
Plan Overview

A Data Management Plan created using DMPOnline

Title: Communication-friendly Validity

Creator: Gjalt-Jorn Peters

Principal Investigator: Christine van Vliet, Gjalt-Jorn Peters

Affiliation: Open University of the Netherlands

Template: OU template (NL)

ORCID iD: 0000-0002-1747-446X

ORCID iD: 0000-0002-0336-9589

Project abstract:

The overall ELSA-project (Ethical, Legal, Societal aspects of AI) aims to combat poverty by fostering individuals' capabilities through technology and AI. Leveraging the SPARK (Supportive, Personalized, Adaptive, Resource Kit) platform, an open-source ecosystem, the project aims to understand and influence financial behavior among individuals, especially concerning governmental financial regulations on parental leave. The focal point of this study is the development and validation of template questions aimed at accurate measurement of psychological constructs crucial for effective behavior change interventions that are eventually to be used in the SPARK.

ID: 136548

Start date: 11-12-2023

End date: 31-03-2024

Last modified: 09-11-2023

Grant number / URL: Not applicable

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Communication-friendly Validity

Algemene informatie

Naam aanvrager en projectnaam

Christine van Vliet - Communicatie vriendelijke validiteit

Lijst namen van personen die toegang hebben tot de data en verantwoordelijkheden met betrekking tot data management

Question not answered.

Subsidiegever en projectnummer

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Partnerorganisaties

ZUYD Hogeschool

Naam van de data manager/data steward die bij de opstelling van dit plan is geraadpleegd

We will consult Mellanie Geijen, our data steward.

Datum van overleg met datamanagementpersoneel

Question not answered.

Data verzameling

Gaat u persoonsgegevens verwerken en/of opslaan tijdens uw project?

- Ja

We will record the interviews (audio). Although the interview questions will not ask about personal data, recorded voice data can potentially be considered as personal data. In addition, we cannot a priori exclude the possibility that participants will spontaneously divulge personal data.

Geef aan op welke manier de data voor het onderzoek wordt verzameld.

- Nieuwe data genereren

Als er bestaande data wordt hergebruikt/gekoppeld: geef aan of u toestemming heeft van de data-eigenaar/eigenaren voor het gebruik/koppelen van zijn/hun data.

- Nee, ik ga geen bestaande data hergebruiken of koppelen

Als er nieuwe data worden geproduceerd: beschrijf de data die naar verwachting zullen voortkomen uit uw onderzoek, inclusief het formaat van de data die worden verzameld of geproduceerd.

We will record audio data, which will subsequently be transcribed and then anonymized.

Welke meetinstrumenten, software worden gebruikt om de data te genereren/verwerken?

We will collect data to study a 19-question measurement instrument developed based on the Reasoned Action Approach.

Hoeveel dataopslag is er in totaal nodig voor uw project?

- 0 – 10 GB

Geef aan of er bij het verzamelen van data samen wordt gewerkt met andere partijen.

- Ja, het verzamelen van nieuwe data doe ik samen met andere onderzoekers, onderzoeksgroepen.

We will collaborate with the Hogeschool Zuyd and Maastricht University.

Geef aan of het project uitgevoerd wordt in een consortium met twee of meer partners. Binnen het consortium zijn duidelijke afspraken gemaakt over datamanagement en intellectueel eigendom.

- Ja, er zijn duidelijke afspraken gemaakt m.b.t. datamanagement en intellectueel eigendom d.m.v. een consortiumovereenkomst.

We have made clear agreements to work according to Open Science principles. In addition, note that unless one invokes database rights, anonymized data cannot be owned by anybody (facts exist in the public domain) and personal data can only be owned by the corresponding person.

Data opslag en beveiliging

Wat zijn de belangrijkste risico's voor data beveiliging?

- Data lek, onbevoegde toegang of onbevoegd gebruik

The biggest risk (negligible given our procedures, yet the biggest) is the risk that one of the audio files is obtained by an attacker.

Welke maatregelen worden gehanteerd om te voldoen aan de veiligheidseisen en om bovenstaand risico's te beperken?

- Toegang beperking
- Encryptie
- Verzameling van gevoelige data beperken
- (Beveiligde) kopie van de ruwe data opgeslagen op de netwerkschijf van de OU
- Procedure voor het verwerken van data en/of training

The audio recordings will immediately be encrypted using 7-zip, with 256-bit AES encryption and using unique passwords with at least 128 bits of entropy. These passwords will be stored in a dedicated, project-specific KeePass2 database that is shared among the researchers using a cloud service (note that these files are safely encrypted, and as such, it is safe to use SurfDrive or even other cloud services for these files), and that itself is protected with another password with at least 128 bits of entropy. That master password will be stored in each researcher's password manager (and will initially be shared using either Signal or the OneTimeSecret service).

Once an audio recording is encrypted, the password is stored in the shared KeePass2 database, and the updated KeePass2 file and the encrypted archive holding the audio file have synchronized with the cloud service, the interviewer will ask another researcher to verify that they can access the password and successfully open the encrypted archive and extract the audio file. Once this is confirmed, the original raw audio file will be deleted from both researchers' computers and the recording device.

