

Data Management Plan (DMP)

#TERRA: Paths to Sustainability in Adolescence
(COMPETE2030-FEDER-00903800)

Authors

Tiago Ferreira

[ORCID: 0000-0002-2884-2547](#)

Center for Psychology at the University of Porto (CPUP)
Faculty of Psychology and Education Sciences, University of Porto

Filipa Nunes

[ORCID: 0000-0002-1315-3327](#)

Center for Psychology at the University of Porto (CPUP)
Faculty of Psychology and Education Sciences, University of Porto

February 26, 2026

Summary. This Data Management Plan outlines the procedures for the collection, processing, documentation, storage, protection, sharing, and long-term preservation of data generated within the #TERRA project. The project produces quantitative survey and longitudinal data on adolescents' pro-environmental attitudes and behaviors, complemented by contextual information from parents and school heads. All data management procedures comply with FAIR principles (Findable, Accessible, Interoperable, Reusable) and applicable data protection regulations. Personal data are stored separately from research datasets and protected through encryption and restricted access protocols. Processed and fully anonymized datasets will be deposited in a public repository with a persistent identifier (DOI) and licensed under CC BY-SA 4.0 to ensure transparency, reproducibility, and responsible data reuse.

1 Information About the Research Project

1.1 #TERRA.: Paths to Sustainability in Adolescence

The #TERRA. project aims to investigate Pro-Environmental (PE) attitudes and behaviors during adolescence, seeking to identify their individual and contextual antecedents. This project combines a cross-sectional analysis of adolescents' PE attitudes and behaviors, based on a representative sample of the Portuguese population, with a longitudinal analysis focused on individual changes associated with age and daily fluctuations in ecological behaviors. This project aims to:

- (1) develop, adapt, and validate assessment instruments that reliably and validly measure PE attitudes and behaviors in adolescence, encompassing both the perceptual and general dimensions of actions, as well as their frequency and nature within ecologically relevant domains such as sustainable consumption, waste management, and energy efficiency;
- (2) characterize and identify the main individual and contextual factors that shape the PE attitudes and behaviors of Portuguese adolescents;
- (3) examine the developmental trajectory of PE attitudes and behaviors throughout adolescence, identifying predictive and risk factors associated with inter-individual changes;
- (4) analyze daily and weekly fluctuations in adolescents' ecologically relevant behaviors, investigating the role of intra-individual predictors (e.g., mood states and life satisfaction) and immediate contextual factors (e.g., daily social and environmental context) that shape variations in these behaviors.

Taken together, research results will enrich research in basic psychology, through the deeper understanding the complex interaction between individual and contextual predictors of PE attitudes and behaviors, providing critical evidence to promote environmentally sustainable behaviors among young people. The #TERRA project will also have a significant social impact, with direct implications across multiple intervention domains. In the educational domain, it will support the design of educational programs aimed at promoting sustainable behaviors, contributing to more effective pedagogical practices in fostering environmentally responsible citizens. In the domain of public policy, the findings will provide solid scientific guidance for national climate change mitigation strategies, enabling more informed decisions aligned with the United Nations Sustainable Development Goals (SDGs). Furthermore, the #TERRA project will impact social awareness by offering evidence to underpin campaigns that strengthen active environmental citizenship and encourage the adoption of more responsible consumption habits.

1.2 Research Team

- Tiago Ferreira (PI), tiagoferreira@fpce.up.pt
- Filipa Nunes (contracted researcher), fnunes@fpce.up.pt
- Hélder Spínola, hspinola@staff.uma.pt
- Marisa Matias, marisa@fpce.up.pt
- Paula Mena Matos, pmmatos@fpce.up.pt
- Richard Inman, richardinman@fpce.up.pt

1.3 Information about Funding Organization

This project was funded by COMPETE 2030 and Portugal 2030 programs and co-financed by the European Union (COMPETE2030-FEDER-00903800).

2 Information about Data

2.1 Reuse of existing data

Are you going to reuse data in this project?

☐ Yes ☒ No

2.2 Generation of new data

Are new data going to be generated in this project?

