Action Number: 101079338

Action title: Towards Excellent Robotics and Artificial Intelligence at a Slovak University

Action Acronym: TERAIS

Deliverable 6.6

Data Management Plan – Final version

WP6 – Project Management

Authors: Daniela Olejarova, Omar Eldadeer, Hassan Ali, Viktor Kocur, Martin Takac, Igor Farkas

Lead participant: UKBA (Comenius University Bratislava)

Date: 30/09/2025

Dissemination level: public

DMP Version: 2.0

Table of Contents

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338

1. Introduction

This deliverable presents the final version of the Data Management Plan (DMP) for the TERAIS project and provides the analysis of the data management policy applied by the Partners to datasets generated within the project as well as other data related to organisation of the events within the project. In particular, the DMP identifies the main datasets and describes research data management during the project, as well as how and what parts of the datasets have openly been shared, made accessible for verification and re-use, and been curated and preserved. The goal of this DMP is to facilitate effective internal data management and make data FAIR (Findable, Accessible, Interoperable, and Reusable).

This document provides guidance to the project partners on data management and it is a useful tool to agree on data processing, facilitate the creation of a common understanding and, where possible, common practices. This deliverable is submitted to the European Commission at M36 (30/09/2025) and represents the final plan. The DMP has, in fact, been a living document, having been updated and further refined with the TERAIS project progress. It is also important to remark that this DMP reflects the provisions established by the project contracts and complements the planned project exploitation, dissemination and IPR procedures.

2. DMP management and update

The document's development and updates have mainly relied on the collection of information about datasets filled out by each project partner responsible for producing such data (see Section 3.2). The form used to collect this information (i.e., Dataset Questionnaire Form, DQF) was prepared and has constantly been updated by UKBA, UHAM and IIT Offices involved in this process under the supervision of Daniela Olejarova as the TERAIS Project Manager assigned with responsibility of overall coordination in this area.

The DMP deliverable, including an editable copy and the DQFs, has been available to all partners on **the project shared drive** (UKBA institutional Google drive). The document has been updated as appropriate along the project duration. The different versions have been numbered and dated for identification. Official versions in a pdf format have been stored on the **TERAIS project website**. When a new dataset was identified along the project implementation, partners submitted a new form containing the new identified dataset and notified the coordinator. UKBA was then in charge of updating the document and its annexes and notifying the Consortium through the project mailing list system.

3. Data summary

3.1 TERAIS datasets

The final version of the TERAIS DMP is based on the description of six datasets, of which key details are summarised in the following table (Table 1).

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338

Table 1. *Summary of the TERAIS research datasets*

DATASET NAME	PARTNERS	ORIGIN	TYPE	FORMAT	SIZE	WP & TASK	ACCESSIBILITY[1]
DATASET 1: **TERAIS_UKBA_Dataset_ExperimentalData**	UKBA IIT	Generated within the project	**Quantitative tabular** data **Quantitative and textual** data **Digital image** data **Digital audio** data **Digital video** data **Documentation and scripts**	.csv, .txt, .omv .xml, .txt, json, .docx, .rtf .tiff, .jpeg, .pdf, .bmp, .png .wav, .mp3 .mp4, .ogv .txt, .htm, .pdf According to other proprietary software used for the analysis other formats might be .m, or in case other open-source formats will not be available, proprietary formats of devices used for data collection.	**max. 2 TB**	**WP3**: Task 3.1	Confidential to partner before publication Open access after publication
DATASET 2: **TERAIS_UKBA_Dataset_Code**	UKBA	Generated within the project	**Code and scripts -** The data will be produced by project participants who will write code. It will be organised using the git versioning system. The code will be written in specific languages such as Python or C++ and follow common standards for the relevant languages.	Most files will be stored in standard files containing text with relevant extensions for code and scripts (.py, .c, .h, .cpp, .hpp, .sh, .bat). Some configuration files may be stored as .json, .csv or .xml files or other human readable formats. Other data might be stored in text files such as .txt, .md. Documentation might be stored in .pdf or .html format. Some code might be stored in the form of compiled binary files.	**100 MB**	**WP3:** Task 3.1	Confidential to partner before publication Open access after publication
DATASET 3:	IIT UHAM UKBA	Generated within the project	Data will be extracted from **sensors with proprietary software** (e.g. Tobii: Tobii-Pro Lab,	non-proprietary formats will be preferred for data saving and data storage. For instance: • **Quantitative tabular data:** .csv, .txt, .omv	**max. 2 TB**	**WP3**: Task 3.1	Confidential to partners before publication

