Reclaiming the promise of mobile sensing for emotion research

A Data Management Plan created using DMPonline.be

Creator: Koen Niemeijer

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / URL: PDMT2/23/021

ID: 203300

Start date: 23-10-2023

End date: 22-10-2024

Project abstract:

Understanding the intricate dynamics of emotions in daily life is crucial, both from a scientific and clinical perspective. Yet, obtaining a good picture of these emotion dynamics is challenging due to their contextual and person-dependent nature, and because the standard approach of repeated selfreport (so-called Experience Sampling Methods) places a significant burden on participants. Mobile sensing, utilising data from smartphones to analyse behavioural and contextual patterns, promises to revolutionise mental health research and clinical practice by offering continuous and objective data. Despite this optimism, ours and others' research indicates that mobile sensing's direct predictive power for momentary emotions is limited. Our project proposes two innovative approaches to reclaim the promise of mobile sensing for emotion research. The first approach investigates whether mobile sensing's predictive performance varies based on individual differences, especially different level of neuroticism and depression. We hypothesise that individuals with higher emotional variability, such as those with elevated neuroticism or clinical depression, might exhibit more distinct behavioural patterns, thus enhancing prediction accuracy. The second approach seeks to integrate mobile sensing with ESM by using sensing data to trigger ESM prompts at key moments of interest and avoid prompting at inopportune moments, reducing the burden on participants while improving the data quality.

Last modified: 19-01-2024

Reclaiming the promise of mobile sensing for emotion research

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Volume
WP1+3 mobile sensing data	Mobile sensing data collected in December 2021, e.g. physical activity, location, and environmental context	Existing	Digital	Observational	.json	700GB
WP1+3 ESM data	Self-report questionnaires capturing emotional states at multiple time points daily	Existing	Digital	Observational	.csv	5MB
WP1+3 personality questionnaires	Questionnaires assessing personality traits and risk of emotion disorders	Existing	Digital	Observational	.csv	1MB
WP1+3 informed consent	Informed consents participants	Existing	Digital	NA	.csv	104 ICs
WP2 mobile sensing data	Mobile sensing data, e.g. physical activity, location, and environmental context	New	Digital	Observational	.json	50GB
WP2 ESM data	Self-report questionnaires capturing emotional states at multiple time points daily	New	Digital	Observational	csv	5MB
WP2 personality questionnaires	Questionnaires assessing personality traits and risk of emotion disorders	New	Digital	Observational	csv	1MB
WP2 informed consents	Informed consents participants	New	Digital	NA	.csv	105 ICs
WP4 mobile sensing data	Mobile sensing data, e.g. physical activity, location, and environmental context	New	Digital	Observational	.json	50GB
WP4 ESM data	Self-report questionnaires capturing emotional states at multiple time points daily	New	Digital	Observational	.csv	5MB
WP4 personality questionnaires	Questionnaires assessing personality traits and risk of emotion disorders	New	Digital	Observational	csv	1MB
WP4 informed consents	Informed consents participants	New	Digital	NA	.csv	100 ICs
R code	Code for data processing, analysis, and model development.	New and existing	Digital	Software	.R	5MB

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:

For the datasets related to WP1+3, we are referencing information that is detailed in the publication available at https://doi.org/10.2196/43296. The data described within this source are for reference only, as the actual datasets are not available for public sharing.

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.

- Yes, human subject data (Provide SMEC or EC approval number below)

Datasets pertaining to WP1+3: G-2020-2200-R3(AMD)
Datasets pertaining to WP2: Approval from EC pending
Datasets pertaining toWP4: Approval from SMEC pending

Will you process personal data? 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).

- Yes (Provide PRET G-number or EC S-number below)

Datasets pertaining to WP1+3: G-2020-2200-R3(AMD)
Datasets pertaining to WP2: Approval from EC pending
Datasets pertaining toWP4: Approval from SMEC pending

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.

- Yes

The datasets derived from participant data across all work packages are exclusively designated for academic and research purposes and are not earmarked for commercial exploitation. However, the technological advancements achieved, specifically the development of an R package, will be used for commercial valorisation through a spin-off initiative by m-Path. This R package represents the tangible output of our research and has potential applications in technology transfer, where the methodologies and tools developed can be adapted for broader use within the industry by m-Path.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

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 in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

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).