☒ Yes ☐ No

2.3 Type of data the research project will generate

The project's participants include adolescents (high school students), their parents or legal tutors, and school heads. The project encompasses five studies: a literature review (Study 1), a pilot study (Study 2), a national representative cross-sectional study (Study 3), a longitudinal study (Study 4), and a daily diary study (Study 5). Study 1 will generate data derived from bibliographic records, study characteristics (e.g., population, context, measurement instruments), and extract quantitative indicators from the reviewed literature. The remaining studies will generate quantitative data. Study 2 will be conducted with a convenience sample of 150 adolescents. The national representative study will involve 600 adolescents. The longitudinal study will involve 500 adolescents, their parents ($n \approx 1000$), and school heads ($n \approx 25$). Finally, the diary study will comprise 280 adolescents, randomly selected among the participants from longitudinal study. Data for study 3 will be collected online and via phone interviews using self-report measures. This data collection will be carried out by an external company specialized in survey studies. The remaining data will be collected by the research team through self-reports from adolescents, as well as reports from parents and school heads about variables related to specific adolescents, their families, and their schools.

2.4 Data format

We will adopt standardized formats that facilitate data exchange between operating systems and software applications, namely *.xls/.xlsx and *.csv.

2.5 Data volume

The project is expected to require a moderate amount of data storage, estimated between 10 and 100 gigabytes in total. This range reflects the anticipated volume of data necessary to support all activities and deliverables throughout the project lifecycle.

3 Documentation and Metadata

3.1 Documentation

In alignment with the Data Documentation Initiative (DDI) Alliance standards, the project will provide comprehensive and structured documentation to ensure transparency, interoperability, and long-term usability of the research data. This documentation will serve as a critical resource for both internal team members and external researchers who may wish to access or reuse the data. The documentation package will include:

- Methodology Details: A clear description of the research design, data collection procedures, sampling strategies, and any instruments or tools used;

- Variable Definitions and Data Dictionaries: Detailed explanations of all variables, including naming conventions, units of measurement, coding schemes, and any derived variables;
- Metadata Files: Structured metadata following DDI standards to describe the dataset's context, provenance, and relationships between files.
- Readme Files: Human-readable summaries providing quick guidance on file structure, formats, and instructions for accessing and interpreting the data.
- Codebooks: Comprehensive listings of variables, categories, and coding rules, including any transformations applied during data cleaning or analysis.
- Version Information: Documentation of dataset versions, including changes made over time, to maintain reproducibility and traceability.

This approach ensures that the data will be easily discoverable, understandable, and reusable, supporting compliance with FAIR principles (Findable, Accessible, Interoperable, Reusable) and promoting open science practices.

3.2 Metadata

Are there metadata associated with the data?

☒ Yes ☐ No

3.2.1 Which metadata will be available

In adherence to the best practices and standards, metadata will include the following elements:

- (1) Title and description;
- (2) Creator and contributors;
- (3) Date of creation and update;
- (4) Keywords and tags;
- (5) License information;
- (6) Data source and temporal scope;
- (7) Methodology and data collection details;
- (8) File format and size;
- (9) Version information;
- (10) Related publications;
- (11) Funding information;
- (12) Contact information.

The research data will be accompanied by comprehensive metadata to ensure clarity, discoverability, and compliance with best practices. These metadata will include: (1) Title and description; (2) Creator and contributors; (3) Date of creation and update; (4) Keywords and tags; (5) License information; (6) Data source and temporal scope; (7) Methodology and data collection details; (8) File format and size; (9) Version information; (10) Related publications; (11) Funding information; (12) Contact information.

3.2.2 Metadata standard(s)

To maintain interoperability and adherence to international standards, the project will adopt the Data Documentation Initiative (DDI) framework for metadata description. This structured approach ensures that the dataset is fully aligned with FAIR principles and ready for effective reuse by other researchers.

