SE6223N

ID: 197234

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Adult spinal deformity (ASD) is a complex disorder of the spine resulting in three-dimensional deformities, often leading to severe functional impairment during activities of daily life. Spinal fusion surgery is increasingly used to realign the spine within normative values with rods and screws. Surgical planning is predominantly based on two-dimensional static radiographic parameters and subjective surgical expertise. However, the often absent functional improvement and high revision rates highlight the structural lack of objective insights in the highly complex musculoskeletal function of the spine in current clinical decision-making. Computational simulation based on musculoskeletal models (MSKMs) is an innovative tool recently recognized for obtaining such objective insights in ASD. However, as ASD involves important subject-specific alterations in muscle anatomy and stiffness of the intervertebral (IV) joints, MSKMs should first integrate such alterations to generate biofidelic insights in spinal function. Thereto, this project aims to develop a modeling framework that generates MSKMs of the spine that uniquely replicate each patient's aberrant bony anatomy, muscle anatomy, and IV stiffness based on medical imaging. Associated simulations of clinically relevant motor tasks will enable the first objective assessment of functional impairment at the joints of interest for patients with and without functional improvement towards the further improvement of surgical outcome.

Last modified: 27-04-2023

An image-based personalized musculoskeletal modeling and simulation framework for functional analysis of activities of daily living in adult spinal deformity patients
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

The research makes use of a retrospective dataset containing the following:

1. Demographic data (information such as age, gender, and medical comorbidity) (KB)
2. Clinical and surgical exam results (exam outcomes performed by the surgeons identifying kinematics and balance in dynamic activities, surgery details) (KB)
3. Motion capture data: video recordings and 3D recorded motion trajectories of reflective markers (.mp4, .C3D format) (MB)
4. Medical imaging data: MRI, EOS, CT, X-ray images, 2D and 3D static parameters (.DCM, .MCS, .MXP format) (GB)
5. Surface electromyography (EMG measurements of erector spinae muscle bilaterally during activities of daily living) (MB)

The research will generate the following data:

1. Estimates of kinetics, kinematics, ground reaction forces, muscle forces, and loading in patients during a set of motor tasks (from 3. above) (.mot, .sto, .XML data formats) (MB)
2. Image-processing results (muscle volumes, CSA, centerlines, attachment points, centers of masses, wrapping surfaces, fat infiltration, and moment arms) (.stl) (MB)
3. Computational models and algorithms (musculoskeletal modeling and simulation data) (.osim, .mat, .py) (GB)

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of the responsible person (If already designated, please fill in his/her name.)
Prof. Dr. Ir. Lennart Scheys will be the person responsible to preserve data for 5 years after the end of the project.
2. Storage capacity/repository

- during the research
- after the research

During and after the research, data will be kept in dual backups (external hard drives and KU Leuven server). Moreover, data and metadata is also stored on KU Leuven's data warehouse through DOPLr (Data and Organization Platform for Research), a secured, in-house developed data management platform compatible with *Integrated Rule-Oriented Data System* (*iRODS*). This platform only grants access to allowed persons and to specific studies. The external hard drives will be in the possession of Prof. Scheys after the end of the project. Access to the data on the KU Leuven server is only granted to members of the Institute for Orthopaedic Research and Training (IORT), KU Leuven.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

We do not wish to deviate from the principle of preservation of data for a minimum preservation term of 5 years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

This project uses a retrospective dataset, for which ethical approval has already been obtained (S58082). All included patients gave informed consent to the study. As no new personal data will be created (i.e. pseudo-anonymized computer models and kinematic outputs from motion simulations) and the research goals fully overlap with the available ethical approval, no extra measures had to be undertaken for the ethics committee. The retrospective dataset and all generated data are only accessible to specific persons working on the study on secured platforms (DOPLr, KU Leuven server, external hard drives). If data is to be shared with external parties, such as students, the data is anonymized and an NDA has to be signed.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

No other issues related to data management are anticipated or relevant to mention.

