GELIFES - STANDARD RDMP TEMPLATE

The template below reflects the general institute policy on data management, and can be used for most research projects. The relevant sections are pre-filled with the default institute standards (**bold** text) to be used as-is or to be extended when necessary, e.g., in case of joint projects with external institutes or organizations. Obligatory fields are indicated by * and if applicable, must be completed for each project.

Since 25 May 2018, studies with human subjects require registration to be compliant with the EU General Data Protection Regulation (GDPR). The RUG manages a central register with 13 legally approved, obligatory questions in **section 2.2** that must be completed for each study processing personal data from human subjects.

Send the completed RDMP to gelifes-data@rug.nl

WHAT to file	RDMP	Archiving primary data	Archiving secondary data
MSc research projects			
RUG students intern	**yes**	yes	yes
RUG students extern	**yes**	no	yes
Non-RUG students	**no**	no	no
MSc colloquia & essays	**no**	no	no
PhD projects			
RUG PhDs intern	**yes**	yes	yes
RUG PhDs extern	**yes**	no	yes
Non-RUG PhDs	**no**	no	no
All other projects			
RUG-affiliated postdocs	**yes**	for each first/equal author publication	for each first/equal author publication
RUG-affiliated staff	**yes**	for each first/equal author publication	for each first/equal author publication
RUG-affiliated honorary profs	**yes**	for each first/equal author publication	for each first/equal author publication

WHEN to file it	RDMP	Data archive
MSc research projects	**within 1 month after starting the project**	within 1 month after grading the project
PhD projects	**within 6 months after starting the project**	within 1 month after formally handing in the thesis
		for each published chapter: within 3 months after the publication appears online

All other projects	**within 6 months after starting the project**	within 3 months after the publication appears online

1. **Administration**	
1.1 Name of the student/ researcher	*Add your s-number (MSc students) or p-number (PhD students & staff).* * Name: Wenjun Meng * S/P-number: S5613329
1.2 Name of the research group	* Research group: Voice Technology, supervised by Associate Prof. Dr. Matt Coler
1.3 Name of the project	* Project title: Beyond Adult Speech: Exploring SepFormer's Performance in Child Speech Separation
1.4 Description of the project	*Briefly (3-5 sentences) describe the project and its expected results.* * Project description: This thesis investigates the performance of SepFormer, a state-of-the-art speech separation model, in processing child speech, which has been less explored compared to adult speech. The study aims to evaluate the effectiveness of SepFormer in separating speech in datasets comprising child speech, with the hypothesis that SepFormer's performance will significantly decline due to the unique acoustic properties of child speech. The research utilizes the PhonBank database and employs evaluation metrics such as Scale-Invariant Signal-to-Noise Ratio and Signal-to-Distortion Ratio to assess performance.
1.5 Funding agency	*For 2nd or 3rd money flow projects: indicate the funding agency and grant number.* * Funding agency: N/A * Grant number: N/A
1.6 Institute project code	*If applicable:* * Project code: N/A
1.7 Project duration	* Start date: April 5th 2024 * Expected end date: June 11th 2024 Extended end date: N/A
1.8 DMP version	* Version 1.0: April 5th 2024 Last updated: N/A Version 2.0: N/A
1.9 Project data manager(s)	*Indicate all persons responsible for the project's data management. Also include daily supervisors if different from the primary PI.* * (First) promotor/PI responsible for the project: Associate Prof. Dr. Matt Coler * Daily supervisor(s) if other than PI: N/A
1.10 External collaboration	*In case of joint projects: indicate which external institutes/organizations are involved and who (PI) is responsible for data management at the external institute.* * Institute/organization: N/A * PI(s) responsible for externally collected data: N/A
2. Data collection	
2.1 Legal issues	*For (legally) sensitive or confidential project data: indicate which special procedures, permits and/or licenses are required.* * Project data that require specific legal procedures, permits and/or licenses: ☑ Human subjects: GDPR ☐ Animal lab and/or field experiments: add license(s) ☐ Imported plant or animal specimen: add license(s) ☐ Genetically modified organisms (GMO): add license(s) ☐ Radioactive chemicals/isotopes: add license(s) ☐ Other confidential/sensitive data: Specify ☐ No confidential/sensitive data collected