[1] *In case of patenting or potential commercial exploitation, different levels of confidentiality can be foreseen (in case of Collaborative Projects):*

- ***Beneficiary institution access:*** *The disclosure of information is not provided at any level and/or any time. This option is applied when, regardless of the long-term value and scope for wider use, the dataset contains data that would lose their value if disclosed (e.g. experimental results and validation). The beneficiary who chooses to protect this information from any external access aims at patenting/exploiting it or at protecting trade secrets and must ensure confidentiality beyond the clauses agreed in the Consortium Agreement;*
- ***Confidential to partner:*** *This option is applied when, regardless of the long-term value and scope for wider use, the dataset contains personal data that cannot be protected once disclosed. These may include among others videos and images collected during the project tests;*
- ***Confidential to Consortium including EC services (Consortium only):*** *This option is applied for data containing confidential information (e.g. exploitable results) requiring IP protection, aimed at possible exploitation. In certain cases to be defined along with the development of the project, the owners may allow Open Access upon dissemination of the associated results in peer-reviewed scientific publications;*
- ***Open Access:*** *This option is applied when data has no IP restrictions and will be openly available and re-usable.*

DATASET NAME	PARTNERS	ORIGIN	TYPE	FORMAT	SIZE	WP & TASK	ACCESSIBILITY[1]
TERAIS_Dataset_IIT_ExperimentalData			Shimmer, Optitrack, OptoForce, EyeLink Sr-Research), open software (e.g. Pupil Core) through **applications developed ad hoc** (e.g., Python or Matlab libraries to interface with device and collect data in real-time), or through **proprietary software for collecting questionnaire data** (e.g. Survey Monkey). Data will be stored with storage tools (e.g., Network Attached Storage), following guidelines of Istituto Italiano di Tecnologia. **Statistical analysis will be conducted with software such as R, Jamovi, SPSS, Statistica, and Matlab**.	• **Qualitative and textual data:** .xml, .txt, .json, .docx, .rtf • **Digital image data:** .tif, .jpg, .pdf, .bmp, .png • **Digital audio data:** .wav, .mp3 • **Digital video data:** .mp4, .ogv • **Documentation and scripts:** .txt, .htm, .pdf According to other proprietary software used for the analysis other formats might be .m, or in case other open-source formats will not be available, proprietary formats of devices used for data collection.			Open access after publication
DATASET 4: **TERAIS_Dataset_IIT_Code**	IIT UHAM UKBA	Generated within the project	**Codes**	According to the programming language, files will be stored, for instance, in the following formats: • **Code:** .cpp, .h, .py, .r, .m • **Quantitative tabular data:** .csv, .txt, .omv • **Documentation and scripts:** .txt, .htm, .pdf • **Qualitative and textual data:** .xml, .txt, .json	**100 MB**	**WP3**: Task 3.1	Confidential to partners before publication Open access after publication
DATASET 5: **TERAIS_Dataset_UHAM_PointingGestures**	UHAM IIT	Generated within the project	**Experimental data** – The data will be collected using the robot's proprioceptive and external sensors. Videos of the robot's eye cameras and top- and side-view cameras will be collected as well as logs of the interaction between the experimenter and the robot.	video files (ex: .mp4), csv data, and text data (The size of the data will depend on the number of experiments and the experiments' durations. Based on previous experience, the size might be around 5 GB for each minute of recording. This is only a mere estimation, however, it is always possible to compress the data.)	**5 GB** for each minute of recording	**WP3**: Task 3.1	Confidential to partner before publication Open access after publication
DATASET 6: **TERAIS_dataset_UHAM_Code**	UHAM, UKBA, IIT	Generated within the project	**Codes**	Most files will be stored in standard files containing text with relevant extensions for code and scripts (.py, .c, .h, .cpp, .hpp, .sh, .bat). Some configuration files may be stored as .json, .csv or .xml files or other human readable	**100 MB**	**WP3:** Task 3.1	Confidential to partner before publication

