

Document Title	Data Management Plan
Project Title and acronym	PINNACLE: <i>Public Engagement in AI Evaluation for Education and System Benchmarking</i>
Pillar	III. Transformation of RTDI System
Programme	BRIDGE2HORIZON
Grant Agreement	BRIDGE2HORIZON/0823E/0203
Deliverable No.	D10
Work package No.	WP2
Work package title	Dissemination and Exploitation Activities
Authors	Evgenia Christoforou
Contributors	Jahna Otterbacher, Emily Theophilou
Reviewers	Maria Kasinidou, Styliani Kleanthous
Status (D: draft; RD: revised draft; F: final)	F
File Name	D10_DataManagementPlan.docx
Date	30 March 2025

Draft Versions - History of Document				
Version	Date	Authors / contributors	e-mail address	Notes / changes
1.0	10/2/25	Evgenia Christoforou	evgenia.christoforou@ouc.ac.cy	Initial version
1.1	10/3/25	Evgenia Christoforou	evgenia.christoforou@ouc.ac.cy	Review version
1.2	30/3/25	Evgenia Christoforou	evgenia.christoforou@ouc.ac.cy	Final version

Abstract	
This deliverable provides the Data Management Plan for the PINNACLE Project.	
Keyword(s):	Data Management Plan (DMP), Protection of Personal Data, Sharing of Datasets

Contents

Contents	4
1. About this Deliverable	5
2. Data Management	5
3. Data Summary	6
4. Making Data FAIR	10
4.1. Data findability, including provisions for metadata	10
4.2. Data accessibility	10
4.3. Data interoperability	11
4.4. Data re-use	11
5. Allocation of Resources	11
5.1. Estimate of costs	11
5.2. Data management responsibilities	11
6. Data Security	12
7. Ethical aspects	12

1. About this Deliverable

This document provides the Data Management Plan (DMP) plan for the PINNACLE project. The deliverable provides an overview of how research data that are generated or collected during and after the project will be managed. This deliverable also identifies standards and how data will be shared.

2. Data Management

Data Management Plans (DMPs) are a key element of good research data management. A DMP is a living document that describes how research data will be collected, generated, processed, documented, stored, shared and preserved during a project, as well as how the data will be managed and made available, where appropriate, after the project has ended.

Aspiring to become increasingly involved in pursuing funding at the European level, in the spirit of the BRIDGE2HORIZON program, the PINNACLE team embraces the practices advocated by the European Commission, and which are required within Horizon Europe.

Under **Horizon Europe**, research data management is mandatory for projects that generate or reuse data. The European Commission requires beneficiaries to manage digital research data responsibly in accordance with the **FAIR principles** and to establish and regularly update a Data Management Plan throughout the lifetime of the project. The DMP should address the types of data generated or reused, their documentation and quality, storage and preservation, access conditions, and measures to make the data **Findable, Accessible, Interoperable and Reusable (FAIR)**.

Good research data management also contributes to scientific transparency and reproducibility. Relevant practices include **open access** to scientific publications and, where possible, research data; depositing data and associated metadata in **trusted research data repositories**; assigning persistent identifiers and enabling appropriate **data citation**; documenting the data, methods and other research outputs needed to validate research findings; and maintaining an up-to-date **Data Management Plan**. Horizon Europe follows the principle that research data should be “**as open as possible, as closed as necessary**”, recognizing that legitimate interests, intellectual property rights, confidentiality, security, ethical requirements and data-protection obligations may restrict access to particular datasets.

The guidelines provided by Michener², together with the current European Commission guidance for **Horizon Europe**, have been followed in the preparation of this DMP. In particular, this DMP adopts a **full research data life-cycle approach**, covering data generation or collection, processing, documentation, storage, access, sharing, preservation and, where appropriate, long-term reuse.

The FAIR principles themselves originated from discussions among stakeholders from academia, industry, funding organisations and publishing, including the 2014 workshop *Jointly Designing a Data FAIRport* held in Leiden. They were formally published by Wilkinson et al. (2016) as the **FAIR Guiding Principles for Scientific Data Management and Stewardship**¹. FAIR describes research data and associated digital objects as **Findable, Accessible, Interoperable and Reusable**, with particular emphasis on rich metadata, persistent identification, appropriate access mechanisms, interoperability standards and sufficient documentation to support future reuse. These principles were incorporated into European Commission research data policy under Horizon 2020 and are now embedded more comprehensively in the open-science and research-data-management requirements of **Horizon Europe**.

3. Data Summary

This section describes the data that will be produced during the PINNACLE project, including the purpose of the data, the formats and origins, the anticipated size, and to whom the data might potentially be useful. According to the project plan, an initial dataset will be collected by recruiting crowdworkers through the **Prolific** platform. This dataset will contain participants' responses to the initial task/activity templates developed within the project. These templates will subsequently be refined and adapted for use with participants in the project's educational courses.

