EYE•TEACH

D6.1: Data Management Plan

Author(s):	EYE-TEACH Consortium
Editor(s):	Daria Pritup (University of Turku)
Responsible Organisation:	University of Turku
Version-Status:	V2
Submission date:	31/03/2026
Dissemination level:	PU - Public

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

TABLE OF HISTORY OF CHANGES		
Version	**Publication Date**	**Changes**
1.0	30.06.2025	**Data Management Plan** Initial, submitted version 1.
2.0	31.03.2026	**Data Management Plan** Version 2, changes: • Table 1: inclusion of WP4 meeting recordings and WP2/WP3 behavioural data (p.6-7) • Treatment of management and governance documentation collected in the project (p. 7) • 1.2 Data origins (p. 7) to include age group for child participants • 2.2 Making data accessible (p.9) to include public OSF project link (DOI) and data access to include multi-factor identification • Table 2 Data quality assurance processes (p. 11-12) to include bogus-item check and additional responsible partners • Digital outputs (p. 13) to include OSF • 4.1 Costs and responsibilities (p. 13-14) to include information about Joint Controllership Agreements and Data Processing Agreements • 5.1 Data handling security (p.15) to include external hard drive backup • Table 3 (p. 16-17) to include additional partners: CNR and CNK • Chapter 6 Ethics and legal compliance updated with a summary (p. 17-18) and sections 6.2 and 6.3 (p. 19-22) on research protocols and ethics self-assessment • Table 4 to include partner CNR and DPO contact information for all named partners and additional procedures

Table of Contents

Introduction

Eye-tracking and AI for Enhanced Teaching (EYE-TEACH) is an EU-funded initiative aimed at transforming educational practices across Europe. By integrating AI and eye-tracking technologies into teachers' day-to-day pedagogical practices, we empower educators and give them novel tools for enhancing their students' reading comprehension. This technology offers insights into reading behaviours and comprehension levels, enabling personalised teaching strategies.

This Data Management Plan informs the research of the project across the data lifecycle, from planning to collecting, analysing, sharing and storing to potential re-use. The document explains how research data will be handled by the project's research partners and sets out clear and responsible ways of managing data that follow open science standards. In line with the FAIR principles (Findable, Accessible, Interoperable, Reusable), the aim is to make data available to others while balancing openness with the requirements of GDPR and ethical safeguards.

This is a living document, updated by M15 and M36 of the project's lifecycle.

1. Data Summary

EYE-TEACH re-uses published and unpublished eye-movement datasets to identify potential predictor variables of reading comprehension. Datasets are identified via a systematic review of published research and inquiries from researchers in the field. The datasets to be re-used should include both eye-movement variables and reading comprehension measures for a participant measured within the same reading task.

EYE-TEACH also collects several types of data as outlined in Table 1. The size of the data depends on the type of file format, the number of participants in each substudy, and the length of the testing sessions (see approximations in Table 1).

1.1. Purpose of data generation and re-use

Survey and interview data are generated to map the needs of teachers and educational organisations regarding the use of AI-assisted ET-analytics tools, the factors impacting teachers' willingness and readiness to adopt such technologies, and the current use of data in guiding educational practices. We will also explore and identify eye movement metrics that are most suitable

for tracking reading comprehension processes in different task settings and educational contexts by re-using existing eye tracking datasets.

Moreover, survey and interview data are collected to co-create a pilot system with the teachers and education professionals. Validity tests of the eye movement metrics in the specific reading comprehension task will be done with eye tracking and comprehension data generated in laboratory experiments. The user interface of the pilot system will be tested in usability tests involving eye-tracking data, audio/video recordings, screen recordings, and survey and interview data. Feasibility tests will be carried out to examine the potential challenges and opportunities of this type of systems in educational use, producing survey or interview data. Validity of the eye movement metrics collected with the pilot tool will be tested by collecting eye-tracking and reading comprehension data with the pilot system.

Stakeholder contact information will be gathered for communication and dissemination activities.

Table 1. Data to be collected in the EYE-TEACH project.

