

DATA MANAGEMENT PLAN

PROJECT	
Project number:	[project number]
Project acronym:	[InSpeech]
Project name:	[Empirical and theoretical assessment of the links between inner speech and insight]

DATA MANAGEMENT PLAN	
Date:	[15/04/2025]
Version:	[1.0]

1. Data Summary

The given research is a new experiment made specifically for the current MSCA project and will generate new data, as a result of the empirical phase. The research will not re-use any existing data. The aim of the experiment is to gather participants' data on how they experience inner speech while facing a complex problem (problem solving task). The data will be both quantitative and qualitative. During the experiment the participants will solve word puzzles (Compound Remote Associates Test) with or without a parallel activity (saying words repeatedly or listening to texts) and then undergo a short semi-structured interview to deeper explore the specific experience of inner speech they had. Thus the generated data will consist of 3 parts (for each participant):

1. The participants' data (will be pseudonymized): name, email address, gender and age.
2. Performance data on CRA tests: solution times (in milliseconds), accuracy (numerical - 0 or 1), number of correctly solved puzzles (numerical), type of inner speech suppression used (auditory, articulatory or non), type of problem solving strategy for each solution (insight or analytical).
3. Qualitative data from the semi-structured interview (answers for the questionnaire).

The project is expected to generate a maximum of 20 MB of data, including maximum of 10 MB of the quantitative data (CRA test in .crv format) and maximum of 10 MB of qualitative data (interview responses in .crv format). The generated data will be used for the further analysis of the research, constituting research outcomes and developing the theoretical framework as one of the key results of the study. The data will be also useful for other researchers in the field of inner speech and problem solving, in order to develop further exploratory research. The data can also be used outside of academic purposes to contribute to the awareness of inner speech experience in general population.

The generation, analysis and preservation of the data will follow the FAIR principles, the Guidelines on FAIR Data Management in Horizon 2020 (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf) and the KU Leuven RDM policy (<https://www.kuleuven.be/rdm/en/policy>) for data management.

2. FAIR data**2.1. Making data findable, including provisions for metadata**

The research will ensure the findability of the generated data. It will thus be uploaded to a trusted open data repository Zenodo and will be assigned with a persistent identifier (PI). The rich metadata will also be added to the dataset to ensure the possible re-use. The rich metadata will be provided for the dataset to allow the easier discovery of the data. The metadata will include the information on the project (its description and title), main contributors, types of resources, funding information, affiliated institutions, information on creation and persistent identification. The metadata will also include a readme file to support the re-use of the data. The metadata will be supported with the relevant keywords to ensure the optimal discovery and potential re-use of

the data.

2.2. Making data accessible

Repository:

The generated data for the research will be deposited in a trusted repository – Zenodo, which will also ensure the persistent identifier issued for the deposited data, and provide the possibility to resolve the identifier to the digital files.

Data:

All the data generated during the experiment will be made openly available except for the names and email addresses of the participants to ensure the security and safety of participants' personal data. The possible embargos on making the generated data openly available in the repository may apply due to the specific guidelines of the publishing houses. Depositing the data to the open repository Zenodo will make the data accessible through free and standardized protocols such as HTTP and OAI-PMH, and will not require the identification of the person accessing the data. The generated data shared on the repository will not include any personal or sensitive information on the participants and thus will not require a data access committees nor additional evaluations.

Metadata:

The metadata related to the research will be openly available and licensed under a public domain (CC0), in accordance to the Grant Agreement. All the relevant documentation or links to the software necessary to re-use the data will be a part of the metadata set. The metadata will be made searchable and indexed via DOI using DataCite servers. The researchers will ensure the long-term availability, validation and re-use of the generated data, making it openly available on Zenodo for a minimum of 10 years or during the lifetime of the repository. The metadata will remain available even in case the actual data is no longer available and will be retained for the lifetime of the repository. According to the principles of Zenodo (<https://about.zenodo.org/principles/>), the metadata will be stored in the highly accessible servers of the CERN laboratory, and the lifetime of the data accessibility is linked to the lifetime of the laboratory/servers with the current experimental programme being defined for at least 20 years.

