Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

Data Management Plan Template

Via this template you can create (or upload) your data management plan (DMP). We know that things change, so this DMP template should be edited throughout your research project.

Questions: If you have questions for the data steward of your department, add them at the end of the form as a comment under “Activity”.

Save: click "save" at the bottom of the form to save the information you provided, you can modify your answers in the draft status.

This template has been approved by NWO, ZonMw, and the European Research Council.

General Project Information

1. Project Title / Study name *

Measurement of Morphological Richness in Tagalog Agrammatic Speech

2. Are you a? *

- ⊙ Researcher (PhD, academic staff)

3. Primary Contact Name (Student or Researcher) *

Jonathan Gerona

4. Primary Contact Email (Student or Researcher) *

j.gerona@tue.nl

When applicable, at least one supervisor should be mentioned.

5. Supervisor(s) Name(s)

Liz Guzman-Ramirez

6. Supervisor(s) email address(es)

l.guzman.ramirez@tue.nl

7. Department *

Other

7a. Please write your affiliation *

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

Neurolinguistics Research Group

Need to share your DMP? Add the email of the person on the right side panel under the "Shared with" label (on low-resolution computers, this option appears at the end of the form).

8. What is the purpose of this study? (select all that apply) *

☑ Scientific purposes (possibly leading to a publication)

9. Start date of data collection *	10. End date of data collection *	11. Estimated end date of the project *
11/1/25	2/28/26	9/30/26

12. Does your project receive external funding (e.g., NWO, relevant for special regulations from funders)? *

⊙ Yes

Some questions have been added to the form to fulfill your funder requirements.

13. Is your funder? *

☑ European Commission (EC)

13b. Please provide your grant number(s). *

123456789

14. What are your main research questions? *

The overall aim is to characterise the dimensions of morphological deficit in the verb usage among Tagalog-speakers with agrammatism. To achieve this, we take advantage of the unique features of Tagalog verbal conjugation and build upon linguistically-appropriate measures to assess the following parameters:

1. Degree of Finiteness
2. Inflectional Accuracy
3. Affixal Error Patterns
4. Morphological (Syntagmatic & Paradigmatic) Richness

15. Are there humans directly or indirectly participating in your study? *

⊙ Yes

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

Since humans are (in)directly participating in your research, **an Ethical Approval is required**. We will automatically generate the ERB template for you to complete after your DMP has been approved by a data steward.
If you are a student from ID, the link to the ERB application form that you must use will be provided after your DMP has been submitted.

15a. Please check the box that indicates the relevant study population *

- ☑ General healthy population
- ☑ Patients, specifically (write the feature in the box "other")

15b. Please specify *

Individuals with a history of cerebrovascular accident (stroke) and are diagnosed with Broca's aphasia

15c. Age category of participants *

- ☑ 18 years or older

16. Do you collect/process personal data?

- ⊙ Both personal and non-personal data

Personal data are for example, name, address, phone number, email address, IP address, gender, age, video or interview recordings, etc.

Special category personal data are race, religion, health data, political views, genetic or biometric data, sexual preference, trade union membership, criminal convictions and offenses.

17. Please select which of the following (special category) personal data do you collect/process? *

Gender, Age, Date of birth, City or area of residence/postal code, Biometric data, Other..., Health data, Name, Contact details (e-mail, telephone number, ...)

If your type of data is not in the list, please select "Other" and specify them in the box below.

17a. Please elaborate what data will you collect/process. *

- Demographic information (age, gender, date of birth [to compute for the chronological age], occupation, languages spoken) of each participant to describe the groups (this

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

includes health information such as type of stroke, site of brain lesion, handedness, visual or hearing augmentations, and existing speech-language pathology diagnosis, if available);

- Contact information (name, contact details, and residential address) will also be collected as the elicitation sessions may take place in 2 separate days at the participants own home;
- Biometric data: Audio recordings of spontaneous speech samples elicited from all participants.

18. Will you pseudonymize/anonymize the data? *

- ☑ Pseudonymize
- ☑ Anonymize

Anonymization irreversibly removes direct (e.g., name, address) and indirect identifiers (e.g., age, occupation) from data to prevent identification.