Work Package(s)	*Type of data*	*File format*	*Software*	*File size*	*Sensitive / Confidential (Yes/No)*
WP1	Interviews (original audio)	.mp3	Soundrecoder, Scribewave (for transcribing)	~30Mb per participant (depends on the length of the interview)	Yes
WP4	Advisor meeting recordings (original video)	.mp4	Teams	~500Mb per meeting	Yes
WP1, WP4	Interviews (transcripted)	.txt	Word, Excel, NVivo	~25-50Kb	No
WP1, WP2, WP3	Eye movement recordings (original files)	.edf or other, .mp4	SR Research Dataviewer, SR Research Experiment Builder, Tobii Pro Lab, iMotions, BeGaze, Python, Experiment Center	~5Mb per participant (depends on the length of the recording)	Yes (e.g., webcam eye-tracker data, pupil recordings, video recordings, audio recordings) No (numeric data)
WP1, WP2, WP3	Eye movement data (preprocessed)	.csv	Tobii Pro Lab, iMotions, Excel, R, SR Research, Python, BeGaze, Blickshift Analytics	~100-500Mb per participant (depends on the length of the recording)	No
WP1	Survey data (original files)	.xlsx .sav .csv	Qualtrics	~1-5Mb	No
WP1	Survey data (preprocessed)	.xlsx .csv	Excel R, Blickshift Analytics	~1-5Mb	No
WP2, WP3	Behavioural data (comprehension	.xlsx .csv	Excel	Varies by task	Yes (e.g., audio recordings)

	data and cognitive/ language tasks data)	.md .html .mp4 .txt .docx	R, Python, Blickshift Analytics, HTML-Javascript, Word		No
WP3	User information (mother tongue, school level)	.xlsx	Excel, Blickshift Analytics	~1-5Mb	No
WP3	Text materials (metrics)	.xlsx	Excel	Varies by task	No
WP5, WP6	Contact information (mailing list of the EYE-TEACH ecosystem, events, board members)	.xlsx	Excel	~20-100Kb per file	Yes

The project also generates internal management and governance documentation, such as consortium meeting agendas, minutes, action lists, and decision logs. These materials are not treated as primary research datasets, but as confidential project records supporting coordination, traceability, and implementation. Non-confidential governance documentation may include for example written input from consortium partners.

1.2. Data origins

The data generated in the project originates from volunteer study participants recruited to the project: children (ages 10-13), adults, teachers, or education professionals. Re-used data originates from participants of previous studies.

1.3. Data utility

The data might be useful for researchers, educators, and technology professionals interested in developing novel tools for educational contexts.

2. FAIR data

The data generated within the project will be managed in line with the FAIR and open access principles. This means that the data will be as findable, accessible, interoperable and reusable as possible, and it will also be as open as possible, yet as closed as necessary to ensure data protection and anonymity concerns, especially for potential minor participants.

2.1. Making data findable

As a public dataset in the OSF repository (see section 2.2), the data will receive a persistent identifier through registration of a DOI with Datacite. Keywords will be provided in the metadata to optimize the possibility for discovery and re-use of the data.

OSF leverages key metadata to describe public scholarship and utilizes a metadata model that facilitates FAIRness (Findable, Accessible, Interoperable, Reusable) as well as enabling connections across the research lifecycle. The OSF Metadata Profile[1], which uses many common metadata standards, describes the community vocabularies and persistent identifiers that the OSF uses, the relationships available between metadata fields, the metatags used to enable enhanced web discovery, and an overall map of the metadata implementation. The metadata uses the Datacite metadata schema, including title, description, authors, license, subject, language, resource type, publication date, modification date.

OSF provides an integration with Stanford University's CEDAR embeddable editor, which allows for annotating research artifacts using specialized metadata templates. CEDAR makes it possible for community creators of specialized metadata templates to create machine-readable and FAIR schemas. The embeddable editor displays CEDAR schemas within the OSF. Researchers will select the relevant specialized template on OSF which best fits their domain of research and fill out the additional metadata form. On public projects, the contents of the extra form are displayed

[1] OSF Metadata Profile, accessed 22 May 2025: https://help.osf.io/article/573-osf-metadata-profile

alongside the OSF standard metadata. It is also possible to download the additional metadata as a JSON file.