An image-based personalized musculoskeletal modeling and simulation framework for functional analysis of activities of daily living in adult spinal deformity patients
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

An image-based personalized musculoskeletal modeling and simulation framework for functional analysis of activities of daily living in adult spinal deformity patients
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

An image-based personalized musculoskeletal modeling and simulation framework for functional analysis of activities of daily living in adult spinal deformity patients
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Medical Imaging Data	Medical images obtained from CT, MRI, biplanar radiography (EOS), traction radiography	Reuse existing data	Digital	Experimental	.dcm .mcs	<1 TB	
Motion Capture and Force Data	Optical recordings of activities performed by patients, through VICON software. Ground reaction forces (GRF) and electromyography (EMG) data measured during the activities	Reuse existing data	Digital	Experimental Compiled/aggregated data	.avi .c3d .mot .sto .x1d .x2d .xcp	<100 GB	
Biometric data	Gender, age, length, weight, BMI	Reuse existing data	Digital	Observational Experimental	.txt	<100 MB	
Clinical Examination Data	Outcomes of the evaluation of neurological disorders and compensation mechanisms in daily life activities by surgeons	Reuse existing data	Digital	Observational Experimental	.txt	<100 MB	
Surgical Treatment Data	Medical records on the performed treatment: fusion level, perioperative complications	Reuse existing data	Digital	Experimental	.txt	<100 MB	
Treatment Outcome Data	Health-related quality of life questionnaires	Reuse existing data	Digital	Experimental	.txt	<100 MB	
Modeling Data	Computed biomechanical parameters and geometries: skeletal and muscle meshes, muscle and intervertebral stiffness parameters and estimates, and spinal alignment measurements.	Reuse existing data Generate new data	Digital	Experimental	.XML .csv .txt .obj/stl .mcs/mxp .osim	<100 GB	
Simulation Data	Data describing spinal kinetics and kinematics, muscle forces, passive stiffness forces, and loading in musculoskeletal models during specific daily life activities	Generate new data	Digital	Simulation data	.XML .mot .sto .osim	<100 GB	
Research Documentation Data	Documentation generated by the research or collected from online sources and from collaborators, including ethical approval documents, laboratory notes, protocols	Generate new data	Digital	Compiled/aggregated data	.pdf .txt	<100 MB	
Manuscript Data	Data resulting from the research will be published as articles in peer-reviewed journals	Generate new data	Digital	Compiled/aggregated data	.pdf	<1 GB	
Algorithmic and Programming Data	algorithms and scripts created to develop the modeling and simulation frameworks. More specifically for medical image segmentation, image processing, parameter calculation, optimization of model parameters, simulations of motion	Generate new data	Digital	Compiled/aggregated data	.txt .py .mat .cpp	<1 GB	
Software Data	Set of used visualization tools, computer programs	Reuse existing data	Digital	Software	.exe	<1 TB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

The retrospective dataset (including all data types: medical images, biometrics, clinical exams, treatment details, models, and motion capture data) is accessible on DOPLr, with restricted access to the collaborators of the study.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data

Ethical approval is obtained for the study: S58082

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Personal data includes biometric data, names, data held by the hospital (EAD number), motion capture recordings, and medical images.

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

Hospital data from the patients: no personal patient data is allowed to be transferred to any external parties without prior approval from applicable UZ Leuven committees.
However, no third-party agreement restricts dissemination or exploitation of the methodological and technical data resulting from this project (i.e. the developed algorithms).

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

All collected data (i.e. biometrics, clinical and treatment information, medical images, motion capture, models, and resulting outcomes from image processing, modeling, and simulations) are and will be stored in an online research management system (DOPLr),
which runs on a secured and backed up server of KULeuven and enforces FAIR data principles. This system ensures that everything will be traceable and stored long-term (well beyond the 5-year requirement). Moreover, algorithms, scripts, and software usage will be documented, and when finalized they will be additionally described in manuscripts and on GitHub in the IORTLeuven team, which is private and accessible to our members. Finally, the documentation which is required to understand the methodological developments will also be stored on the IORT Leuven platform from KU Leuven Box software, which is secured and backup on the server of KU Leuven.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

From the onset of the project, data documentation will be tailored to their deposition in the DOPLr repositories, with headers corresponding to the requirements of the DOPLr system and compatible with the open-source *Integrated Rule-Oriented Data System* (*iRODS*) system. Technical and analytical methods used to generate the data will be documented in sufficient detail to allow for independent reproduction. These will include software version numbers, segmentation algorithm requirements, processing requirements, computational requirements, etc. When depositing data in a repository, the final dataset will be accompanied by this information in the form of a README.txt document. This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the research group and add context to the dataset for future reuse.

3. Data storage & back-up during the research project

Where will the data be stored?

The KU Leuven Enterprise Box server with database hosted on a secured KUL server with password-protected
access will be used for storage of results from the study, such as excel files, reports, manuscripts, presentations, etc.
All digital data will be dually stored on the KU Leuven server (secured), and on the DOPLr platform (secured), which is specifically developed to store large amounts of data for long periods of time.
Algorithms, scripts, and software will be stored in the private online git of the IORT.

How will the data be backed up?