3.2.3 Metadata vocabular(s).

We will employ controlled vocabularies to guarantee consistency and semantic clarity across metadata elements. In this project, we will rely on vocabularies embedded within the DDI standard, complemented by widely recognized schemas such as Dublin Core for general descriptive elements and ISO 639 for language codes. These vocabularies will ensure that metadata terms are precise, machine-readable, and compatible with international data-sharing platforms. Specifically, the controlled vocabularies will include standardized terms for key metadata components such as “ContributorRole” (to define roles like author, editor, or data collector), “TopicClassification” (to categorize thematic areas), “ModeOfCollection” (to describe data collection methods), “TypeOfInstrument” (to specify instruments used, such as questionnaires), and “TypeOfTranslationMethod” (to indicate translation approaches). Additional vocabularies will cover “DataSourceType”, “DataType”, and “NumericType”, ensuring clarity in data structure and measurement. We will also use terms like “ResponseUnit”, “Variables-Relations”, and “AnalysisUnit” to describe the unit of analysis and relationships between variables, as well as “SummaryStatisticType” and “TypeOfFrequency” for statistical and temporal attributes.

By adopting these controlled vocabularies, the project will achieve semantic interoperability, improve data discoverability, and support FAIR principles, making the dataset ready for long-term preservation and effective reuse.

3.2.4 Metadata language code

The metadata will be provided primarily in English to ensure broad accessibility and interoperability with international data-sharing platforms. To standardize language identification, we will adopt the ISO 639-2 language code system, which uses three-letter codes for languages. For English, the code is `eng`. This code will be applied consistently across metadata fields such as language, title, and description to maintain clarity and machine-readability. For example:

```
<language code="eng">English</language>
<title xml:lang="eng">Pro-Environmental behavior questionnaire </title>
<description xml:lang="eng">This dataset contains information on...</description>
```

Using ISO 639-2 ensures that metadata is structured in a way that supports semantic interoperability, aligns with FAIR principles, and facilitates integration with global repositories. Although the primary language will be English, the metadata structure allows for future multilingual support if needed. For example, Portuguese could be added using the ISO 639-2 code `por`:

```
<title xml:lang="por">Questionário sobre comportamentos Pró-Ambientais</title>
```

This flexibility ensures that the dataset can be adapted for broader audiences without compromising standardization.

4 Storage and Security of the Data and Metadata

4.1 Data and Metadata Storage and Backup

During the project, all data and metadata will be securely stored on the institution's network to ensure reliability and controlled access. In addition, the project's raw data will be maintained on the Center for Psychology at the University of Porto (CPUP) Network Attached Storage (NAS), providing a robust and centralized storage solution. To further safeguard against data loss, a backup copy will be kept on an encrypted external disk, accessible only to the Principal Investigator (PI) and the contracted researcher. This multi-layered approach ensures both security and redundancy, protecting the integrity and confidentiality of the research data throughout the project lifecycle.

4.2 Backup Policy for Project Data

The project follows a strict backup policy to ensure data integrity and security. All raw data will be securely stored on the University's server and the Center for Psychology at the University of Porto (CPUP) Network Attached Storage (NAS). Both storage systems are protected with encryption to prevent unauthorized access.

To minimize the risk of data loss, backups will be performed regularly every three months, creating redundant copies that safeguard against hardware failure or accidental deletion. This approach ensures that the research data remains secure, recoverable, and compliant with institutional standards throughout the project lifecycle.

4.3 Protection and Security of Sensitive Data

To ensure the highest level of data security and protection for sensitive information, the project will implement both the default security measures of the institution's networked research storage and additional safeguards tailored to the nature of the data collected.

Personal information from participants will be stored in a separate, independent dataset on the University's server and the CPUP Network Attached Storage (NAS), both protected with encryption to prevent unauthorized access. Each participant will be assigned a unique alphanumeric identifier, which will be used to link the anonymized datasets containing the variables of interest for the study, ensuring confidentiality and compliance with data protection regulations.

The project's raw data will also be securely stored on the University's server and CPUP's NAS with encryption. To further mitigate risks, data will be regularly backed up, and an additional encrypted external disk will hold a safety copy accessible only to the Principal Investigator (PI) and the contracted researcher.