2.3. Making data interoperable

The researchers will follow the community-endorsed best practices to ensure the interoperability of the deposited data and metadata. The data will be stored in open formats such as CSV. The metadata will be stored in JSON Schema in accordance to Zenodo principles (<https://about.zenodo.org/principles/>) to ensure its interoperability. The metadata will refer to open external vocabularies and ontologies. The metadata stored on repository will follow the FAIR principles and rely on the best practices of interoperability of data and metadata.

2.4. Increase data re-use

In order to validate data analysis and facilitate the re-use of the generated data, the necessary documentation will be provided in a ReadMe file as a part of the metadata set, that will contain the information on methodology, data cleaning, analysis, variable definitions and interviewing process. The data will be stored in a freely available way in the public domain to permit its widest possible re-use. The data will be licenced using clear and accessible standard licenses, CC-BY for the generated data and CC0 for the metadata in line with the Grant Agreement. This will ensure the possibility for the third parties to re-use the generated data and metadata after the end of the project. The data will also be associated with the detailed provenance in line with Zenodo principles (<https://about.zenodo.org/principles/>), and the data will be traceable to a registered Zenodo user.

Data quality assurance process: In order to ensure the quality of the data collected, the collection process will be carefully tested during the pilot study. The data collection will be supervised by the experienced researchers and the necessary training will be provided to the researchers and interns involved in data collection. The data collection of both the quantitative and qualitative parts of the experiment will follow the standardized procedures. After the data collection the manual checks will be implemented to ensure the quality and correctness of the inserted answers during the quantitative study. The qualitative phase with the interview will allow the additional verification of the received data. The documentation of both data and metadata will follow the FAIR principles. The generated scripts for the CRA tests will be peer-reviewed by the qualified team members and will also be shared according to the FAIR principles.

3. Other research outputs

The methodology of the research presupposes the generation of the script for the experiment that will allow the fluent computer-based study using CRA tests. The script will be developed by the researchers and then deposited with the necessary metadata to the open repository Zenodo following the FAIR principles. The researchers will ensure the accessibility and interoperability of the script for the verification, re-use and further research.

4. Allocation of resources

The collection of the data, its analysis and preservation will be the responsibility of the researchers involved in the current project (Aleksandr Fadeev, Eva Van Den Bussche). There are no additional costs related to the making of data in terms of staff, as this will be done by the researchers themselves and the involved interns. The participants of the research will be KU Leuven students of the Psychology department who will receive study credits for participating in the experiment. The software used for the collection of the data, its analysis and preservation is free publicly available software, i.e. Psycho-Py and RStudio. The software used for temporarily storing the data, namely Microsoft Cloud solution, is provided by KU Leuven for the researchers free of charge. The trusted repository used for the preservation and re-use of data, Zenodo, is free and publicly available. The additional charges might occur during the publication process within the Open Access framework, which (together with any other occurring costs related to data management and publication) will be covered by the MSCA project budget according to the Grant Agreement. The long term preservation of the data will be ensured by the existence of the Zenodo servers which are linked to their experimental project which is meant for the minimum of 20 years.

5. Data security

The data collected during the research will be safely stored via the KU Leuven Microsoft Cloud solution to enhance the fluent process of data collection and analysis, as well as to ensure the security of the sensitive data (in the case of this study – names and email addresses of the participants). The personal data of the participants will be pseudonymized, and only the main researchers of the given study will be able to trace back the data to the individual participants. The third parties will not be able to get access to the participants' personal data. After the study the data will be deposited to the trusted data repository Zenodo for the long term preservation and following re-use according to the FAIR principles.

6. Ethics

There are no specific ethical or legal issues that may have an impact on data sharing. The informed consent provided to the participants will give the information regarding the collection, storage, possible re-use and safety of the participants data collected during the study.

7. Other issues

The study is supported through The Marie Skłodowska-Curie Actions (MSCA) and thus follows the recommendations and principles of EU commission on the good practices for data management in Horizon Europe projects outlined in the “Guidelines on FAIR Data Management in Horizon 2020” (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0		Initial version