2.2. Making data accessible

Repository

Anonymised data will be deposited in the Open Science Framework (OSF) repository at https://osf.io. The OSF is a free research collaboration and management platform that launched in 2012. The repository for the EYE-TEACH project can be found at https://doi.org/10.17605/OSF.IO/5PNHR. The EYE-TEACH OSF repository is divided into subprojects for each of the project's work packages and studies. Each dataset will be assigned an identifier through OSF. Each OSF object and file is identified internally through an OSF identifier — a combination of 5 letters and numbers unique within OSF. When combined with the root OSF URL, the identifier forms a GUID (globally unique identifier) for the object.

Data

The project research data will be anonymised and made openly available through OSF. Certain datasets may need to be shared under restricted access conditions or may not be shared at all if

a) the data contains sensitive information and/or cannot be anonymised, or
b) opening the data would be against a partner's legitimate interests, including regarding commercial exploitation.

For data that meets one or all conditions stated above, restricted access may be granted by the Data Controller to other partners via a data sharing agreement.

Access requests to personal/sensitive data will be evaluated and approved by the Data Controller, following consultation with the Data Protection Officer (DPO). Identity of the person accessing the personal or sensitive data is ascertained by personal passwords and multi-factor authentication; the data are stored on password-protected and multi-factor authenticated servers and the Data Controller will provide access only to approved users.

Pre-processed research data will be made available as soon as possible. The pre-processed data are freely accessible via the Open Science Framework project repository.

Data that cannot be anonymised cannot be made accessible even after the project has ended. It will be destroyed. Stakeholder contact information will not be opened.

Metadata

Metadata will be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement. The pre-processed data and metadata will be preserved and remain accessible indefinitely.

The data will be stored in formats that are readily readable with open-source software. Data analysis scripts will be made available with the datasets via OSF.

2.3. Making data interoperable

Standard variable names for eye tracking data will be used whenever possible. Each dataset will be accompanied with a codebook providing detailed descriptions of the variables included in the data to allow data exchange and re-use.

The data will include qualified references[2] to other data, such as

- links to persistent identifiers of other datasets produced in the project (via OSF),
- links to datasets used as input or output in a particular experiment,
- references to datasets used in related publications,
- citations and acknowledgements of other researchers' data or work.

[2] A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)

2.4. Increase data re-use

The documentation of the published pre-processed data will be stored and released via OSF. The documentation will contain read-me files describing the contents of the datasets, codebooks, and data analysis scripts; codebooks that describe the datasets (e.g., variable definitions and units of measurement); and data analysis scripts that describe the data cleaning and analysis procedures.

The published, pre-processed data will be made freely available for re-use by third parties under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence/dedication with equivalent rights, in line with the obligations set out in the Grant Agreement. The published data will be available on OSF also after the end of the project.

The provenance of the data will be thoroughly documented using appropriate standards and tools tailored to each data type. The EYE-TEACH project will ensure that data origin, context, and transformations are traceable, reproducible, and reusable by external parties where applicable.

2.4.1. Data quality assurance processes

Data are collected following a detailed research plan containing a protocol for data collection, including e.g. procedures for calibration of the equipment and data quality checks. All data will be processed according to the best practices established for each measurement. Please see Table 2 for more details.

Table 2. Data quality assurance processes.

Data type	*Origin / Collection method*	*Quality Assurance measures*	*Responsible partner(s)*
Teacher survey data	Multilingual surveys (WP1)	Professional translation and validation by partners; pilot testing; bogus items as attention checks; country-level consistency checks	UANTWERPEN, OUNL
Eye-tracking data – students	Eye movement recordings in lab and classroom (WP2)	Standards for accepted calibration results; automatized or controlled data cleaning procedures; use of validated indicators	UVEG, UTU

