
Identifying Dyslexia Earlier in Life

A Data Management Plan created using DMPonline.be

Creators: Wendy Verheijen, Marlies Gillis

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Project Administrator: Wendy Verheijen

Grant number / URL: PDMT1/23/011

ID: 202885

Start date: 01-11-2023

End date: 30-09-2024

Project abstract:

Diagnosing dyslexia relies on a waiting-to-fail approach, i.e., the diagnosis relies on persistent struggles with reading and writing after some years of reading instruction. By earlier identifying children with a high risk for developing dyslexia, one could start earlier and more effective intervention so that the impact of dyslexia on an individual's life is reduced. Although dyslexia manifests as a reading and writing disorder, it is thought to be rooted in reduced auditory temporal processing. Therefore, investigating the neural responses to sounds might be key to identifying dyslexia earlier in life. We aim to distinguish children with dyslexia from neurotypical children using a wide set of neural metrics derived from two novel approaches, i.e., (1) neural tracking and (2) functional connectivity measures applied to M/EEG data of children listening to natural speech. These approaches allow us to investigate the temporal characteristics of the neural responses and how these are related to the neural connectivity networks across the hierarchical levels of neural speech processing ranging from purely acoustic processing to linguistic speech processing. Using the combination of these approaches on different unique datasets allows us to identify dyslexia objectively using neural metrics.

Last modified: 16-01-2024

Identifying Dyslexia Earlier in Life

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
CTSiN The role of reading experience in atypical cortical tracking of speech and speech-in-noise in dyslexia	Open source available dataset which contains MEG data of children with and without dyslexia listening to speech (presented with or without lip movements / with or without noise)	<i>E</i> (existing)	D (digital)	Audiovisual Sound Numerical Textual	.fif	<300GB	

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 dataset (CTSiN) is available through this link: <https://osf.io/9ce5t/>

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

The study which collected the dataset was approved by the local ethics committee (Comité d’Ethique Hospitalo-Facultaire Erasme-ULB, 021/406, Brussels, Belgium; approval number: P2017/081). Since this project only will reuse the existing data from the public available dataset, no ethical approval by EC is required (confirmed by EC Leuven on 20/12/2023).

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

In this project no additional data are collected. However, we still filled in a PRET form: PRET G-number: G-2023-7308

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 or 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

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).

Currently, the data is stored in a custom format.

Before starting the first analysis, we will convert the dataset to BIDS structure (<https://bids.neuroimaging.io/>). The BIDS data structure is an internationally recognized dataset format to improve replicability.

The generated script will be stored on GitLab.

Will a metadata standard be used to make it easier to find and reuse the data?

If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

The BIDS data structure requires some additional metadata files to increase readability and reproducibility of the dataset. As far as they are available, these metadata will be added.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)

The raw data is stored online (<https://osf.io/9ce5t/>). In order to analyze the data, some files will need to be stored on the (encrypted) hard drive (since calculations from a non-local source are too slow and lead to computational failures).

A copy of the data will be stored on KU Leuven managed drives.

The generated script will be stored on Gitlab. Important intermediate results will be saved on OneDrive.

How will the data be backed up?

- Personal back-ups I make (specify below)

A backup on the encrypted local hard drive is taken automatically twice a week.

GitLab manages the code and its versions, and therefore, also functions as a back up of all code to generate intermediate results.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

During analysis, the data is stored on the researcher's hard drive, which is encrypted.

At the beginning of the project, the dataset will be copied onto the local hard drive and the KU Leuven managed drive to prevent changes in the online dataset from affecting the planned project.

The scripts are backed up through GitLab, which is accessible to all team members and PI, but, not for unauthorized persons.

However, since the raw data are publicly available, no additional measures are necessary.

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

No additional costs are expected.

OneDrive is maintained centrally by KU Leuven. The personal hard drive has already been purchased. KU Leuven researchers can make use of the Gitlab services free of charge.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

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

- Other (specify below)

The data is stored online (<https://osf.io/9ce5t/>).

A copy of the data will be stored on KU Leuven maintained drives.

After finalizing the study, the caches and results will be stored on KU Leuven-managed drives (for min. 10 years). The processing code will be stored in Gitlab (and a Gitlab tag will be generated).

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

Expected costs are: €11,384/year/100GB * 3 * 10years = €341.52. These costs will be covered by a reserve credit (as the annual costs are rather low).

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Other (specify below)

This dataset is already open-accessible data. This dataset can be found here: <https://osf.io/9ce5t/>

Regarding the scripts and results generated during the current project, it is not clear whether these will be shared. This depends on whether the results will be published. It is unclear whether a publication of the current project will be realized in the timeframe of the PDM.

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

NA

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 per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other (specify below)

The data is readily available on: <https://osf.io/9ce5t/>

Regarding the scripts and results generated during the current project, it is not clear whether these will be shared. This depends on whether the results will be published. It is unclear whether a publication of the current project will be realized in the timeframe of the PDM.

When will the data be made available?

- Other (specify below)

The data is readily available on: <https://osf.io/9ce5t/>

Regarding the scripts and results generated during the current project, it is not clear whether these will be shared. This depends on whether the results will be published. It is unclear whether a publication of the current project will be realized in the timeframe of the PDM.

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

Regarding the raw data, available in the public database, the license is not specified on <https://osf.io/9ce5t/>

Regarding the scripts and results generated during the current project, it is not clear whether these will be shared. This depends on whether the results will be published. It is unclear whether a publication of the current project will be realized in the timeframe of the PDM.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

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

NA. The data is readily available on: <https://osf.io/9ce5t/>

Responsibilities

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

I (Marlies Gillis) will convert the data set into BIDS structure and create the metadata to increase the readability of the dataset. The end-responsibility lies with professor Francart (PI).

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

I (Marlies Gillis) perform an automatic backup at least twice a week of the encrypted hard drive and on OneDrive. Similarly, I will store the code on Gitlab.

The data is also stored online at <https://osf.io/9ce5t/>

Who will manage data preservation and sharing?

Regarding the raw data available in the public database, ULB professor Mathieu Bourguignon is responsible for data preservation and sharing. If the data of the current project is shared, I (Marlies Gillis) will manage the data sharing. The end-responsibility lies with professor Francart (PI).

Who will update and implement this DMP?

I (Marlies Gillis) will update and implement this DMP.