The purpose of this initial dataset is to evaluate the **clarity, comprehensibility, and structure** of the developed tasks and to assess whether the collected responses can be meaningfully transformed into variables suitable for **quantitative and qualitative analysis**. The findings from this initial evaluation will inform revisions to the task design, instructions, and data collection procedures. Further adjustments will be made iteratively based on evaluations conducted with both crowdworkers and learners participating in the project's courses.

¹Wilkinson, M. D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3:160018 doi:10.1038/sdata.2016.18

²Michener, William K.(2015) Ten simple rules for creating a good data management plan. *PLoS computational biology* 11.10: e1004525.

We will also collect two datasets corresponding to the two rounds of the **AI in Everyday Life** course offered to members of the general public. Data will be collected from participants **before, during, and after** each course. Prior to the start of the course, participants will complete an initial demographic questionnaire together with validated measures relating to **human–AI interaction, trust, and AI literacy**. During each week of the eight-week courses, additional data will be collected through participants' responses to the weekly learning activities and quizzes. At the end of each course, participants will complete post-course questionnaires assessing changes in **AI literacy, trust, and human–AI interaction**, as well as a final evaluation of the course. These data will enable analysis of participants' experiences, learning outcomes, and changes in relevant attitudes and competencies over the duration of the course.

During the **first round of the course**, data collection will follow a **bottom-up approach**, in which participants will be invited to share their own experiences, observations, and concerns regarding AI applications they have recently used in their everyday lives. This approach is intended to capture participants' perspectives and identify issues that emerge directly from their interactions with AI-enabled systems.

During the **second round of the course**, data collection will follow a more **top-down approach**. Participants will be provided with a structured set of instructions for using and auditing a specific AI application and will be asked to report their observations, experiences, and evaluations. This approach will enable a more systematic assessment of participants' interactions with AI applications and facilitate comparison across participants and tasks.

During the activities described above, participants may inadvertently disclose **personal data**. According to the European Commission, personal data comprise any information relating to an identified or identifiable living individual. Any datasets containing personal data will be stored on **secure, password-protected servers** and will be accessible only to authorized researchers at the host organization (the Open University of Cyprus - OUC). Before any dataset is shared beyond the authorized research team, it will be appropriately **anonymized**, and direct or indirect identifiers that could reasonably enable the identification of individual participants will be removed.

This Data Management Plan (DMP) describes the data collected or generated within the **PINNACLE** project. It does not provide detailed descriptions of third-party datasets that may be used during the project, although any such datasets will be handled in accordance with their applicable licences, terms of use, and relevant legal and ethical requirements.

Overall, the total volume of data generated within PINNACLE is expected to be in the order of **hundreds of megabytes**. This DMP provides the framework for determining and documenting the following aspects of data management and the associated procedures:

Dataset reference and name Unique identifier for the dataset and name.	Collected within WP The WP that generated the dataset	Responsibility Person (and partner) responsible for producing and managing the dataset.
Dataset description Descriptions of the data that will be generated or collected, the purpose of the dataset, whether it relates to scientific publication, origin, nature and scale (size of data).		
Standards and metadata Reference to existing suitable standards or outline on what metadata will be created.		
Data sharing Description of how the data will be shared, including access procedures, embargo periods, software necessary for sharing/accessing the data and indication of repository of where the data will be stored.		
Archiving and presentation (including storage and backup) Description of procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, approximate end volume, associated costs are and how will be covered.		

Next, we shall detail the nature of the datasets that are expected to be generated, within each Work Package (WP) of the project.

Dataset reference and name	Collected within WP	Responsibility
DS1, Crowdsourcing Task – Anonymised Dataset	WP4, Framework Design and User Evaluation Tasks	OUC, CyCAT team
Dataset description This dataset will contain the complete research data collected from crowdworkers recruited through the Prolific platform during the initial evaluation of the task templates developed within the project. The purpose of the dataset is to assess the clarity, comprehensibility, structure, and usability of the tasks before their adaptation for use in the <i>AI in Everyday Life</i> courses. Data may include participants’ responses to the task templates, completion times, demographic variables required for the research, and qualitative feedback concerning the tasks and instructions. The dataset will support both quantitative and qualitative analysis and will inform revisions to the task design and subsequent data collection procedures. The expected data volume is relatively small, in the order of a few to tens of megabytes. The dataset will contain a curated and anonymized version of the crowdsourcing data. Direct identifiers will be removed and indirect identifiers will be removed, generalized, aggregated, or otherwise transformed where necessary to reduce the risk of re-identification. The dataset may include anonymized task responses, selected demographic variables, completion-related information, and qualitative feedback where these can be safely shared. The purpose of the dataset is to support transparency, reproducibility, scientific validation, and potential reuse of the research findings.		
Standards and metadata Structured data will primarily be stored in CSV format.		
Data sharing Subject to informed consent, ethical approval, and verification that the risk of participant re-identification is sufficiently low, the anonymised dataset may be made available to the research community. Where relevant, it may accompany scientific publications arising from the project. The dataset is expected to be deposited in Zenodo or another suitable trusted research data repository.		
Archiving and presentation (including storage and backup) The final anonymized dataset and associated documentation will be preserved in Zenodo or another suitable trusted repository, where it can be assigned a persistent identifier and maintained for long-term access. A working copy will also be retained on secure OUC infrastructure during the project. The dataset is expected to be small and no significant additional preservation costs are anticipated.		