To ensure our data remains clear and usable, we store mobile sensing data in JSON files and a structured SQLite database, accompanied by detailed documentation on the data schema. R functions for data interaction are well-documented for user understanding. The study's data collection procedures are outlined in flowcharts for easy reference.
ESM data is documented in a Word document noting question timing and participant instructions. Personality questionnaires are directly linked to their original sources for proper context and usage.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Collected data will be accompanied by detailed manual files that describe the collection methods and the contents of each data field. The mobile sensing data will also be systematically organized within a structured SQLite database. Due to the sensitive nature of mobile sensing data, which includes GPS locations, sharing of the raw data is not permissible.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- OneDrive (KU Leuven)
- Large Volume Storage

The data from this project will be stored on OneDrive for convenient access and on Large Volume Storage for long-term preservation. We will process the data exclusively on computers with encrypted hard drives. For the transfer of mobile sensing data from participants' phones, we will use pCloud, a secure GDPR-compliant cloud storage provider specialising in sensitive data, based in Switzerland. Data will be promptly deleted from pCloud servers following download to ensure security.

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

To secure our data, we store it within KU Leuven's Large Volume Storage and OneDrive, both compliant with high and medium confidentiality standards, respectively. KU Leuven's ICT solutions adhere to stringent university-wide information security protocols. Access permissions for the raw network storage managed by the faculty's ICT service are strictly regulated, delegated, and audited by designated data managers who are trained for this role, regardless of their IT expertise.
For the code files, which include pseudonyms, access is tightly controlled by a dedicated data manager, with the Principal Investigator serving as an alternate overseer. Researchers involved in the project are granted access exclusively to pseudonymised data, ensuring that personal identifiers are not disclosed.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The anticipated costs for data storage and backup for the duration of the research project are estimated at €1044.20, based on the projected data volume of less than 1 terabyte over a span of 10 years. These expenses have been incorporated into the bench fee component of the funding proposals submitted to the FWO and PDM.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Large Volume Storage (longterm for large volumes)

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Given that the data is less than 1 TB in size, storing the data for 10 years would cost €1044.20. This cost is factored into the requested bench fee from FWO and PDM.

Data Sharing and Reuse

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.

- Yes, as embargoed data (temporary restriction)
- Yes, as restricted data (upon approval, or institutional access only)

Due to the sensitive nature of the mobile sensing data and the fact that certain components cannot be effectively pseudonymized, the collected mobile sensing data in this project will not be made publicly available for external sharing. This ensures the protection of participant privacy and data confidentiality.
The ESM and personality questionnaires data will be subject to an embargo period to secure the research team's time to fully analyse the data and publish the findings. This measure is to prevent premature dissemination of the data which could lead to potential intellectual property issues or misappropriation of the research concepts by third parties.

If access is restricted, please specify who will be able to access the data and under what conditions.

The mobile sensing data will not be made publicly available as these cannot be sufficiently pseudonymised. Participants' personal information (e.g., contact information, names, etc.) will never be shared.
Questionnaire data will be made available after a period of embargo. Codes will be uploaded to the Open Science Framework and/or Gitlab.
The R package and associated functions will be uploaded on CRAN and Gitlab under a GPL-3 license.

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.

- Yes, privacy aspects

Due to the inherent nature of this data, particularly the inclusion of GPS coordinates, there is a high risk of participant re-identification. The specificity and granularity of the data could potentially reveal sensitive personal information. As such, to maintain participant confidentiality and comply with data protection regulations, this subset of the data will not be shared publicly without a data transfer agreement.

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)
- Other (specify below)

The mobile sensing data will not be made available as these cannot be sufficiently pseudonymised. Participants' personal information (e.g., contact information, names, etc.) will never be shared.
Questionnaire data will be made available after a period of embargo. Codes will be uploaded to the Open Science Framework and/or Gitlab.
The R package and associated functions will be uploaded on CRAN and Gitlab under a GPL-3 license.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- GNU GPL-3.0 (code)
- CC-BY 4.0 (data)
- Data Transfer Agreement (restricted data)

Mobile Sensing Data: Access to this dataset is restricted due to its sensitive nature. Sharing of the data will only occur under a formal data transfer agreement that guarantees appropriate security and confidentiality measures. The agreement will be established with a receiving institution before any data transfer, ensuring the protection of participant privacy.
ESM and questionnaire data will be shared under CC-BY 4.0 after publication of the research results.
The R package mpathsenser is currently published under the GNU GPL-3.0 license. Future code resulting from this project will be made available under the same license.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

A permanent identifier is added to data upon deposit in a repository.

What are the expected costs for data sharing? How will these costs be covered?

Sharing data at the Open Science Framework, CRAN and Gitlab is free.

Responsibilities

Who will manage data documentation and metadata during the research project?

Koen Niemeijer and Peter Kuppens

Who will manage data storage and backup during the research project?

Koen Niemeijer, Peter Kuppens, and Kristof Meers

Who will manage data preservation and sharing?

Koen Niemeijer and Peter Kuppens

Who will update and implement this DMP?

Koen Niemeijer and Peter Kuppens