Eye-tracking data – teachers	Collected during design evaluation and dashboard testing with teachers (WP1)	Standards for accepted calibration results; automatized or controlled data cleaning procedures; use of validated indicators	OUNL, UANTWERPEN
Eye-tracking EM metrics data	Curated from internal and external EM datasets (WP2, WP3)	Standard technical quality check with available information (e.g., corrupted files/misalignment errors, data completeness, calibration accuracy check, sampling rate consistency, metadata/documentation verification)	DFKI, UVEG, UTU
AI model data	Collected datasets from partners (WP1, WP2) Public datasets (WP2)	Classification: accuracy, F1-score, ROC-AUC Regression: RMSE, R^2 NLP: BLEU, ROUGE, perplexity Generation: diversity, Human-in-the-loop scoring Multi-label/ranking: NDCG, Hamming loss, MAP Explainability: SHAP, LIME Model cards Bias exploration and reporting	DFKI
Teacher focus group, interview, and questionnaire data	Collected during focus groups, interviews, and during the design and evaluation phase and feasibility trial (WP1, WP3, WP5)	Pilot testing; structured documentation	OUNL, UANTWERPEN, CNK
System pilot feedback focus group data	Feedback from classroom and mock-up system evaluations (WP1, WP3, WP5)	Pilot testing, interview guideline, thick description, validated measures (quantitative) and established methods for thematic analysis (qualitative) CRS-Que Framework, TAM Framework	DFKI, OUNL, AcrossLimits
Focus groups and interview transcripts	Educators, students, ethicists (WP1, WP4)	Pilot testing, interview guideline, thick description	CNR, UANTWERPEN, OUNL, CNK

Ethics and legal documentation	Consent forms (WP4)	Will be described the way the consent will be collected	CNR, UANTWERPEN, OUNL, CNK, UVEG, UTU
Contact information for communication and dissemination	Voluntary sign up on website, against consent disclaimer (mailing list) Consent forms (events)	Established GDPR-compliant consent forms	AcrossLimits, CNR

3. Other research outputs

EYE-TEACH will produce several research outputs (other than data), such as

- published digital/physical materials (articles, guidelines, materials)
- digital outputs (models, software)
- non-print media (videos, images)

The FAIR data management of these research outputs is described below.

Published digital/physical materials

Published materials and guidelines will be made accessible through preprint archives (e.g., EdArXiV), self-archiving, and/or open access publishing and repositories, as well as stored on the project website.

Digital outputs

Digital outputs will be made accessible through software and source code repositories (e.g., OSF, GitHub) or other open repositories (e.g., Zenodo) and attached to DOIs.

Non-print media

Non-print media will be stored on the project website for at least 5 years after project end.

4. Allocation of resources

4.1. Costs and responsibilities

Direct and indirect costs for making data or research outputs FAIR (relating to e.g., storage, archiving, re-use, security, etc.) are covered by each partner institution responsible for collecting data, as project personnel work time will be allocated for these tasks.

The Coordinator (UTU) will be responsible for overseeing data management and that the DMP is followed. Each WP or Task Leader who acts as Data Controller (possibly with other partners that act as joint controllers) is responsible for managing the data collected under their WP/Task according to the project DMP.

For studies in which more than one partner jointly determines the purposes and essential means of processing, the partners will formalise their relationship through a Joint Controllership Agreement in accordance with Article 26 (GDPR). In EYE-TEACH, this applies in particular to the relevant activities jointly carried out by OUNL and UANTWERPENen. Where a partner or third-party processes personal data on behalf of the controller(s), processing will be governed by a Data Processing Agreement defining the subject matter, duration, nature and purpose of the processing, the categories of data involved, and the technical and organisational safeguards to be applied. This applies, where relevant, to processing by CNK on behalf of the responsible controller(s).

4.2. Long-term preservation

Published pre-processed data and its metadata will be permanently stored in OSF. The original data will be stored on a secure cloud server 5-10 years after the project, as mandated by the Grant Agreement. The 5-year retention period ensures that the European Commission can conduct necessary audits and verify project implementation. Only individual Data Controllers (WP/Task leaders) will have access to their respective data. Once it has been analysed, original data will be securely disposed of. Any contact information of research participants will be destroyed after

the participants have concluded their participation. Contact information of community engagement participants will be destroyed after 5 years from project end.

Long-term storage of published data on OSF is guaranteed by the preservation fund established by the Center of Open Science. The preservation fund will ensure the accessibility of data for at least 50 years if OSF's operations were to be curtailed. As long as OSF remains operational, data storage is functionally unlimited.

Costs for sharing and preserving scholarship on OSF is based on the storage volume of contents stored. Research projects that exceed the 50 GB OSF public storage limit or the 5 GB private storage limit will require extra capacity. If additional capacity is needed, it will be purchased by the project.

5. Data security

5.1. Data handling security

Research data will be handled according to information protection and processing instructions of the project DMP and DPIA. The original and sensitive data will not be shared between Data Controllers or Processors. Data Controllers are responsible for storing, sharing and recovering original data. Data shared with Data Processors will always be a pre-processed copy of original data. Data Processors are responsible for storing, backups and recovery of the data shared to them by Data Controllers. For more information on data security measures per partner, see Table 3.