DATASET NAME	PARTNERS	ORIGIN	TYPE	FORMAT	SIZE	WP & TASK	ACCESSIBILITY[1]
				formats. Other data might be stored in text files such as .txt, .md. Documentation might be stored in .pdf or .html format. Some code might be stored in the form of compiled binary files.			Open access after publication

The majority of the identified datasets have long-term value. Datasets are useful to several categories of research communities, and users from research, industry, and society, including:

Research and scientific community

- Robotics & AI
- Social & Cognitive Robotics
- Human-Robot Interaction & Human-Computer Interaction
- Social & Developmental Psychology
- Cognitive & Neuroscience
- Computer Vision

Industry

- Development of Co-bots
- Machine Learning algorithms for autonomous robots and systems
- User Experience Design of technological devices

Society

- Specific beneficiary categories of Social Robots (examples from literature: elderly, children with developmental disorders, children with learning disorders)
- Increased and safer employment of robots in social institutions

Within these contexts, datasets may have several re-uses in novel similar research studies. Detailed expected utility for each dataset to be generated has been reported in the tables of the next section.

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338

3.2. Extended TERAIS datasets' information

Table 2. *Presentation of the extended information on the* ***TERAIS datasets 1-3***

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
Main dataset (name)	**TERAIS_UKBA_Dataset_ExperimentalData**	**TERAIS_UKBA_Dataset_Code**	**TERAIS_Dataset_IIT_ExperimentalData**
Sub-dataset (name)	Empirical Research Dataset	Code Dataset	ExperimentalData
Responsible partner	**UKBA**	**UKBA**	**IIT**
Other partners involved	IIT – Italian Institute of Technology UHAM – University of Hamburg		UHAM – University of Hamburg UKBA – Comenius University Bratislava
Goal	This Dataset will be used for Work Package 3 (Achievement of Scientific Excellence) see Grant Agreement – Annex 1, Part B, par. 1.2.4 "Supporting scientific excellence and international collaboration – RESEARCH and NETWORKING".	This Dataset will be used for Work Package 3 (Achievement of Scientific Excellence) see Grant Agreement – Annex 1, Part B, par. 1.2.4 "Supporting scientific excellence and international collaboration – RESEARCH and NETWORKING".	This Dataset will be used for Work Package 3 (Achievement of Scientific Excellence) see Grant Agreement – Annex 1, Part B, par. 1.2.4 "Supporting scientific excellence and international collaboration – RESEARCH and NETWORKING".
Data origin	Generated within the project	*Generated within the project*	*Generated within the project*
	Please justify below the need of new data to be generated	*Please justify below the need of new data to be generated*	*Please justify below the need of new data to be generated*
	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project, would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable, and trustworthy robots.	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable, and trustworthy robots	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project, would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable, and trustworthy robots.
Data collection	The dataset will contain various types of sensor data including images, videos, sound recordings, depth data, motion capture data. Sensor data will be produced using both proprietary or non-proprietary software based on the relevant hardware (e.g. Tobii: Tobii-Pro Lab, Shimmer, Optitrack, OptoForce, EyeLink Sr-Research, Pupil	The data will be produced by project participants who will write code. It will be organised using the git versioning system. The code will be written in specific languages such as Python or C++ and follow common standards for the relevant languages.	Data will be extracted from sensors with proprietary software (e.g. Tobii: Tobii-Pro Lab, Shimmer, Optitrack, OptoForce, EyeLink Sr-Research), through applications developed ad hoc (e.g., Python or Matlab libraries to interface with device and collect data in real-time), or through proprietary software for collecting questionnaire data (e.g. Survey Monkey) . Data will be stored with