2.2 GDPR registration	*Studies with human subjects require registration to be compliant with the EU General Data Protection Regulation (GDPR) since 25 May 2018. The UG manages a central register with 13 standard, legally approved questions that have to be answered for each study processing sensitive data from human subjects.* *** NOTE: This section is obligatory for all studies processing data from human subjects** 1. Data processing activity: analyzing in research project 2a. Responsible faculty/institute: **FSE-GELIFES** 2b. Local process owner: **Senior PI responsible for the project** 2c. People with access to the data: ☐ 1-5 ☐ 5-10 ☐ 10-100 ☑ > 100 3. Intended use of the data: research 4. Legal basis for processing personal data: **To be filled in by Legal Department** 5a. Human subject category: ☐ Staff ☐ Students ☑ External research subjects ☐ Patients ☐ Clients ☐ Other, specify: Click here to enter text. 5b. Estimated number of subjects: ☐ 1-10 ☑ 10-100 ☐ 100-1000 ☐ 1000-10,000 ☐ 10,000 - 100,000 ☐ > 100,000 6. Personal data collected: ☐ Name and address details ☐ Unlisted address ('secret' address, not in Dutch Personal Records Database (BRP)) ☐ Nationality ☐ Date of birth ☐ Place of birth ☐ Health information ☐ Religion ☐ Criminal record ☐ Biometric information (e.g. fingerprints/iris scan) ☐ Photo/video ☐ Telephone number ☐ BSN (Dutch Citizen Service Number)/V-number (Foreigner registration number) ☐ Email address ☐ Location information ☐ IP address ☑ Other, specify: audio 7a. Secure storage of personal data: **Select the appropriate options in section 3.1** 7b. Retention period for personal data: **Standard archiving conditions apply unless indicated otherwise in section 3.2** 8. External third party providing personal data: ☑ no ☐ yes, specify: Click here to enter text. 9. External third party processing/storing personal data: ☑ no ☐ yes, specify: Click here to enter text. 10. Data sharing with countries outside the EER (EU, Norway, Iceland, Liechtenstein): ☑ no ☐ yes, specify: Click here to enter text. 11. Privacy Impact Assessment (PIA) for data sharing outside EER available: ☑ no ☐ yes ☐ unknown 12. Automated sharing of personal data with other systems: ☐ no ☑ yes, specify: the University of Groningen Thesis Repository 13. Technical and/or organizational security measures taken to protect the personal data: ☐ Encryption ☐ Pseudonymization ☑ Anonymization ☐ Access limited to specific user groups

2.3 Description of the raw (primary) data	*Indicate what type of data will be collected and whether a very large amount of data is expected (> 500 GB). Also indicate where the data will be collected (own/external institute), and whether the data is new or re-used from another project.* * Primary data collected: ☐ Scanned or electronic field logs & score forms ☐ Scanned or electronic lab journals & score forms ☐ Pictures of gels or microscopic observations ☐ Output from data loggers ☑ Video & audio recordings ☐ Webcam/photo identification files ☐ Satellite/aerial imagery ☐ Sequencing & genotyping data ☐ Micro array & hi throughput data ☐ Other, specify: Click here to enter text. ☐ No primary data * Expected amount of data: ☑ < 10 GB ☐ 10-200 GB ☐ > 200 GB Additional remarks: Click here to enter text.
2.4 Description of the processed (secondary) data	*Indicate whether & how the raw data will be processed: what software, algorithms, workflows will be used?* * Secondary data produced: ☐ Spreadsheets, databases & graphics ☐ Output from statistical packages ☐ Output from geographic information systems ☑ Simulated datasets ☐ Program code (e.g., C/C++, NetLogo, Matlab, Maple, Mathematica) ☐ Batch scripts (R, Python) ☐ Associated parameter files used to produce or process primary data ☐ Other, specify: Click here to enter text. ☐ No secondary data * Expected amount of data: ☑ < 10 GB ☐ 10-200 GB ☐ > 200 GB Additional remarks: Click here to enter text.
3. Data storage	
3.1 Data storage	*Indicate how & where primary and secondary data are stored during the project: how often are stored data backed-up; how is version control realized? If relevant: indicate how the privacy of test subjects is guaranteed.* * Long-term storage with automated back-up: ☑ **RUG servers (wiki, iRODS, low-cost storage, eLabjournal)** ☐ **UMCG servers (Lifelines or similar)** ☐ **DANS DataverseNL (open access)** ☐ Other, specify: Click here to enter text. ☐ None * Cloud-based storage & sharing: ☑ **RUG-Unishare** ☐ GitHub ☐ Google Drive ☐ MS OneDrive ☐ DropBox ☐ WeTransfer ☐ Other, specify: Click here to enter text. ☐ None * Short-term storage: ☑ **RUG PCs** ☐ NAS/external HD ☐ Laptop ☐ Flash memory ☐ Other, specify: Click here to enter text. ☐ None Special storage facilities: Click here to enter text.

3.2 Data archiving	*Indicate how, where & for how long primary and secondary data will be archived, including specific software that is used or other relevant information needed to verify the research results.* **All primary and secondary data are archived in a single zip-file, also including the final report, thesis or publication of the project and one or more metadata files that document the data archive. The archive is stored in the institute repository; archives will be stored for at least 10 years.** Additional archiving requirements/agreements: Click here to enter text.
4. Data	
	If applicable: Indicate whether specific agreements apply with regard to data ownership or access, e.g. contractual obligations to third parties (commercial or other), embargo requirements for publication, etc. **The research institute acts as the legal representative of the University of Groningen and holds ownership of all research data collected, generated or otherwise acquired within the institute unless different arrangements are clearly documented and approved by the institute director.** **The purpose of the institute data repository is archive-only; access to archive files is restricted to senior staff members ultimately responsible for the corresponding research projects.** Additional agreements on data ownership and/or access: Click here to enter text.
5. Data documentation	
5.1 Metadata	*Indicate how & for whom metadata (data documentation) is organized.* **The data archive includes a standard metadata file (read_me_first.txt) that is set up according to the guidelines in the institute data archiving procedure; for archives consisting of multiple folders a read_me_first.txt metadata file is included for each folder**.
5.2 File formats & naming standards	*Indicate which data formats will be used in the project; if non-standard data formats are used: explain how these formats can be read (include the appropriate software or a link to its source).* **The project aims to use the preferred or acceptable data formats listed in the guidelines in the institute data archiving procedure. All file names, metadata and other description files and comment lines in code are in English. The data archive follows the naming standards provided in the institute data archiving procedure.** Non-standard data formats used: Click here to enter text.