Pseudonymization replaces a person's unique identifier with an artificial pseudonym, still allowing identification with a separate key.

19. Please describe how will you pseudonymize/anonymize the data *

Each of the recordings of elicitations and transcriptions will be assigned with a pseudonymous code that is linked to a demographic dataset that is only accessible to the main researcher using an encryption key. The processed and aggregated data are anonymised for publication.

19a. Which tool will you use to pseudonymize/anonymize the data?

ARX

21a. What is the non-personal data you collect? *

Linguistic parameters drawn from the spontaneous speech analyses such as:

1. Percentage of finiteness & accuracy
2. Types and proportions of affixes errors
3. Degree of affixal synthesis for syntagmatic richness
4. Mean size of morphological templates for paradigmatic richness

22. Where will the data come from? (Select all that apply) *

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

- ☑ My own previously collected data (secondary data use)
- ☑ New data collected only by me or my research team

23. Does the study involve sharing and/or receiving data from third parties? *

- ☑ No

Third parties include any entity that shares or receives data, such as companies or collaborators at other universities.

Research Data Management Information

During Research

25. Which of the following tools will you use to process the data (including project documentation)? (Select all that apply, if they are not in the list click on "Other" and add them below, you can add several ones in the same box) *

- ☑ Manual transcription
- ☑ Microsoft Excel
- ☑ Voice/video recording using phone in flight mode
- ☑ Other...

25a. Please provide the name of the tool(s): *

CLAN (Computerised Language Analysis for AphasiaBank)

26. Where will the data (including project documentation) be stored *during* the study? (Select all that apply) *

03 Microsoft OneDrive, 01 TU/e Network Drive

Check the PAR Solution Searcher to find a solution for your project.

Questions? Reach out to your Data Steward by adding a comment at the end of this form.

After approval, you can always update your DMP: if you need to update your DMP just click on the top-right side in the label “Update my DMP” (button only appears after approval), this will open the DMP and you can edit any field and resubmit for revision.

27. How much data (including project documentation) will you approximately collect or reuse? *

- ⊙ 10's of GBs

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

27a. Do you need any special processing/tools for your data? *

- ☑ Data encryptor (Cryptomator Hub)

You can request a Cryptomator hub here.

After Research

28. Will you store (meta)data for future research? (Select all that apply) *

- ☑ Yes, by publishing the (meta)data in a trusted repository under restricted access
- ☑ Yes, by publishing articles (fully theoretical research project)
- ☑ I need advice

Funders (e.g. NWO and ZonMw) require that as minimum research data underpinning publications are made FAIR and published in a trusted repository.

28a. In which repository will you publish your data? *

- ☑ 4TU.ResearchData
- ☑ Other...

28b. Please add the name of the repository *

AphasiaBank

28c. Do you already have one or multiple DOIs? Please add them here.

Did you know that you can reserve a DOI?

Find more information on how to reserve a DOI here.

29. What documentation and/or software code will be deposited with the data? *

- ☑ README file
- ☑ metadata file
- ☑ Logbook
- ☑ Analysis script

Include all documentation which is necessary for you or your colleagues to be able to interpret the data in the future.

For further information, please refer to TU/e RDM.

30. What (license) conditions will apply to the data you will publish? (Select all that apply) *

- ☑ CC-BY 4.0 (free reuse with credits, most used license)

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

31. QUESTION FROM FUNDER: Indicate which metadata standard will be provided to help others identify and discover the data. (Select all that apply) *

☑ DataCite (used by 4TU.ResearchData and Zenodo)

32. QUESTION FROM FUNDER: When will the data be available for re-use? *

⊙ As soon as article is published

Funders (e.g. NWO and ZonMw) require that datasets must be made FAIR at the time of publication of the corresponding research paper.