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
	Core, LeapMotion, IntelRealSense), or through applications developed ad hoc (e.g., Python or Matlab libraries to interface with device and collect data in real-time, such as reaction time). The dataset will also contain questionnaires given to participants which will be collected through proprietary software for collecting questionnaire data (e.g. Survey Monkey). The dataset will also contain analysis of other data contained within the dataset or other forms of derived data. Such data may be produced via statistical analysis conducted with software such as R and Jamovi. Sensor data might be processed by proprietary or non-proprietary software or software developed within the project (e.g. estimation of human pose from images, transcription of recorded voice).		storage tools (e.g., Network Attached Storage), following guidelines of Istituto Italiano di Tecnologia. Statistical analysis will be conducted with software such as R, Jamovi, SPSS, Statistica, and Matlab.
Dataset type	*Other* **Raw experimental data and Results/Analysed Data**	*Tools* **Code**	*Other* **Raw experimental data and Results/Analysed Data**
File formats	Non-proprietary formats will be preferred for data saving and data storage. For instance: • Quantitative tabular data: .csv, .txt, .omv • Qualitative and textual data: .xml, .txt, .json, .docx, .rtf • Digital image data: .tif, .jpg, .pdf, .bmp, .png • Digital audio data: .wav, .mp3 • Digital video data: .mp4, .ogv • Documentation and scripts: .txt, .htm, .pdf, .md According to other proprietary software used for the analysis other formats might be .m, or in case other open-source formats will not be available, proprietary formats of devices used for data collection.	Most files will be stored in standard files containing text with relevant extensions for code and scripts (.py, .c, .h, .cpp, .hpp, .sh, .bat). Some configuration files may be stored as .json, .csv or .xml files or other human readable formats. Other data might be stored in text files such as .txt, .md. Documentation might be stored in .pdf or .html format. Some code might be stored in the form of compiled binary files.	Non-proprietary formats will be preferred for data saving and data storage. For instance: • Quantitative tabular data: .csv, .txt, .omv • Qualitative and textual data: .xml, .txt, .json, .docx, .rtf • Digital image data: .tif, .jpg, .pdf, .bmp, .png • Digital audio data: .wav, .mp3 • Digital video data: .mp4, .ogv • Documentation and scripts: .txt, .htm, .pdf According to other proprietary software used for the analysis other formats might be .m, or in case other open-source formats will not be available, proprietary formats of devices used for data collection.
Expected volume of data	**maximum 2 TB**	**100 MB**	**maximum 2 TB**

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
Expected time of release	We will release parts of the dataset in individual repositories upon publication of papers that reference it. We may postpone the release of some parts of the datasets until all of the papers relevant for a given study are published.	We will release parts of the developed codebase in individual repositories upon publication of papers that reference it. We will also release code in cases where it would be beneficial for users of our other published datasets as a tool to facilitate a better use of the published datasets. We may postpone the release of some parts of the code until all of the papers relevant for a given study are finished.	We will release datasets after publication of papers that reference it. In any case, data will be released one year after the end of the project, to ensure having time for publishing results on such data.
DOCUMENTATION AND DATA QUALITY			
Metadata and documentation	Beyond giving as much information as possible in published papers to ensure reproducibility of the studies, metadata will be incorporated into the dataset to help others identify, discover, and reuse data. The DCMI standards or similar will be used for documentation with respect to the following aspects: title, abstract, description, access right, creator, audience, availability, bibliographic citation, contributors, format, identifier. • Data will be organised in folders, one for each single experiment referred to the dataset. According to the experiment, the access to data could be restricted, shared, or made open. • Data will be divided into raw data and analysed data. According to the experiment and the nature of data, only analysed data might be made accessible. • Data will be divided in folders according to their nature and format. • Readme text and, if possible, graphic representation of the dataset will be released. • Links to papers or additional materials explaining the procedure of the study will be provided.	We will use the git versioning system, which keeps track of the metadata for the repository. Codebase documentation will also be provided for the scientific community to be able to utilise the code for their own research. When released the code will be hosted in public git repositories either hosted on third-party services such as GitHub or on GitLab servers hosted by one of the participating institutions. Specific releases of code versions without the associated git metadata will be released and assigned a DOI.	Beyond giving as much information as possible in published papers to ensure reproducibility of the studies, metadata will be incorporated into the dataset to help others identify, discover, and reuse data. DCMI, Datacite, DDI standards or similar will be used for documentation. Dataverse ensures exporting metadata and data schema in several formats. • Data will be organised in folders, one for each single experiment referred to the dataset. According to the experiment, the access to data could be restricted, shared, or made open. • Data will be divided into raw data and analysed data. According to the experiment and the nature of data, only analysed data might be made accessible. • Data will be divided in folders according to their nature and format. • Readme text and, if possible, graphic representation of the dataset will be released. • Links to papers or additional materials explaining the procedure of the study will be provided.