The Data Management Team will retain raw data only until a stable version of the processed datasets is finalized. These processed datasets will be maintained on the University's servers for five years after the project's conclusion to support reproducibility and long-term access. All security procedures will be reviewed and approved by the University's Data Protection Unit, ensuring compliance with institutional and legal standards.

5 Personal Data, Intellectual Property Rights, and Ownership

5.1 Processing of personal information and intellectual property rights and ownership?

The project will involve the processing and storage of personal data, as participants' information is necessary for the research objectives. All personal data will be handled in strict compliance

with applicable data protection regulations and institutional policies.

5.2 How will compliance with legislation and regulation on personal data be ensured?

In Study 1, we will not be collecting any personal data. Study 2, 3, 4 and 5 will collect data from adolescents (high school students). Study 3 will be conducted by an external company. Data will be collected through self-report measures online and via phone interviews to ensure representativeness and avoid the reproduction of participation asymmetries. The remaining studies' data collection will be conducted by the research team also through self-report measures. Study 2 will collect data from adolescents via a convenience sample at a single time-point, while Study 4 will collect longitudinal data from adolescents, their parents, and school heads. Adolescents will report data at three time-points across 18 months, while data from parents and school heads will be collected at the baseline. In Study 5 (daily diary), we will collect data from adolescents three times a day for two weeks. The data collections (adolescents, parents, and school heads) will be conducted either through an online survey in the Qualtrics, a software fully compliant with the GDPR requirements, or through a paper and pencil questionnaire, according to the participants' preference. Personal information (i.e., name, contact, and school name of the participants) will be requested from the participants only in the informed consent document, at the beginning of the project. The signed informed consents with personal information will be stored in a closed cabinet in one of the CPUP rooms. Only the PI and contracted researcher will have access to this room. A dataset containing this personal information will be created and stored at the University's server and Center's for Psychology at University of Porto (CPUP) Network Attachment Storage (NAS), protected with encryption against unauthorized access with a password only accessible to the two main researchers. A unique alphanumeric identifier will be randomly assigned to each participant at the beginning of data collection and used to link data from different informants and assessments. Participants' privacy will be granted by data deidentification, storing the personally identifiable data and the project's interest data in different datasets linked by the unique participant's code. All the generated data will be uploaded by the research team to the University's server immediately and CPUP's NAS after data collection. The Data Management Team will compile data from different sources and time points in central datasets and proceed with data organization cleaning and quality assurance. Any physical material used for collecting data (e.g., paper and pencil questionnaires) will not contain participants' personal information – they will be identified using the participants' code and discarded as soon as possible after it has been processed.

5.3 How will ownership of the data and intellectual property rights to the data be managed?

The anonymized datasets will be at a public data repository under the Creative Commons ShareAlike (CC BY-SA 4.0) international license. This licensing protocol will allow others to use, share, and adapt the processed data for any purpose if they give appropriate credit and share the resulting work under the same license (i.e., CC BY-SA 4.0). Participants will be asked to give their consent to the open access availability of the data.

6 Sharing and Preservation of the Data for the Long Term

6.1 Preservation of data for the long term

All data resulting from the project will be preserved for a minimum of five years following the project's conclusion to ensure compliance with institutional and funding requirements, as well as to support reproducibility and future research. However, specific preservation rules will apply to

different data types. The dataset containing personal information will be retained only until the final data collection moment, ensuring that participant identities can be linked to their unique alphanumeric identifiers if necessary. After this point, the personal data will be securely deleted to uphold privacy and data protection standards. The raw data will be preserved only until a stable version of the processed datasets is finalized. Once processed datasets are validated and complete, raw data will be securely removed to minimize storage risks and maintain data integrity. This approach balances the need for long-term preservation with ethical and legal obligations regarding personal data protection.

