Quantitative automated evaluation and multi-modal monitoring tool for movement disorders in dyskinetic cerebral palsy

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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Contact informaton	Subject identification list including names, phone numbers, e-mail (collected for contact purposes)	New	Digital	other	.xlsx	<100MB	
ICF	Informed consent forms templates en signed	New	Physical or digital	other	.docx .pdf	<100MB	Printed papers (one ring binder located 4.100 campus Bruges

Descriptive documents	Protocol, Study operations manual (for measurements, instructional material for assessors, protocol pre- and postprocessing steps, instructions for scoreres (KU Leuven formats/templates used if available)	New and existing	digital	other	.docx .pdf	<100MB	
Patient characteristics (data collection form)	Diagnosis, medication related to movement disorder, age (year/month), weight, body length, gender, Gross Motor Function Classification System (GMFCS) and Manual Ability Classification (MACS)	Existing	digital	Observational	.csv	<100MB	
Retrospective video data	Common 2D videos collected in previous trials: Dyskinesisa impairments scale tasks	Existing	digital	Observational	.mp4	<5TB	
Retrospective MMA data	X y coordinates of bodypoints extracted from retrospective videodata	New	digital	Compiled/aggregated	.csv	<1GB	
MMA features retro	MMA features calculated for 5 seconds time windows	New	digital	Compiled/aggregated	.csv	<1GB	
Video data new	Common 2D videos collected by standardize home-based protocl	New	digital	Observational	.mp4	<5TB	
MMA data new	X y coordinates of bodypoints extracted from new videodata	New	digital	Compiled/aggregated	.csv	<1GB	
MMA features new	MMA features calculated for 5 seconds time windows	New	digital	Compiled/aggregated	.csv	<1GB	
Ruw IMU	Inertial measurement unit data of head, wrists and ankles	New	digital	Observational	.csv	<1GB	
IMU features	IMU features calculated for 5 seconds time windows	New	digital	Compiled/aggregated	.csv	<1GB	

Ruw HR data	Heart rate data	New	digital	Observational	.csv	<1GB	
HR features	Heart rate features calculated for 5 seconds time windows	New	digital	Compiled/aggregated	.csv	<1GB	
Clinical scoring 5 sec timewindows	Clincial scores obtained by scoring videos within 5 seconds timewindows for dystonia and choreoathetosis severity by three raters	New	digital	Compiled/aggregated	.csv	<100MB	
Videos clinical scales	Videos of the Dystonia Impairment Scale (DIS)	New	digital	Observational	.mp4	<1TB	
Clinical scoring	DIS, BADS	New	digital	Compiled/aggregated	.csv	<100MB	
Publication data	Full dataset of extracted dataset, IMUs and HR data, combined with clinical scores	New	digital	Compiled/aggregated	.csv	<100 GB	
Software codes	Software codes for data cleaning, data validation, feature calculation and machine learning	New and existing	digital	Compiled/aggregated	.py .mat	<1 GB	
Metadata	Metadata on processed MMA, IMU and HR data and clinical scoring	New	digital	Compiled/aggregated	.csv	<100MB	

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:

If available we will reuse certain data from the participants medical records: Diagnosis, medication related to movement disorder, age (year/month), weight, body length, gender, Gross Motor Function Classification System (GMFCS) and Manual Ability Classification (MACS)

Existing protocols and software codes:

DIS protocol: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8749.2011.04209.x

BADS protcol: https://doi.org/10.1017/s0012162299000870

Software codes: https://doi.org/10.48804/8LMRXT ; https://github.com/RehabAUmc ; https://github.com/SamDeWinter/Thesis ; https://github.com/ThomasVE2000/Thesis-movement-assessment-of-the-upper-extremity-in-dyskinetic-cerebral-palsy

Videodata KU Leuven:

https://drives.kuleuven.be/ Large Volume Storage/CRD-REV001_Neuro-Monbaliu/AVI-DYS/DATA-AVI-DYS retro

Videodata Amsterdam UMC:

L:\Archief\active\rev-onderzoek-01\ML4DYS-Veni\2. DATAVERZAMELING\Videodata_retro

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

Within the research children and young adults with dyskinetic cerebral palsy and typically developing peers will be included (methodological observational stuyd). Ethical approval will be obtained prior to the start of measruements. Reseach will be conducted only on the basis of prior informed consent by the participants and/or their legal representatives (dependend on the age).

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

Three types of personal data will be gathered:
(i) Personal information for contact purposes (directly identifiable data included in the subject identification list e.g. name, phone number, email), which will not be used in any further analysis. Participants will be asked whether this information can be stored in a database for future research, via a separate informed consent procedure in accordance with the General Data Protection Regulation.
(ii) Personal and medical information for research purposes (pseudonymous data), consisting of socio-demographical data (Diagnosis, medication related to movement disorder, age (year/month), weight, body length, gender, Gross Motor Function Classification System (GMFCS) and Manual Ability Classification (MACS)), via the study-related informed consent procedure in agreement with the General Data Protection Regulation.
(iii) Personal information for research purposes (non-pseudonymous data), concerning video recordings used for markerless motion tracking and clinical scoring
Prior data collection before GDPR/Pret will beregistered.