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
Keywords	Human-Robot Interaction; Human-awareness; Trust; Explainability; Cognitive Robotics; Social Robotics	Human-Robot Interaction; Explainability; Trust; Human-awareness; Cognitive Robotics; Social Robotics	Human-Robot Interaction; Human-awareness; Trust; Explainability; Cognitive Robotics; Social Robotics. According to the type of dataset and research, tools such as MeSH on Demand will be used to ensure the findability of the dataset.
Data quality	Experiments will be conducted based on previous experience with experimental settings and data collection, as well as pilot and preliminary data analysis. Questionnaires and experimental stimuli will be piloted using cognitive interviews with participants in order to ensure validity and reliability of the measures. When possible, video or image data will be recorded by more cameras (e.g. external cameras, robot's cameras). Human gaze and kinematic data will be recorded after previous calibration.	Code will be verified by its execution on relevant hardware. Code segments may be reviewed by multiple people. Some of the functionality might be verified using unit tests.	Experiments will be conducted based on previous pilot studies exploring problems in experimental settings and data collection, and early data analysis. When possible, video or image data will be recorded by more cameras (e.g. external cameras, robot's cameras). Human gaze and kinematic data will be recorded after previous calibration.
STORAGE AND BACKUP DURING THE RESEARCH PROCESS			
Storage and backup solutions	Data will be stored and backed-up in the institutional storage system hosted at UKBA, which will utilise the RAID storage system to ensure redundancy. Data will also be backed up to the institutional OneDrive cloud storage or institutional Google Drive storage. Data that will be shared for future reuse by other researchers will be uploaded on open science repositories such as Zenodo or OSF.	The git versioning system supports the codebase to be stored both in remote storage (GitLab, GitHub) and locally on the devices used by researchers. The GitHub remote storage is hosted by a third-party which ensures data security. GitLab remote storage will be hosted by UKBA on a server with RAID storage system which will provide data redundancy. Specific releases of code versions without the associated git metadata will be released on platforms such as Zenodo.	Computer data and configuration will be backed-up on an encrypted external storage system at least every week and software installed on computers will be updated regularly. Data will be tested regularly to ensure their usability in case of necessity.
Data security and protection	Data will be processed with a secure LAN network. Every system will be accessible only via authentication with credentials assigned to every person responsible for data processing. All systems will be monitored. Data will be protected following the technical and organisational safety measures to ensure protection from unauthorised	The codebase will not contain sensitive data therefore we will use standard practices. The private repositories will be accessible by researchers working on the project via their account protected with passwords or authorization tokens/keys. The public repositories will be visible to anyone, but only	Data will be processed with a secure LAN network. Every system will be accessible only via authentication with credentials assigned to every person responsible for data processing. Passwords will consist of at least 8 characters and will be modified by the person on first use and at least every three months. All systems will be monitored. Data will be protected following the technical