6.2 Availability of data for reuse

6.2.1 What data will be made available for reuse?

The project will not make all raw data publicly available. During the research phase, raw data will be securely stored on the University's server and the CPUP Network Attached Storage (NAS), with access restricted exclusively to the Data Management Team. To ensure privacy and minimize risks, raw data will be discarded as soon as possible after it has been processed. By the end of the project, the processed and fully anonymized datasets will be made openly available without any restrictions in a public access data repository. These datasets will be assigned a persistent identifier (DOI) to guarantee proper citation and long-term accessibility.

6.2.2 Availability of Data for Reuse

The processed and fully anonymized datasets will be made available for reuse upon completion of the project. This timing ensures that all data have been validated, documented, and prepared according to FAIR principles before being shared. Once published in the designated repository, the data will remain accessible for a minimum of five years after the project's conclusion, supporting transparency, reproducibility, and future research initiatives.

6.3 Repository for Data Archiving and Persistent Identification

The processed and anonymized datasets will be archived and made publicly available in the Open Science Framework (OSF), a trusted platform widely used for open research practices. OSF is registered in the re3data directory, ensuring visibility and compliance with recognized standards for research data repositories. To guarantee long-term accessibility and proper citation, OSF supports a system of persistent identifiers, assigning a Digital Object Identifier (DOI) to each publicly shared project. This feature ensures that the datasets can be reliably referenced and integrated into scholarly communication networks.

6.4 License for Data Reuse

All processed and anonymized datasets will be archived in a public data repository under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license. This license allows others to share, adapt, and build upon the data for any purpose, including commercial use, provided that appropriate credit is given and any derivative works are distributed under the same license terms. By adopting CC BY-SA 4.0, the project promotes open science, encourages collaboration, and ensures that data remains freely accessible while maintaining attribution and license continuity.

6.5 Strategy for Publishing Analysis Software

To ensure transparency, reproducibility, and adherence to open science principles, all code developed for data analysis during the project will be made open access. The code will be published in a publicly accessible repository, accompanied by clear documentation to facilitate reuse by

other researchers. Additionally, version control information will be provided to track updates and changes throughout the project lifecycle. This approach guarantees that the analytical processes are fully transparent, enabling verification of results and fostering collaborative improvements.

7 Responsibilities and Resources

7.1 Responsibilities

The project's Data Management Team will consist of three members of the management team: Tiago Ferreira (Principal Investigator), Filipa Nunes (team member and contracted researcher), and Marisa Matias (team member). This team will oversee the development and continuous updating of the Data Management Plan (DMP) and manage the entire data curation process, including data collection, organization, quality assurance, documentation, preservation, and sharing. Data collection will be carried out by the research team and trained research assistants, who will receive specific assessment training under the direct supervision of the project's coordination team. Data analysis will be conducted by the research team. All members of the Data Management Team are affiliated with the Center for Psychology at the University of Porto (CPUP).

7.2 Resources

The Data Management Team will ensure that all processes adhere to FAIR principles throughout the project lifecycle. Data curation will be integrated into the overall research management workflow. All datasets will be stored in CSV format, an interoperable standard that facilitates reuse by other researchers. Comprehensive documentation, including codebooks, will be created during the research process to enhance accessibility and reusability.

Before publication, an independent researcher (not part of the Data Management Team) will review the datasets and codebooks to confirm that they contain sufficient information for understanding and that the analysis results are reproducible and verifiable, ensuring interoperability and usability by external research teams. The final datasets and codebooks will be published in Open Science Framework (OSF) under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license, making the data findable, accessible, and reusable. Each dataset will be assigned a DOI to enable proper citation. The Principal Investigator (Tiago Ferreira) will hold ultimate responsibility for data sharing.

To guarantee data integrity and security, the project will use institutional servers and CPUP's Network Attached Storage (NAS) for primary storage, both protected by AES-256 encryption and secure authentication protocols. Backup copies will be maintained on encrypted external drives accessible only to authorized personnel. For version control of analysis code and documentation, the project will employ Git-based platforms (e.g., GitHub or GitLab), ensuring transparency, traceability, and collaborative development. These measures provide robust protection against unauthorized access and support reproducibility through controlled versioning.