

FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

1. General Project Information	
Name Grant Holder & ORCID	Antonietta Gabriella Liuzzi - 0000-0001-8960-5601
Contributor name(s) (+ ORCID) & roles	Rik Vandenberghe – Supervisor
Project number ¹ & title	12AEP24N - THE NEUROBIOLOGY OF MEANING COMPOSITION: A MULTI-METHOD RESEARCH APPROACH
Funder(s) GrantID ²	FWO
Affiliation(s)	X KU Leuven <input type="checkbox"/> Universiteit Antwerpen <input type="checkbox"/> Universiteit Gent <input type="checkbox"/> Universiteit Hasselt <input type="checkbox"/> Vrije Universiteit Brussel <input type="checkbox"/> Other: ROR identifier KU Leuven: 05f950310

¹ “Project number” refers to the institutional project number. This question is optional. Applicants can only provide one project number.

² Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

Please provide a short project description	<p>Comprehending written sentences requires complex combinatorial routines: syntactic parser processing, logical semantics and conceptual structure. By taking advantage of the rare opportunity to record neural signal from depth electrodes (Stereo-electroencephalography (SEEG)) in neurosurgery patients - along with EEG and functional Magnetic Resonance Imaging (fMRI) recordings – this project will reveal the neurobiological foundation of the relationship between syntactic and semantic composition. Scalp EEG recordings during fast periodic visual stimulation (FPVS) will provide a marker of discrimination of sentences among word lists, meaningless sentence structures and meaningful scrambled sentences (Exp 1). The paradigm validated in the EEG experiment (Exp. 1) will be used during SEEG recordings. This will allow us to functionally map the ventral occipito-temporal cortex (VOTC) for selective responses to sentences at several level of combinatorial processing (Exp 2). Two fMRI experiments on healthy individuals will determine whether any level of awareness (passive or active) generates a semantic representation consistent with the semantic representation derived by a sentence transformer model (Exp 3). Finally, FPVS-EEG recordings on progressive aphasia patients (PPA) will determine whether spared sentence comprehension, despite single-word comprehension deficit, of semantic variant PPA is related to a specific combinatorial routine: thematic role assignment (Exp 4).</p>
--	--

2. 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 ³.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Informed consent	Informed consent for healthy volunteers and patients	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input type="checkbox"/> Digital <input checked="" type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input type="checkbox"/> Images <input type="checkbox"/> Sound <input type="checkbox"/> Numerical <input checked="" type="checkbox"/> Textual <input type="checkbox"/> Model <input type="checkbox"/> Software <input type="checkbox"/> Other:	N/A	<input checked="" type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	N/A
Sentences	2 sets of 192 sentences will be manually created. Sentences will be validated by 30 students	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input type="checkbox"/> Images <input type="checkbox"/> Sound <input type="checkbox"/> Numerical <input checked="" type="checkbox"/> Textual <input type="checkbox"/> Model <input type="checkbox"/> Software <input type="checkbox"/> Other:	Excel	<input checked="" type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	
EEG and SEEG paradigm	Oddball paradigm for EEG and SEEG generated with SinStim Software	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input checked="" type="checkbox"/> Images <input type="checkbox"/> Sound <input checked="" type="checkbox"/> Numerical <input checked="" type="checkbox"/> Textual <input type="checkbox"/> Model	Java Excel PNG	<input type="checkbox"/> < 1 GB <input checked="" type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	

				<input checked="" type="checkbox"/> Software <input type="checkbox"/> Other:			
fMRI paradigm	Oddball paradigm for fMRI generated with Presentation NMBS	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input checked="" type="checkbox"/> Images <input type="checkbox"/> Sound <input checked="" type="checkbox"/> Numerical <input checked="" type="checkbox"/> Textual <input type="checkbox"/> Model <input checked="" type="checkbox"/> Software <input type="checkbox"/> Other:	.sce .log PNG MATLAB SCRIPTS	<input checked="" type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	
EEG and SEEG data	24 healthy volunteers, 30 epileptic patients and 30 PPA patients will perform a EEG experiment	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input type="checkbox"/> Images <input type="checkbox"/> Sound <input type="checkbox"/> Numerical <input type="checkbox"/> Textual <input type="checkbox"/> Model <input type="checkbox"/> Software <input checked="" type="checkbox"/> Other:	.bdf Excel JAVA	<input type="checkbox"/> < 1 GB <input checked="" type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	
fMRI data	48 healthy volunteers will perform two fMRI experiments	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input type="checkbox"/> Images <input type="checkbox"/> Sound <input type="checkbox"/> Numerical <input type="checkbox"/> Textual <input type="checkbox"/> Model <input type="checkbox"/> Software <input checked="" type="checkbox"/> Other:	.dcm (dicom) .nii (nifti) MATLAB SCRIPTS	<input type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input checked="" type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	

³ Add rows for each dataset you want to describe.