Dataset reference and name DS2, AI in Everyday Life Round 1 – Restricted Research Dataset	Collected within WP WP6, AI Evaluation in Education	Responsibility OUC, CyCAT team
Dataset description This dataset will contain the complete research data collected during the first round of the eight-week AI in Everyday Life course . Data will be collected before, during, and after the course. It may include demographic information, responses to validated or established questionnaires concerning AI literacy, trust, and human-computer interaction , weekly activity responses, qualitative reflections, and final course evaluation data. During this first round, the weekly activities will follow a bottom-up approach , whereby participants will describe their own experiences, observations, and concerns regarding AI applications they have recently used or encountered in everyday life. The dataset will be used to investigate participants' experiences, learning outcomes, attitudes, and changes over the duration of the course.		
Standards and metadata Structured data will primarily be stored in CSV and Excel format.		
Data sharing This dataset will remain restricted to authorized members of the OUC research team. Because participants may provide free-text descriptions of personal experiences, the dataset may contain personal or potentially identifying information. Raw qualitative responses will therefore not be shared publicly. The restricted dataset will serve as the source for the preparation of the anonymised dataset described in DS3.		
Archiving and presentation (including storage and backup) The dataset will be stored on secure, password-protected OUC infrastructure with access limited to authorized researchers and covered by institutional backup procedures. The data will be retained in accordance with the project's ethics approval, applicable data-protection requirements, and OUC institutional policies.		

Dataset reference and name DS3, AI in Everyday Life Round 1 – Anonymized Dataset	Collected within WP WP6, AI Evaluation in Education	Responsibility OUC, CyCAT team
--	---	--

Dataset description

This dataset will be derived from DS2 and will contain an anonymised and curated version of the data collected during the first round of the *AI in Everyday Life* course. It may include anonymised questionnaire responses, selected demographic variables, weekly activity responses, qualitative material that can be safely anonymised, and course evaluation data. Particular care will be taken when reviewing free-text responses, as participants may inadvertently disclose personal information or information relating to third parties. The dataset will support scientific transparency, reproducibility, secondary analysis, and validation of research findings.

Standards and metadata

Structured data will primarily be stored in CSV and Excel format.

Data sharing

Subject to participants' informed consent, ethical requirements, and successful anonymisation, this dataset may be made available to the research community and may accompany scientific publications arising from the project. The dataset is expected to be deposited in **Zenodo** or another suitable trusted repository. Qualitative responses that cannot be sufficiently anonymised will either be excluded, redacted, summarised, or provided only in aggregated or derived form.

Archiving and presentation (including storage and backup)

The final anonymized dataset and associated documentation will be preserved in Zenodo or another suitable trusted repository, where it can be assigned a persistent identifier and maintained for long-term access. A working copy will also be retained on secure OUC infrastructure during the project. The dataset is expected to be small and no significant additional preservation costs are anticipated.

Dataset reference and name	Collected within WP	Responsibility
DS4, AI in Everyday Life Round 2 – Restricted Research Dataset	WP6, AI Evaluation in Education	OUC, CyCAT team
Dataset description This dataset will contain the complete research data collected during the second round of the eight-week AI in Everyday Life course . As in the first round, data will be collected before, during, and after the course and may include demographic information, responses to questionnaires concerning AI literacy, trust, and human-computer interaction , weekly activity data, qualitative responses, and final course evaluation data. In the second round, data collection will follow a more top-down and structured approach . Participants will be provided with instructions on how to use and audit selected AI applications and will report their observations,		

experiences, assessments, and concerns using the project's activity templates. The dataset will support systematic analysis and comparison across participants, activities, and course activities.

Standards and metadata

Structured data will primarily be stored in CSV and Excel format.

Data sharing

This dataset will be restricted to authorised OUC researchers. Raw participant-level data and qualitative responses will not be shared publicly. Particular attention will be given to information participants may enter into, receive from, or report about third-party AI applications, including any personal, confidential, or potentially identifying information. DS4 will serve as the source dataset for the preparation of the anonymised dataset described in DS5.