One of the researchers at the Open University will then upload the raw audio file to the transcription service (AmberScript). The transcript that will be produced is then downloaded, and that raw transcript is added to the encrypted archive. That transcript is then anonymized by the researcher, and if the participant indicated that they wanted to check this, it is shared with them to potentially further anonymize it. The approved version is then made public under the Open Data Commons Open Database License (ODbL; <https://opendatacommons.org/licenses/odbl/>).

This public version will then be coded using the Reproducible Open Coding Kit, and the coded file will also be made public under the Open Data Commons Open Database License (ODbL; <https://opendatacommons.org/licenses/odbl/>).

Beschrijf waar de data worden opgeslagen en geback-up't tijdens het project.

- Overig (licht toe)
- Netwerkopslag voor onderzoekdata bij de instelling

We will use researchdrive and a Git repository, which we aim to sync with OSF, during the project.

Hoe worden de beveiliging en bescherming van gevoelige data tijdens het onderzoek geregeld?

- Standaard veiligheidsmaatregelen van de netwerkopslag voor onderzoeksdata bij de instelling
- Extra veiligheidsmaatregelen (licht toe)

Strong encryption (see above).

Wordt er tijdens het project ook papieren output gegenereerd/verwerkt? Waar wordt deze output opgeslagen?

- Ja (licht toe)

Informed consents are produced. These will be safely archived at the Open University conform to the OU policies.

Data documentatie

Gebruikt u een systeem voor de naamgeving van bestanden of mappen?

- Nee, ik gebruik mijn eigen methode voor naamgeving, welke ik verder zal toelichten

For each interview, a unique identifier is composed of the day and the number of the interview on that day (e.g. always '1' if

there was only one interview; '1' and '2' for the first and second interviews on that day, etc), for example "2024-01-12---x". For the encrypted archives, we will use "2024-01-12---x.zip".

For the data files, we will use "2024-01-12---x---transcript--raw.txt" for the raw transcript as provided by AmberScript; "2024-01-12---x---transcript--clean.rock" for the anonymized version; "2024-01-12---x---transcript--UIDs.rock" for the version with Utterance Identifiers added by the ROCK package; and "2024-01-12---x---transcript--coded.rock" for the coded version.

Other filenames will not be dated (note that we use git for version control), but all filenames will use multi-platform-safe filenames (we will follow the guidelines in <https://psy-ops.gitlab.io/psy-ops-guide/files-and-directories.html>).

Geef aan welke metadata toegepast worden aan de hand waarvan anderen de data kunnen identificeren en vinden.

No metadata will be added.

Geef aan op welke manier de data gedocumenteerd/beschreven wordt.

The data will be described in a publication about the project.

Toegang tot data en het delen van data

Gedurende het project, voor het publiceren van de resultaten, met wie wordt de data gedeeld?

- Mede-onderzoekers binnen het project
- Mijn consortium partners
- Onderzoekers in hetzelfde vakgebied
- Mijn promotor(en)/begeleider(s)
- Anders, licht toe

Onze anonymized, the data will be made public. We will not wait until the project is finished.

Zijn er redenen om het delen van de data te beperken?

- Anders, namelijk ...

The anonymized data will be made public. The raw data will be kept for two years and then destroyed.

Bevat het informed consent informatie over het delen, bewaren of verwijderen van data?

- Ja, licht toe

The informed consent is based on the Open Consent Insert (<https://bit.ly/open-consent>) and designed to publish the data. In addition, if possible, we will provide for the possibility for participants to make their data manifestly public.

Data archiveren en publicaties

Zijn er eisen aan het openbaar maken van (een deel van) de data na afloop of tijdens het project?

- Ik zal (delen van) mijn data open access delen (zonder restricties)

All anonymized transcripts will be shared as soon as possible.

Welke data worden beschikbaar gesteld voor hergebruik?

- Alle data uit het project worden vrij beschikbaar gesteld
- Overig (licht toe)

(although only the anonymized transcripts)

Wanneer komen de data beschikbaar voor hergebruik en hoe lang zullen ze beschikbaar zijn?

- Data beschikbaar na verschijnen publicatie

We aim to approximate Born Open Data.

Hoelang wordt de data bewaard?

- Alle data uit het project worden minimaal 10 jaar bewaard

The anonymized data will be kept indefinitely.

In welke repository worden de data gearhiveerd en beschikbaar gesteld voor hergebruik, en onder welke licentie?

Open Science Framework; ODbL license (<https://opendatacommons.org/licenses/odbl/>)

Beschrijf uw strategie voor het publiceren van de analysesoftware die in dit project wordt ontwikkeld.

We will publish all software in public Git repositories under GPLv3 or similar licenses.

Wie is verantwoordelijk voor de data na afronding van het project?

The researchers will remain responsible, as well as the Center for Open Science.

Datamanagementkosten

Welke middelen (bijvoorbeeld financiële middelen en tijd) worden ingezet voor datamanagement en om te zorgen dat de data vindbaar, toegankelijk, interoperabel en herbruikbaar zijn, oftewel: FAIR (findable, accessible, interoperable en reusable)?

We will use our own time to make the data available as FAIR as possible given our means and the deplorable state of possibilities for interoperability in the social sciences.

Planned Research Outputs

Interactive resource - "Transcripts"

The interview transcripts

Planned research output details

Title	DOI	Type	Release date	Access level	Repository(ies)	File size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Transcripts		Interactive resource	Unspecified	Open	Open Science Framework		Creative Commons Attribution 4.0 International	None specified	No	No