All data is automaticallybacked-up on Box and the KU Leuven server. Additionally, extra hard drives are used to store data as a third safety mechanism.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

As all data is stored on DOPLr and KU Leuven servers, and DOPLr runs on these servers, we can confirm that there is and will be sufficient storage and backup capacity during the project:

- the “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp e-series storage systems, and a CTDB samba cluster in the front-end
- the “J-drive” is based on a cluster of NetApp FAS8040 controllers with an Ontap 9.1P9 operating system

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The Box software is only accessible to Box accounts, and specifically, our data will only be accessible to group members of the IORT.
We will also use secured storage on DOPLr. DOPLr is password-access protected by users, which in turn must first apply to become an official user of the software. Furthermore, it offers version management of all data. Our study is in turn only accessible to collaborators in the research (while other researchers can store their data in another separate secured environment, only accessible by them).

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The total estimated cost of data storage during the project is ~3,000 EUR. This
estimation is based on the following costs:
The costs of digital data storage on the "L-Drive" are 173,78 EUR/TB/Year and 519 EUR/TB/Year for the “J-drive”.
The costs of digital data storage on KU Leuven Entreprise Box are 25 EUR/Year/Person.
These costs will be covered by the department.
Data storage on DOPLr is free.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

The minimum preservation term of five years after the end of the project will be applied
to all datasets.

Where will these data be archived (stored and curated for the long-term)?

All digital data will be archived on the secured university’s network drive and DOPLr. Developed algorithms and software will be stored on the secured university's network drive and GitHub. Additionally, data will be stored offline on external hard drives after the completion of the project.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The costs for data preservation on the KU Leuven servers and DOPLr is estimated to be 868,90 EUR (based on 5 years for 173,78 EUR/TB/Year for the "L-drive" and free access to DOPLr). These costs will be covered by the department.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in a restricted access repository (after approval, institutional access only, …)

Personal patient data linked to a supporting manuscript will only be published after anonymization as supplementary information.
Computer model datasets (musculoskeletal models) will be deposited in the open-access repositories of SimTK, after de-identification. SimTK is a free project-hosting platform for the biomedical computation community.
Other digital datasets that support publications (including image, model, and simulation data) will be made publicly available via an open research data platform such as Mendeley Data.
Research documentation: All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents (daily logs, raw data) deposited in the KU Leuven servers and DOPLr are accessible to the PI and the research staff and will be made available upon request.
Manuscripts: All scientific publications will be shared openly. At the time of publication, research results will be summarized on the IORT's website (IORT) and post-print pdf versions of publications will be made available there if allowed by copyright agreements, possibly after an embargo as determined by the publisher. Before the end of the embargo or in cases where sharing the post-print is not allowed due to copyright agreements, a pre-print version of the manuscript will be made available. Publications will also be automatically added to our institutional repository, Lirias 2.0, based on the author's name and ORCID ID (the metadata will be added, not the full manuscripts).
Algorithms, scripts, and software: All the relevant algorithms, scripts, and software code driving the project will be made available to restricted repositories such as IORT/Git, and can be available upon request.
Data that do not support publication will be either deposited in an open-access repository or made available upon request by email.

If access is restricted, please specify who will be able to access the data and under what conditions.

A-priori approval has to be requested and provided by a KU Leuven responsible on the data and model repositories.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Ethical aspects

All personal patient data cannot be shared by 3rd parties unless an agreement with UZ Leuven is formally generated, and all patient data is anonymized.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Manuscripts: All scientific publications will be shared openly. Personal patient data linked to a supporting manuscript will only be published after anonymization as supplementary information.
Other digital datasets that support publications (including image, model, and simulation data) will be made publicly available via an open research data platform such as Mendeley Data.
Computer model datasets (musculoskeletal models) will be deposited in the open-access repositories of SimTK, after de-identification. SimTK is a free project-hosting platform for the biomedical computation community.
Algorithms, scripts, and software: All the relevant algorithms, scripts, and software code driving the project will be made available to restricted repositories such as IORT/Git, and can be available upon request.

When will the data be made available?

As a general rule, all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, and potential IP requirements.

Which data usage licenses are you going to provide? If none, please explain why.

Data: Creative Commons Attribution-NonCommercial-ShareAlike (CC-BY-NC-SA) license

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories whenever possible.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Data documentation and metadata will be managed by the PhD student and technical staff associated to this project.

Who will manage data storage and backup during the research project?

Data management, storage and back up will be performed by the PhD researcher associated with this project, under supervision of the PI.

Who will manage data preservation and sharing?

The PI is responsible for data preservation and sharing, with support from ICTS, gbiomed-IT staff, and UZ-IT staff.

Who will update and implement this DMP?

The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.