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should be described under documentation/metadata.
[RDM Guidance on data](#)

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.	N/A
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.	<input checked="" type="checkbox"/> Yes, human subject data; provide SMEC or EC approval number: S65397 <input type="checkbox"/> Yes, animal data; provide ECD reference number: <input type="checkbox"/> Yes, dual use; provide approval number: <input type="checkbox"/> No Additional information:
Will you process personal data ⁴ ? 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).	<input checked="" type="checkbox"/> Yes (provide PRET G-number or EC S-number below): PRET_G-2021-3844 / S65397 <input type="checkbox"/> No Additional information:
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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please comment:

⁴ See Glossary Flemish Standard Data Management Plan

<p>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 to what data they relate and what restrictions are in place.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:</p>
<p>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 to what data they relate and which restrictions will be asserted.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:</p>

3. Documentation and Metadata

<p>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).</p> <p><i>RDM guidance on documentation and metadata.</i></p>	<p>EEG, SEEG and fMRI tasks will be created by means of Sinstim, MATLAB and Presentation NMBS software. SinStim is a JAVA software with a user-friendly interface for running EEG and SEEG tasks. While running, the software generates an excel file with details such as timing, stimulus presented, stimulus onset and responses. Presentation NMBS software provides scenario files and logfiles. Scenario files are used to display a fMRI task on a screen; logfiles are automatically compiled while the task is running and contain details such as type of stimulus presented, stimulus onset, stimulus duration as well as if and when a motor response was provided during the task.</p> <p>Additionally, a .txt file will be created with a clear description of the main parameters adopted to create the EEG/SEEG and the fMRI tasks. The files will store information such as event sequence and related timing, randomizations within run and across runs, number and type of stimulus adopted. In addition, for each paraments, a path to specific scripts will be included.</p> <p>EEG and SEEG pseydonymized data will be exported as .bdf files. These files contain the EEG and SEEG signal with related information: size, frequency of acquisition, number of sequences, onset of stimulus and onset of motor responses. MRI/fMRI pseudonymized data will be exported as Dicom (.DCM) files. Dicom files provide metadata containing information about the image data, such as the size, dimensions, bit depth, modality used to create the data, and equipment settings used to capture the image. DICOM metadata are readable in Matlab by means of the dicominfo function, which returns all information in a Matlab structure.</p> <p>For each volunteer/patient, a .txt file will be created where details occurring the day of the data acquisition will be reported (e.g. whether a subject decided to interrupt the study. Whether technical problem occurred etc.).</p> <p>Statistical analysis will be performed by means of Letswave 6, SPM, Cosmo, SPSS, Matlab or Python. Similarly to DICOM, SPSS provides metadata embedded in the software.</p> <p>When data will be analyzed by means of Matlab or Python, a .txt file will be generated describing the procedure. In case of several scripts will be created, the .txt. file will contain a specific description for each script.</p>
---	--

<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?</i></p>	<p><input checked="" type="checkbox"/> Standard back-up provided by KU Leuven ICTS for my storage solution</p> <p><input checked="" type="checkbox"/> Personal back-ups I make (specify)</p> <p><input type="checkbox"/> Other (specify)</p>
<p>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.</p>	<p><input checked="" type="checkbox"/> Yes: Sufficient storage & backup capacity are ensured by the host lab and the ICT services of Biomedical Sciences.</p> <p><input type="checkbox"/> No</p> <p>If no, please specify:</p>
<p>How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p> <p><i>CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.</i></p> <p><u>Guidance on security for research data</u></p>	<p>Data stored on the J-drive are accessible by Lab members only.</p> <p>Data on the KU Leuven One-Drive for Business are accessible only by means of my personal U-number and password.</p> <p>Informed Consent Forms will be stored in the laboratory's secure environment for private data. This environment is always locked. Access to the key is provided only to authorized researchers from the lab. No external person can have access to the key.</p>
<p>What are the expected costs for data storage and backup during the research project? How will these costs be covered?</p>	<p>The costs regarding data storage, backup and preservation are covered by data storage resources available at LCN, managed by Prof. Rik Vandenberghe and Prof. Patrick Dupont.</p>

