Stress Tracking in Real life with multimodal Algorithms using data from Commercial wearables in a Student and Staff population.

A Data Management Plan created using DMPonline.be

Creators: Bert Bonroy, Nele De Witte, n.n. n.n.

Affiliation: Thomas More Hogeschool

Template: Thomas More DMP (Flemish Standard DMP)

Principal Investigator: Nele De Witte, n.n. n.n., Bert Bonroy

Grant number / URL: C3/23/043

ID: 207099

Start date: 01-03-2024

End date: 28-02-2026

Project abstract:

Over one in three Flemish residents own a wearable for continuous monitoring of behaviour and physiology. However, few widely available devices use scientifically validated algorithms to detect stress. This discourages companies to integrate this technology in current stress and mental health applications. Since stress is a complex and individual experience affecting both the autonomic nervous system and subjective experience, we hypothesise that combining datasets containing multimodal daily life data (physiological and self-assessment) with machine learning and datamining techniques will provide new scientific evidence for monitoring stress using commercial wearables. About 45% of Flemish students experiences regular to constant study strain and 20% has a mental health problem with a serious impact on daily life, so we will include students of the KU Leuven association. Additionally, we will include staff members to increase diversity and potential generalisation of the findings. The proposed research will allow companies to bring current mental health applications to a higher level by including a novel yet evidence-based data stream for stress identification.

Last modified: 27-05-2024

Stress Tracking in Real life with multimodal Algorithms using data from Commercial wearables in a Student and Staff population.

Thomas More DMP

---

1. 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	**Description**	**New or reused**	**Digital or physical**	**Digital data Type**	**Digital data format**	**Digital data volume (MB/GB/TB)**	**Physical volume**
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	

Data from wearable devices / activity trackers	Wearables are already well established today, allowing us to comfortably wear sensors on the body (e.g. wrist, chest, head) for long periods of time and to collect continuous data inside and outside a lab setting. In terms of wearables, there are already a lot of affordable devices available today, ranging from simple activity trackers (e.g. Fitbit, Garmin, Polar, Withings, Apple, Moonbird) to more expensive and high-end physiological measurements (e.g. Empatica, Byteflies). Research shows that such wearables can already be successfully used to manage stress via biofeedback, an intervention method in which physiological signals are measured and audio-visually fed back. What data we collect will depend on the chosen wearable and what data it makes available. The following physiological parameters can be collected: heart rate, resting heart rate, heart rate variability, activities (sitting, walking, exercising, ...) sleep, and breathing.	Generate new data	Digital	Observational	RegreSQL database, JSON	<100GB	

Subjective User feedback	Questionnaire data consists of (1) some items that will be surveyed several times a week and (2) some longer questionnaires that will be taken less frequently. Questionnaires will be taken via smartphone, but will be stored in the platform. Examples of topics that will be covered in these questionnaires are mental well-being (including stress), contextual information (e.g. Relevant physical or environmental factors). Survey questions will consist of an intake questionnaire consisting of general demographics, a brief stress assessment which will be completed multiple times per week (the items of this assessment will depend on an ongoing systematic literature review on the assessment of stress), and a questionnaire which will be completed around two times per month consisting of validated stress and wellbeing scales (e.g., Perceived Stress Scale).	Generate new data	Digital	Observational	Database, JSON, CSV	<1GB	
Questionnaires on qualitative project data	Questionnaires to evaluate the use of the StressTRACS application	Generate new data	Digital	Observational			
Processed Stress data	The wearable data and the subjective user feedback will be processed by the, in the project developed, stress detection algortithm which results in stress data.	Generate new data	Digital	Compiled/aggregated data	Database, JSON, CSV	<1GB	
Algorithm	The algorithm developed in this project will be developed in MATLAB	Generate new data	Digital	Software	MATLAB	<1GB	

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.

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data and/or personal data

In this study, 3 data subsets will be collected. First, the quantitative data involves data collected via the wearables (subset 1) and via smartphone or an online platform Subjective User feedback (subset 2). These data are anonymous, since they cannot be linked to individual participants. To do so, a study email addresses will be created by the researchers and will be randomly distributed to participants. These email

addresses will then be used to create dummy wearable and platform accounts, all using the same ‘fake’ personal data, namely male gender, height of 170 cm and weight of 70 kg. In addition, the same email addresses and accompanying ID codes are used to collect research data in the smartphone app and platform data. Wearable and smartphone data are automatically uploaded onto the data collection platform (created by the Belgian company). In addition, as only anonymous data is collected on the platform, no personal data is shared with this company. A programmer from the platform's owner could view the data but cannot link it to an existing person.

The questionnaires and qualitative data subsets are collected via the online and GDPR compliant survey program Questionpro. The questionnaires and qualitative data that will be collected include socio-demographic information (e.g., age and gender), information related to feasibility and acceptability of the StressTRACS application. Data collected via Questionpro is only available to the StressTRACS project researchers and is password-protected through the accounts of the individual researchers.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

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

Economic valorisation relates to the algorithm that will be developed which will be build on the Subjective User feedback and Wearable / activety tracker data.

Do existing third party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/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

2. Documentation and Metadata

Describe the documentation that will be created for the data. 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).

For dataset "Data from wearable devices / activity trackers": The data will be stored in a structured database. The columns in the structured database will describe the data in detail (type of experiment, types of variables, measuring frequency, measuring units, type of sensor/questionnaire used, etc.) making them self-explanatory.