33. How long will data (including project documentation) be stored after the end of the project? *

☑ 10 years at the TU/e archive (Following the Netherlands Code of Conduct for Research Integrity)

34. QUESTION FROM FUNDER: What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)? *

Eindhoven University of Technology aims to comply with the FAIR principles and share with the scientific community any data obtained in research projects, if ethical and legal regulations permit doing so. Where possible, data will be published via a trustworthy repository, such as 4TU.ResearchData. Via the repository, data will be:

1. ***Findable*** *by indexing data by search engines on the internet, including rich metadata according to, e.g., the DataCite schemas, and receiving a persistent identifier (DOI)*
2. ***Accessible*** *by using an open internet protocol, including clear authorisation procedures, and, where possible, the data will be shared when related articles are published under an open access license (to be determined, but most likely OSI-approved (for code) or Creative Commons),*
3. ***Interoperable*** *by using standards for metadata (e.g., DataCite), by adding documentation (e.g., README files, codebooks), using preferred formats, and using a standard vocabulary if available, and*
4. ***Reusable*** *by including rich metadata, making sure that all data can be opened and used by generally available software (analysis) tools, by adding documentation with instructions for reuse, and by publishing it under an open access license.*

A guide from the VU on RDM costs can be found here.

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

Privacy Checklist Information

35. Will your project involve the processin of personal data on a large scale? *

No

Processing over 10,000 data subjects is generally "large scale." Around 1,000 or more can also qualify. Consider data volume, sensitivity, duration, and geographic scope. Examples: monitoring highway driving, collecting Covid patient data, hospital operations, tracking public transport users in a city.

36. Does this processing activity involve the use of new or innovative technologies? *

no

Examples include combining fingerprints and facial recognition for physical access control, utilizing bodycams in public spaces, and employing innovative technical methods in research, all of which haven't been used by TU/e yet.

37. Does your study involve systematic (c.q. automated) monitoring of persons? *

no

Think about data processing to observe, monitor, or control individuals without their awareness, like highway camera monitoring, tracking email or phone usage, and specific machine learning or AI applications.

38. Will the study include data processing activities that prevent data subjects from exercising their rights or using a service o contract? *

no

Examples include data collection in public spaces that cannot be avoided (e.g. train stations), or operations aiming to allow or not allow usage of a service or entering a contract, such as loan applications.

39. Will the study process personal data t score, rank or profile persons? *

yes

Examples: scoring driving behavior on highways, banks assessing creditworthiness, and organizations profiling website users for marketing.

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

39a. Please explain which personal data will be used and how it will be used? *

For ad hoc analyses of intersubject variability, participants' performance may be ranked to help explain why some individuals outperform others, and to explore associations with readily available personal variables (type of occupation, number of languages spoken, etc).

40. Does your data processing include activities that involves composing "blacklists", such as financial records or credit scores, genetic data, biometric data, health data, camera surveillance data, location/GPS data, internet-of-things data, employee monitoring, etc. *

no

Uncommon in research, but you might indirectly encounter it. It involves processing personal data on criminal convictions, offences, unlawful behaviour, or bad payment history, shared with third parties.

41. Will the processing activity involve automated decision-making with legal effect or similar substantial effect? *

No

Examples:
automated screening participants against a healthcare reference database to determine whether they will receive experimental treatment
automatic screening and possible rejection of applicants based on an algorithm
reaching a decision, based on an automated credit check, on whether or not to give out a loan to a data subject.

Decisions without (major) consequences do not fall under this criterion.

42. Are or will datasets be linked or combined to perform this processing activity? *

No

Consider the combination of two or more databases built with different purposes, which, when combined, exceed the reasonable expectations of the data subjects. Could data subjects have expected these different databases to be combined?

It is less likely for data subjects to expect databases from different parties to be combined; consider a collaboration

Data Management Plan

Submitted by Jonathan Gerona on September 30, 2025, 10:14:07 AM UTC

between the university and a commercial party.

43. In case you answer "Yes" to any of the questions from the Privacy Checklist and you already have an approved Data Protection Impact Assessment (DPIA) for this project, please attach it here:

44. I want my DMP to be manually revised by my data steward? *

Yes

Depending on your answers this DMP might be automatically approved, if you want it to be manually checked by the data steward from your department respond “Yes” to this question.

45. Approval from your supervisor

By checking this box, I confirm that this DMP has been discussed and approved by my supervisor(s).

Please make sure you add your supervisor’s email via the “+Share” button in the right-hand panel.

Click "submit" to send it for review (by your department’s data steward).

Please wait a minute or so and **refresh** the page manually to see the next steps.