5. Data Preservation after the end of the Research Project

<p>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...).</p> <p>Guidance on data preservation</p>	<p><input checked="" type="checkbox"/> All data will be preserved for 10 years according to KU Leuven RDM policy</p> <p><input type="checkbox"/> All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans</p> <p><input type="checkbox"/> Certain data cannot be kept for 10 years (explain)</p>
<p>Where will these data be archived (stored and curated for the long-term)?</p> <p><i>Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.</i></p>	<p><input type="checkbox"/> KU Leuven RDR</p> <p><input type="checkbox"/> Large Volume Storage (longterm for large volumes)</p> <p><input type="checkbox"/> Shared network drive (J-drive)</p> <p><input checked="" type="checkbox"/> Other (specify): Archive Disk (K-Drive)</p>
<p>What are the expected costs for data preservation during the expected retention period? How will these costs be covered?</p>	<p>Expected costs for data preservation (Archive Disk) are 1500 euro. The costs are covered by data storage resources available at LCN, managed by Prof. Rik Vandenberghe and Prof. Patrick Dupont.</p>

6. Data Sharing and Reuse

<p>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.</p> <p><i>NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION:</i></p> <p>https://wiki.surfnet.nl/display/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS</p>	<p><input type="checkbox"/> Yes, as open data</p> <p><input type="checkbox"/> Yes, as embargoed data (temporary restriction)</p> <p><input checked="" type="checkbox"/> Yes, as restricted data (upon approval, or institutional access only)</p> <p><input type="checkbox"/> No (closed access)</p> <p><input type="checkbox"/> Other, please specify:</p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>Access will be considered after a request is submitted explaining the planned reuse. Pseudonymized data will be made available only for research purposes. Any commercial reuse will be excluded.</p>
<p>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.</p>	<p><input type="checkbox"/> Yes, privacy aspects</p> <p><input type="checkbox"/> Yes, intellectual property rights</p> <p><input type="checkbox"/> Yes, ethical aspects</p> <p><input type="checkbox"/> Yes, aspects of dual use</p> <p><input type="checkbox"/> Yes, other</p> <p><input checked="" type="checkbox"/> No</p> <p>If yes, please specify:</p>
<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p><input type="checkbox"/> KU Leuven RDR</p> <p><input type="checkbox"/> Other data repository (specify)</p> <p><input checked="" type="checkbox"/> Other (specify): OpenNeuro (https://openneuro.org/) and GitHub (https://github.com/)</p>

When will the data be made available?	<input checked="" type="checkbox"/> Upon publication of research results <input type="checkbox"/> Specific date (specify) <input type="checkbox"/> Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. <i>A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.</i> Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.	<input type="checkbox"/> CC-BY 4.0 (data) <input type="checkbox"/> Data Transfer Agreement (restricted data) <input checked="" type="checkbox"/> MIT licence (code) <input type="checkbox"/> GNU GPL-3.0 (code) <input checked="" type="checkbox"/> Other (specify): Datasets (the preprocessed/defaced/anonymized fMRI data in this case) hosted on OpenNeuro fall under the Creative Commons CC0 license, and the code hosted on GitHub will be hosted with a MIT license.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. <i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i>	<input checked="" type="checkbox"/> Yes, a DOI will be added upon deposit in a data repository <input type="checkbox"/> My dataset already has a PID <input type="checkbox"/> No
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected for data sharing.

7. Responsibilities

Who will manage data documentation and metadata during the research project?	Antonietta Gabriella Liuzzi
--	-----------------------------

Who will manage data storage and backup during the research project?	Antonietta Gabriella Liuzzi
Who will manage data preservation and sharing?	Antonietta Gabriella Liuzzi, Patrick Dupont, Rik Vandenberghe
Who will update and implement this DMP?	Antonietta Gabriella Liuzzi