Any participant contact information will be stored separately from the data in a crypted format on a designated data-controlling researcher's password-protected computer and backed up in the cloud file service. These files will be backed up in respective universities' cloud services or external hard drives on a regular basis.

Table 3. Data security measures per partner (Data Controllers and Data Processors).

Partner	*Data storage and backup location(s)*	*Data sharing*	*Data recovery*
UTU	Individual password-protected computers with personal online network folders Individual, password-protected and internally hosted Seafile cloud storage	Access rights (UTU staff) or password-protected link (other partners) to shared, password-protected Seafile cloud storage	Daily automatic backups on personal online network folder Automatic synchronisation to Seafile cloud storage Manual backups to Seafile cloud storage
UANTWERP	OneDrive for Business Microsoft Teams or Sharepoint UANTWERPEN central servers	Microsoft Teams or Sharepoint (with password protected link) for shared data among WP1 team members Belnet FileSender for sharing data with other users	For recovery, we will fall back on the standard recovery methods provided by Microsoft as also described on the Pintra ICT webpage
UVEG	Individual password-protected computers with personal online network folders Individual, password-protected files on DISCO (a personal virtual storage service provided by UVEG)	Access rights (UVEG staff) or password-protected link (other partners) via Consigna (data sharing service provided by UVEG)	Daily automatic backups on personal online network folder Manual backups to DISCO virtual storage
OUNL	Individual password-protected computers with online network folders	SURF Researchdrive MS Teams / Sharepoint	Daily automatic backups on personal online network folder
DFKI	Individual password-protected computers with personal online network folders	MS Teams / Sharepoint	Standard recovery methods provided by Microsoft
AcrossLimits	GoogleDrive for Business	Internal folder accessible only to AcrossLimits	Standard recovery methods provided by Google

		team members or invited users	
CNR	OneDrive for Business Microsoft Teams or Sharepoint	MS Teams / Sharepoint	Standard recovery methods provided by Microsoft
CNK	OneDrive for Business Microsoft Teams or Sharepoint	MS Teams or Sharepoint (with password protected link) for shared data among team members	Standard recovery methods provided by Microsoft

5.2. Data storage security

Long-term storage of published data on OSF is guaranteed by the preservation fund established by the Center of Open Science. The preservation fund will ensure the accessibility of data for at least 50 years if OSF's operations were to be curtailed. As long as OSF remains operational, data storage is functionally unlimited.

Safe storage on OSF is ensured by project privacy settings, password security and two-factor authentication, European storage location, GDPR compliance, backup preservation, and encryption of data transfer.[3]

6. Ethics and legal compliance

Detailed ethics-compliance documentation for EYE-TEACH, including a summary of ethical clearances, study protocols, consent and assent materials, DPIA-related procedures, and the project's ethical framework and guidelines, is maintained in WP4 documentation, in particular deliverables D4.1 and D4.2. The present DMP does not duplicate that material in full. Instead, it summarises the ethical and legal requirements that directly affect data management, including lawful data collection, minimisation, storage, access control, sharing restrictions, preservation, and deletion of research data.

[3] OSF Security and Privacy, accessed 22 May 2025: https://help.osf.io/article/391-security-and-privacy; OSF Account and Security FAQ, accessed 22 May 2025: https://help.osf.io/article/547-account-and-security-faq-s

Data handling within EYE-TEACH is subject to safeguards, including restricted access to identifiable data, separation of contact data from research data, controlled sharing of sensitive datasets, and destruction or continued restriction of datasets that cannot be effectively anonymised. Partner institutions remain responsible for local ethics compliance and data protection implementation, with coordination between UTU and CNR.

6.1. Ethical risks

The ethical and legal issues impacting data sharing in EYE-TEACH have been described in-depth in the Description of the Action (DoA)[4]:

The first identified legal and ethical risk of EYE-TEACH relates to the processing of personal and biometric (sensitive) or biometric- based eye-tracking data, especially when collected from children. Whether eye-tracking data is biometric or only biometric-based depends on the system used for collecting the data. Some eye-trackers store the video image of the eye, in which case the data is biometric. Some systems parse the recorded video image online and store only some parameters, specifically, pupil size and pupil coordinates at each video sample. Online parsed data can be considered as biometric-based data. Biometrics collection is allowed under Art9.2 and Art89 of GDPR, but requires explicit consent.