Archiving and presentation (including storage and backup)

The dataset will be stored on secure, password-protected OUC infrastructure with access limited to authorized researchers and covered by institutional backup procedures. The data will be retained in accordance with the project's ethics approval, applicable data-protection requirements, and OUC institutional policies.

Dataset reference and name DS5, AI in Everyday Life Round 2 – Anonymized Dataset	Collected within WP WP6, AI Evaluation in Education	Responsibility OUC, CyCAT team
Dataset description This dataset will be derived from DS4 and will contain an anonymized and curated version of the data generated during the second round of the <i>AI in Everyday Life</i> course. It may include anonymised questionnaire responses, selected demographic variables, structured AI-auditing activity responses, qualitative observations that can be safely anonymized, and course evaluation data. The dataset will support scientific reproducibility, validation of project results, and secondary research where appropriate.		
Standards and metadata Structured data will primarily be stored in CSV and Excel format.		
Data sharing Subject to informed consent, ethical approval, third-party terms of use where applicable, and successful anonymisation, this dataset may be made available to the wider research community. It may accompany scientific publications and is expected to be deposited in Zenodo or another suitable trusted research data repository. Any personal information, confidential information, or data presenting a material risk of		

re-identification will be removed, transformed, aggregated, or excluded prior to sharing.

Archiving and presentation (including storage and backup)

The anonymized dataset and associated documentation will be deposited in Zenodo or another appropriate trusted repository for long-term preservation and reuse. A secure working copy will also be retained by OUC during the project. Given the expected size of the dataset, preservation costs are expected to be negligible and covered by existing institutional and repository infrastructure.

4. Making Data FAIR

4.1. Data findability, including provisions for metadata

This section outlines the discoverability and identifiability of the data and the use of persistent and unique identifiers.

Data documentation will accompany datasets submitted to the project's Zenodo (<https://zenodo.org/>) repository at the end of the respective research study. Records of datasets will be published in Zenodo, which will issue DOIs for registered datasets and promote discovery.

4.2. Data accessibility

This section specifies the extent of open access, how the data is made available, what methods and tools are used.

We plan to use Zenodo (<https://zenodo.org/>) as the general data repository for PINNACLE. Data will be made available through a shared research platform (Zenodo) with the relevant permissions in place. The project's team will have exclusive use of the data until the main research findings are published.

At present we do not foresee any delays in data sharing following publication of the main research findings.

We will continue to review appropriate data management tools during the project using the Registry of Research data Repositories (<http://www.re3data.org>) list.

4.3. Data interoperability

This section covers what data and metadata vocabularies, standards and methodologies are used to facilitate interoperability.

Our research data produced will be saved in csv format, a generally acceptable method that allows data to be interoperable.

4.4. Data re-use

This section specifies data licensing, availability and length of time for re-use.

Wherever possible, data will be shared as soon as possible using the Creative Commons (<https://creativecommons.org/licenses/>) International Licence with Attribution (CC BY). Experimental data will, in some cases, only become available after the end of the project or publication of the results, whichever comes first, and will be shared under the same CC BY licence. The CC BY licence allows maximum dissemination and re-use but allows others to distribute, remix, tweak and build on the work as long as the original work is credited.

5. Allocation of Resources

5.1. Estimate of costs

We do not expect any cost associated with the findability, accessibility, interoperability or re-use of the data due to the use of existing institutional and freely available data management software. At this moment, it is hard to estimate what would be the costs involved in the long-term preservation of the data.

5.2. Data management responsibilities

The Project Coordinator will lead the data management activities and will be supported by the CyCAT team. The team at OUC will maintain contact with the relevant local bodies (e.g., the Cyprus Open Data Portal) in order to make our research data, results and (non-confidential) documents available and open. In addition, these results will be

distributed via the project's Web and social media presences. Finally, the team at OUC will make sure that all PINNACLE publications have green open access.²

6. Data Security

Any personal information that might be gathered by the PINNACLE project will be strictly accessible by the team working on the specific work packages that will be responsible for storing and maintaining these data using OUC resources. For secure storage as well as long term preservation and curation of the non-sensitive data created, PINNACLE will rely on a certified repository (Zenodo).

7. Ethical aspects

Ethical issues associated with PINNACLE are detailed in two deliverables: the Operational and Risk Management Plan (D1) as well as the Ethical Review and Approval (D2). Human subject and other sensitive data will be handled according to the ethical guidelines described in D1 and D2 and managed by the Project Coordinator (OUC).

² Note, that 'green' access allows the author to deposit the published article on an institutional repository or personal web pages; however, often only after an embargo period.