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

Data and software will be made as much as possible in accordance with privacy regulation made publicly available.

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.

- No

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).

Study operations manual for measurements and instructional material for assessors will be established. Users will be trained in

the data collection.
An electronic CRF will be used to collect all patientcharacteristics (REDCap). Videos, IMUs and HR data will be collected using an app allowing timesynchronized collection.
Steps of pre- and postprocessing of data and feature calculation will be well documented within the software code, making every step from raw data to features reproducable. Version control within programming will be used. The final code will be made available on KU Leuven RDR together with a README file
The syntax of the statistical analysis (SPSS) will be stored.

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.

- No

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

Where will the data be stored?

Patientcharacteristics such as Diagnosis, medication related to movement disorder, age (year/month), weight, body length, gender, Gross Motor Function Classification System (GMFCS) and Manual Ability Classification (MACS) will be collected and stored in the eCRF (REDCap). The results of scoring of the videos of three raters will also be uploaded to the eCRF.
Measured raw data will be stored at Large Volume Storage (L-drive) of KU Leuven, specially developed to store large amount of data for longer periods of time and suitable for sensitive data.
Amsterdam UMC data will be stored locally. Only pseudomized data will be merged for machine learning. Data transfer of pseudomized data will be transfered in a secure way (i.e. by surfsend.surf.nl or filensender belnet)
Study documentation will be stored on a shared reaseach team folder on KU Leuven OneDrive.For an effective data managment ManGO of KU Leuven will be used.
The paper copies of the informed consent forms will be stored in a secured locker at the Department of Rehabilitation Science, Campus Bruges of KU Leuven and at the Department of Rehabilitation Medicine at Amsterdam UMC, respectively. Only authorized personnel will have access to this locked storage room.

How will the data be backed up?

Standard back-up will be provided via the KU Leuven secure network drive with automatic daily backup procedures. Addidionally, a mirror of the data is provided in a second ICTS data center for buisness continuity or disaster recovery purposes.
At Amsterdam UMC similar backup procedure are in place.

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

Sufficient storage and backup capacitiy for data - as described in part 1 of this DMP - is provided on the KU Leuven and Amsterdam UMC servers and networks.

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

The digital pseudonymous data (i.e. coded and containing no personal information) and the non-pseudonymous data (i.e. video data) will be stored in secure university envrionment. The PI of this project will be the only one who can grant access to this network drive. Access to the study folder on the on L:/drive is secured by a two-factor authenticated password protection.
Data within the eCRF is only accessable for unauthorized person with logging of changes.

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

Data will be around 5-10 TB. The expected cosst are 600-1200 Euro/year and will be covered by the project.

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 standard advice of the Research Ethics Committee of UZ/KU Leuven for data storage is to keep data on a highly secured network drive (i.e. K-drive) for a minimum of 25 years after completion of the study. This is with a view to reproduction, verification, and potential reuse of research data. Video data will only be kept if we have additional consent of the participant/guardian.

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

Data will be archived on the secured university's network drive (K-drive) and RDR.

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

For the final year of the project an additional part in the budget is foreseen for data archiving (K-drive) for a period of 5 years. The expected costs are €12,84/year/100Gb. Costs related to archiving will be covered for 50% by the Department of Biomedical Sciences and for 50% by the project.

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

Yes, the data will be made as much as possible available within RDR KU Leuven. The full pseudonymous dataset and the metadata will be made available after publication of the data and upon request with the PI. Importantly, only pseudonymous data of participants who granted their approval for re-use, either within the research group (closed data) or outside the research group (open data), will be made available (also see 'Who will be able to access the data and under what conditions?'). Additionally, after publication, the data softwarecodes will be made available with a detailed README.txt file.

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

If parents/participants have consented, upon request, raw video data (non-pseudonymous data) can be shared for educational purposes at KU Leuven or within the European Network of Dyskinetic Cerebral Palsy for reserach with the aim to improve assessment or treatment of children with cerebral palsy.

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, Privacy aspects

Raw video data cannot be shared in public due to privacy restrictions.

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

KU Leuven RDR with different restrictions.

When will the data be made available?

Published data will be made available at the time of publication of the appertaining reserach paper.

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

Software code: GNU GENERAL PUBLIC LICENSE
Data: CC-BY 4.0 data 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?

We plan to publish data at RDR KU Leuven. These costs are covered by KU Leuven.

6. Responsibilities

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

The coordinating reseracher: Helga Haberfehlner

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

Helga Haberfehlner; backup automatically

Who will manage data preservation and sharing?

PIs: Elegast Monbaliu, Jean-Marie Aerts

Who will update and implement this DMP?

Helga Haberfehlner