The second identified risk is obtaining informed consent from participants regarding their interaction with AI systems. Due to the combination of a relatively unknown biometric-based measurement—if indeed considered biometric (see paragraph above)—to the general public (eye movements) and artificial intelligence, the project must ensure that the information provided to participants is understandable enough for them to give informed consent. Special consideration needs to be given to child participants, not only as research subjects but also as end-users of the developed AI pilot system. This requires gathering informed consent from parents/guardians during research activities and providing training for teachers and education professionals in the ethics of such systems as well as the processes of gathering informed consent.

[4] EYE-TEACH Grant Agreement, Description of the Action, chapter 4. Ethics self-assessment (p. 38-45).

Thirdly, the profiling of the AI system also presents ethical and legal risks. AI algorithms will build profiles based on ET data, and classify and make decisions on student performance. Automated profiling is not prohibited under Recital 71 of the GDPR, but the subject has the right not to be subject to a decision based solely on automated processing, including profiling. Again, as with personal and sensitive data, the participants must be informed of the existence of automated profiling/decision-making, the meaningful information behind the processing of data, and the impact or consequences of profiling.

Biometric data, profiling and automated decision-making all attract special consideration according to GDPR, especially for children as data subjects. This will be reflected in the level of scrutiny and security that will be applied when securing our data. WP2, WP3 and WP4 will select existing datasets for training the AI prototype, evaluating both scientific and ethical and privacy aspects.

Each Data Controller will apply for ethical approval and ask for ethical guidance from their respective partner institution. Each Data Controller will additionally complete a privacy register for the processing of personal data.

6.2. Research protocols

Before data collection begin, all EYE-TEACH studies involving human participants must complete the internal **Research Protocol Template** developed by the CNR team as part of the WP4 activities and revised by the Ethics and Legal Advisory Board (ELAB) of EYE-TEACH. The ELAB consists of external experts providing insights for the project during biannual online and in-person meetings with members of the consortium. The ELAB and its activities will be described in deliverable D4.2.

In summary, the Research Protocol Template, which will be described in detail in deliverable D4.2, documents (i) scientific rationale and study design, (ii) participant characteristics (including minors), (iii) risk–benefit assessment, (iv) informed consent and assent procedures, (v) AI ethics and trustworthy AI aspects, (vi) data protection and roles (controller/processor), and (vii) AI Act applicability and risk classification. Completed and signed protocols, together with ethics approvals, consent materials, DPIA summaries (where applicable), and data processing agreements, are stored centrally by CNR as part of the project's ethics and data governance record.

The template does not replace the official documentation required by national ethics committees, nor does it serve as a standalone ethics approval form. It does not substitute any documentation required under EU or national regulations, such as a Data Protection Impact Assessment (DPIA) under the GDPR. Instead, it functions as an internal harmonisation and quality assurance tool that facilitates ethical reflection and the preparation of official documentation.

6.2.1. Informed consent

As detailed in deliverable D4.1, in all studies, the consent process distinguishes between (I) consent to participate in the study and (ii) consent for data processing. Participants may withdraw from the study at any time without consequence. They may also request withdrawal of their consent for data processing; in such cases, their personal data will be erased unless this would render the research objectives impossible or seriously impaired, in line with GDPR Art. 17(3)(d). Information sheets should include a clear contact point for exercising rights of access, rectification, erasure, and restriction of processing. Informed consent for data sharing and long-term preservation (up to 10 years) should also be included in questionnaires dealing with personal data.

For minors, we obtain parental/legal guardian consent and age-appropriate assent from the child, in line with CNR's child protection and consent toolkits and EU guidance. Children's dissent is respected even when parental consent has been given.

