

## PLAN ZARZĄDZANIA DANYMI [w języku angielskim]

**1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych**

Sposób pozyskiwania i opracowywania nowych danych i/lub ponownego wykorzystania dostępnych danych

The project aims to obtain new data from at least  $n = 40$  participants. After the university's ethics committee approval is achieved, all participants will be asked to sign written informed consent prior to their participation in the project. Data will be acquired through video and voice recording and managed by the PI and team members. The data will be anonymized and transcribed, i.e., converted to text files. The text and raw data files will be annotated using the ELAN and TRANSANA Multiuser software. The codebooks will have a name and contain the date of the file's creation.

Pozyskiwane lub opracowywane dane (np. rodzaj, format, ilość)

At least 40 videos (HD, 4k) will be recorded. The video files will be saved in MPEG4 format. The corresponding audio files will be in WAV format for archiving purposes, whereas the transcriptions will be saved in text files (RTF and TXT formats). Selected images from the videos used for presentations and publications will be saved in PNG format. Tabular data will be saved in XLSX format. Approximately 2T will be necessary for data storage. The files may be converted to other formats for easy sharing (for example, WAV to MP3, XLXS to CSV, and PNG to JPG).

**2. Dokumentacja i jakość danych**

Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym

All data will be assigned metadata: content and technical. Metadata will be as close as possible to files and folders. Intuitive structure and naming of the files and folders will be created. All data corresponding to every interaction will be given a unique ID number, and the same ID number will be used to organize the datasets. The data from every interaction (including corresponding media files, transcriptions, and annotated files/codebooks) will be saved in the folders described with text codes. The metadata will be in English. A file README.TXT will be created for each folder. Dates in the following format, "YYYYMMDD," will be used to order original and modified data. Modifying the data will require a new signature, date stamp, and comment (to the README.TXT) regarding the newly made changes. The README.TXT will also contain all software/equipment-specific information necessary to annotate the data, data collection dates, data collection settings, people involved in data collection and processing, and funding sources that support data collection.

Stosowane środki kontroli jakości danych

The data quality will be evaluated by the PI and co-investigators. High-quality data will be selected. Data that will be excluded will be generated once more. Transcriptions in text formats will be corrected by other team members, if necessary. The equipment will be tested at the beginning of the project and always before videorecording (e.g. settings of cameras). Data files will be controlled every four weeks.

**3. Przechowywanie i tworzenie kopii zapasowych podczas badań**

Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

All data (video and audio) will be downloaded from the devices' memory cards to an acquisition computer. Data will be stored and backed up following the 3-2-1 backup rule. Accordingly, we will keep three copies of data: the original data will be stored on an acquisition computer, and the two backup copies will be stored on different storage media: one on a notebook and the other on a portable device of 2TB volume. During data collection, data backup will be carried out every week. After that, data backup will be carried out every four weeks. Finally, we will ensure that one copy is stored offsite (in a locked cabinet in a university office).

Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

The project requires the ethical approval of the study's research plan by the research ethics committee of the host institution. For data collection and backup, we will use a portable device stored in a locked cabinet in the host institution, accessible only to the PI, and a password-protected computer. All data will be used solely on the equipment provided by the host institution and protected with ESET software. The participants' personal data, written informed consent files, and other sensitive data will be available only to the PI and the collaborating psychologist and stored in a locked cabinet in the host institution. Files and folders containing sensitive data will be password-protected. All team members will sign declarations regarding confidentiality and personal data protection.

#### 4. Wymogi prawne, kodeks postępowania

Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych

##### **If personal data are processed, how will compliance with legislation on personal data and data security be ensured?**

The team will receive training on personal data collection and management from the host institution. Data collection and management will be carried out in accordance with the internal regulations of the host institution and the General Data Protection Regulation (GDPR). All data collected in the project will be anonymized, analyzed, and shared based on informed consent and/or after notifying all concerned persons. All sensitive data will be password-protected or locked in a cabinet in the host institution. The PI will have full data access and assign consent to specific team members. Participants will be informed about the project and about their right to withdraw at any time.

Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością. Obowiązujące przepisy

Copyright for the project's contents belongs to the PI and the host institution, i.e., Nicolaus Copernicus University in Toruń.

The issues of managing intellectual property rights are regulated by Resolution No. 11 of the Senate of the Nicolaus Copernicus University in Toruń on January 22, 2013. The copyright owner will aim at wide accessibility and dissemination of selected data and results obtained in the project. Selected anonymized data will be shared by license Creative Commons for non-commercial purposes only.

#### 5. Udostępnianie i długotrwałe przechowywanie danych

Sposób i termin udostępnienia danych. Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga

The anonymized data collected and analyzed in the project will be made available to the academic and general public community only if the participants grant written informed consent. Raw data will be stored on a portable disk for ten years. Selected data will also be shared via the NCU Open Data Repository on the RepOD platform within the last three months of the project. The data will be available based on a Creative Commons open license. The data will be useful for linguists, social interaction researchers, and psychologists.

Sposób wyboru danych przeznaczonych do przechowania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)

All data collected in the project will be stored on a portable device and locked in a cabinet in the host institution. Selected data from the project, such as codebooks/annotated textual and tabular data/ and selected interactional multimodal data will be made available via a repository provided by the host institution (RepOD), if the participants have granted their written consent.

Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych

Internet access and an Internet browser are required to access data stored in the host institution repository on the platform RepOD. The FAIR Principles will be met by describing data with metadata, publishing data in the repository under an open Creative Commons license, and receiving a DOI.

Sposób zapewniający stosowanie unikalnego i trwałego identyfikatora (np. cyfrowego identyfikatora obiektu (DOI)) dla każdego zestawu danych

All the data items shared via the repository will have URI and DOI, allowing its accurate and easy citations.

#### 6. Zadania związane z zarządzaniem danymi oraz zasoby

Osoba (np. funkcja, stanowisko i instytucja) odpowiedzialna za zarządzanie danymi (np. data steward)

The data will be managed by one of the team members appointed by the PI and under the supervision of the PI at the time of the project. After that, the PI will be in charge of the data management. The team will be able to consult experts on data protection (Data Protection Officer) and data sharing (University Library) employed by the host institution.

Środki (np. finansowe i czasowe) przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych

The purchase of a portable disc drive of 2T is planned in the project's budget.