For dataset "Subjective User feedback": The data will be stored in a structured database. The columns in the structured database will describe the data in detail (type of experiment, types of variables, measuring frequency, measuring units, type of sensor/questionnaire used, etc.) making them self-explanatory.

For dataset "Questionnaires on qualitative project data": This data is collected via QuestionPro. The answers are linked to the questions of predefined validated questionnaires, allowing the data to be described unambiguously.

For dataset "Processed Stress data": The data will be stored in a structured database. The columns in the structured database will describe the data in detail allowing it to be linked to the various experiments
For dataset "Algorithm": The code in which the algorithm will be developed will be commented on as best as possible.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Data from wearable devices / activity trackers and Subjective User feedback combined with the output of the Processed Stress data (as result of the developed algorithm) will be stored structured in the online data collection and representation platfrom (developed for this project).
The data from the wearable devices / activity trackers are described by the manufacturer.

3. Data storage & back-up during the research project

Where will the data be stored during research? Explain in the comment section.

- In a cloud service offered by Thomas More (Microsoft 365: SharePoint)
- In an external cloud service

As part of the StressTRACS project there is a need for a data platform to collect, process and visualise the project data. The contractor will set up a platform that can collect, process, and display data from the subjects for the duration of the project.
The data come from commercially available activity trackers (wearables) on the one hand, and data from questionnaires that will be offered to users via the platform and accompanying smartphone application on the other. The data will be processed by the algorithms that the researchers of the StressTRACS project will develop, which will provide additional information to be displayed to the user.
The data from the wearables and from the subjective data filled in by the participants in the smartphone app will also be stored in a regresql database hosted on a kubernetes cluster managed and hosted by Thomas More Mobilab & Care.
The data platform provides the posibility to download all data. These downloads will then be stored on Thomas More Microsoft 365 SharePoint (StressTRACS).

How will the data be backed up?

Data will be backed up using cloud service offered by Thomas More (Microsoft 365: SharePoint/OneDrive/Teams). These cloud services take automatic backups.
The StressTRACS data platform to collect, process and visualise the project data includes regular back-ups of the data.

Is there currently sufficient storage and backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

Cloud service offered by Thomas More has 5 TB storage and backup capacity through the corporate Office 365 licence. This is sufficient for the data created during the project.
The StressTRACS data platform to collect, process and visualise the project data provides sufficient storage and backup capacity during the project as it is part of the subcontracting.

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

Access to the data stored in the cloud service offered by Thomas More (corporate Office365 storage) is arranged via the Azure Active Directory and Multi Factor Authentication.
Data stored in the data centres of Thomas More is physically secured via an access system. Digital access is governed by rights and roles.

The StressTRACS data platform to collect, process and visualise the project data will be developed by a subcontractor. A Data Sharing Agreement will be signed with the subcontractor to ensue that the data are securely stored and not accessed or modified by unauthorized persons.

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

No extra costs are to be expected since the cloud service offered by Thomas More will be used.
The development costs of the StressTRACS data platform are budgeted and covered in the project application (substantiated by a quotation).

4. Data preservation after the end of the research project

Which data will be retained for at least ten 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 is kept for 10 years

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

The data will be stored on SharePoint (with automatic back-up procedures) for at least 10 years, conform the Thomas More RDM policy.

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

No extra costs are to be expected since the cloud service offered by Thomas More will be used.
At the end of the project, a data dump is done from the StressTRACS data platform. This data will then stored on the Thomas More infrastructure.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, upon request by mail
- Yes, in a restricted access repository (after approval, institutional access only, …)

If you answered 'Yes, in a restricted access repository' or 'Yes, upon request by mail', please specify who will be able to access the data and under what conditions. Explain the procedures to control access.

The anonymised data will be available to other researchers on request after signing a data sharing agreement. This data sharing agreement will be made available on the project website. A link to this procedure will be added on all project publications. The project leader will review all data requests. Access will be considered based on the planned reuse. Only uses for research purposes will be allowed and commercial reuse will be excluded

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a third party, legal restrictions)? Please explain per dataset or data type where appropriate.

- Yes, Privacy aspects

The study protocol could potentially lead to the collection of privacy-sensitive information (e.g., in the subjective user feedback). Any such data will be handled with care and removed before data can be shared with other parties (under the conditions mentioned above).

If the data (or part of the data) will be made available for reuse after/during the project, explain where the data will be made available. If already known, provide a repository per dataset or data type.

Data will be available on request after signing a data sharing agreement. The procedure for requesting access to data is available on the project website

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.

Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY-NC), so that users have to give credit to the original data creators

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

We commit to making data available (in part) in a repository to increase visibility.

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

There are no expected costs for data sharing.

6. Responsibilities

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

Project contributors will manage data documentation and metadata during the research project. Project leader bears the end responsibility.

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

Project contributors will manage data storage and backup during the research project. Project leader bears the end responsibility.

Who will manage data preservation and sharing?

Project contributors will manage data preservation and sharing during the research project. Project leader bears the end responsibility.

Who will update and implement this DMP?

The project leader bears the end responsibility of updating and implementing this DMP. The DMP will be updated during and at the end of the project.

Which additional resources are needed for the execution of the Data Management Plan?

NA

Did you read the Research Data Management Vision text of Thomas More?

- Yes

7. Internal project registration

Please provide the internal Thomas More project number.

K-OMO-230023

Stress Tracking in Real life with multimodal Algorithms using data from Commercial wearables in a Student and Staff population.

GDPR

GDPR

Have you registered personal data processing activities for this project?

- No