6.2.2. AI Act applicability and risk classification

The AI Act (Regulation (EU) 2024/1689) is explicitly considered in the Research Protocol Template. The aim is to determine whether activities using AI are covered by the scientific research exemptions. More specifically, the studies do not appear to benefit from the exemption for systems "specifically developed and put into service for the sole purpose of scientific research and development" (Art. 2(6)). However, they could benefit from the exemption for systems intended for market release that are still in the research and testing phase, provided they have not been tested in 'real-world conditions' (Art. 2(8)). Where the exemption does not apply, AI systems in educational settings are treated as high-risk by default (Annex III), provided they explicitly avoid any functionality that would amount to emotion recognition in

education, which is prohibited under Art. 5(1)(f). Instead, they focus on indicators such as reading behaviour and cognitive load. Despite the exemption, it would still be advisable to carry out a Fundamental Rights Impact Assessment (FRIA), in accordance with Article 27. This should take into account the Data Protection Impact Assessment (DPIA) where required, as defined in paragraph 4 of the AI Act.

6.2.3. Data protection

At project level, we distinguish three data states:

- **Personal data (including pseudonymised data)**: any dataset from which individuals can be identified directly or indirectly; fully subject to GDPR.
- **Anonymised data**: datasets that have undergone an irreversible anonymisation process. The GDPR no longer applies after the anonymisation, but, anonymisation is a form of data processing in itself. Moreover, ethical obligations and good practice still apply.
- **Anonymous at source**: data that never contains identifiers and cannot reasonably be linked to individuals.

For each study, the Research Protocol Template requires the PI to classify the data status, identify the data controller(s), list any data processor(s) (e.g., survey platforms), and confirm that appropriate controller–processor or joint-controller agreements are in place.

6.3. Ethics self-assessment of the AI-assisted ET-analytics tool

EYE-TEACH has conducted an initial ethical self-assessment of its AI-assisted ET-analytics tool using the Assessment List for Trustworthy Artificial Intelligence (ALTAI), as reported in D4.1 Annex 1. The assessment forms part of the project's ethics-and-privacy-by-design approach and covers the seven requirements of trustworthy AI: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity and fairness, societal well-being, and accountability. In practice, this includes measures for human oversight, transparency of algorithmic outputs, privacy-by-design, bias mitigation, explainability, auditability, and user training. The ALTAI self-assessment will be updated during the project,

including in connection with the M18 and M36 ethics reviews, and complements the project's GDPR, DPIA, institutional ethics review, and WP4 governance processes.

7. Relevant national and departmental procedures for data management

Data Controllers will make use of the following national or departmental procedures for data management (see Table 4).

Table 4. List of relevant national and departmental procedures for data management.

Partner	*Procedures*
DFKI (Germany)	Federal Data Protection Act (BDSG) DPO contact: datenschutz@dfki.de
OUNL (Netherlands)	Uitvoeringswet Algemene verordening gegevensbescherming (UAVG) Netherlands Code of Conduct for Research Integrity OUNL Code of Conduct DPO contact: FG@ou.nl
UA (Belgium)	Belgian Data Protection Authority: Act on the protection of natural persons with regard to the processing of personal data Recommendations formulated by the European Data Protection Supervisor UANTWERPEN Open Science and Research Data Management Guidelines: Data Management Plan (Department Research Affairs & Innovation, UANTWERPEN) DPO contact: privacy@uantwerpen.be
UTU (Finland)	Finnish Data Protection Act (Tietosuojalaki 1050/2018) University of Turku Data Protection Policy University of Turku Library: Lifecycle planning for research data Digital preservation services offered by the Ministry of Education and Culture in Finland: Fairdata.fi DPO contact: dpo@utu.fi

UVEG (Spain)	Guidelines provided by the Data Protection Office of the University of Valencia More information and external guidelines on data protection Regulations provided by the Spanish Ministry of Education, Vocational Training, and Sports DPO contact: lopd@uv.es
CNR (Italy)	Italian Data Protection Authority (Garante): Legislative decree 30 June 2003, n.196 containing the "Personal Data Protection Code" Regulation setting out the requirements relating to the processing of special categories of data – Section 5. Requirements relating to the processing of personal data for scientific research purposes Ethical guidelines for processing for statistical or scientific research purposes (2018) Law No. 132 of 23 September 2025, Provisions and powers delegated to the Government regarding artificial intelligence CNR's "Guidelines for Research Integrity" DPO contact: dpo@cnr.it
CNK (Poland)	Ordinance No. 42/2023 of the Director of the Copernicus Science Centre dated 30 June 2023 on the introduction of the Data Protection Policy at the Copernicus Science Centre. Personal Data Protection Office DPO contact: iod@kopernik.org.pl