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
	access, data theft and/or data leakage. For this reason, data will be encrypted to be unreadable in case of data breach.	researchers working on the project will have the ability to modify them in the same manner as described for private repositories.	and organisational safety measures to ensure protection from unauthorised access, data theft and/or data leakage. For this reason, data will be encrypted to be unreadable in case of data breach.
LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT			
Personal/special data categories	*YES*	*NO*	*YES*
Protection of personal / special data categories	According to the provisions of the General Data Protection Regulation (GDPR), article 6 par. 1 point a), personal data of participants will be processed only after receiving their signed informed consent. As Data Controller, UKBA implements technical and organisational measures to ensure the correct Data security, in line with Articles 25 and 32 of GDPR above mentioned. Personal Data will be pseudonymized and stored in a repository system provided by UKBA. The mapping table linking the participant's ID information to corresponding codes will be stored in a different and encrypted repository system, accessible only to experimenters. Thereby, all data collected during the experiment for each participant will be processed in connection with the code assigned to participants rather than with their ID information. Personal Data will not be kept any longer than the time necessary for the objective they had been processed for (see Article 5, par. 1, point e) of GDPR). Access to the data will be provided only to project collaborators involved in the specific study and students collaborating in the study (under the researchers' supervision).	No personal or sensitive data will be included in this dataset.	IIT complies with Regulation no. 679/2016 with subsequent amendments (hereafter "GDPR"). For the studies carried out in Genova, IIT is the Data Controller of participants' personal data. In this role, IIT determines the purposes and methods of Data management and adopts the technical and organisational security measures described below to ensure that the Data Management is carried out in compliance with GDPR. These measures are re-examined and updated when necessary. As Data Controller, the Italian Institute of Technology (IIT) implements technical and organisational measures to ensure the correct Data security, in line with Articles 25 and 32 of GDPR above mentioned. After the participant signs the documents of informed consent and privacy, the documents will be stored in a secure archive via Enrico Melen 83, IIT, Genova. Such documents will be accessible to the Principal Investigator and to specifically authorised IIT personnel. The principal investigator and the research team are responsible for research data within the whole cycle: collection, processing, transferring and storing. In every phase, the principal investigator and the research team will ensure the correct application of technical and organisational security measures for data management. The Principal Investigator and the authorised personnel received specific training about personal data management within IIT, with particular attention to Data Security and Protection.

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
			Personal Data will be pseudonymized, encrypted, and stored in a repository system provided by IIT. The mapping table linking the participant's ID information to corresponding codes will be stored in a different and encrypted repository system, accessible only to experimenters. Thereby, all data collected during the experiment for each participant will be processed in connection with the code assigned to participants rather than with their ID information. In case anonymization is possible, it will be preferred to pseudonymization for data sharing. Complete anonymization will be also preferred to pseudonymization for data storage in case no relevant information will be lost. Personal Data will not be kept any longer than the time necessary for the objective they had been processed for (see Article 5, par. 1, point e) of GDPR). access procedure for authorised persons to process personal data.
Ethical issues issues (research data involving experiments with humans)	*YES* No minors will take part in the studies. Participants will be recruited by standard recruitment procedures, including word of mouth, social networks, and students. According to the provisions of the General Data Protection Regulation (GDPR), article 6 par. 1 point a), personal data of participants will be processed only after receiving their signed informed consent. The template of the informed consent is attached. As Data Controller, UKBA implements technical and organisational measures to ensure the correct Data security, in line with Articles 25 and 32 of GDPR above mentioned. After the participant signs the documents of informed consent and privacy, the documents will be stored in an institutional storage system. These documents will	*NO* No human subjects' data will be included in this dataset.	*YES* Participants will be recruited with a recruiting email sent to a local mailing list to which they previously subscribed. According to the provisions of the General Data Protection Regulation (GDPR), article 6 par. 1 point a), personal data of participants will be processed only after receiving their signed informed consent. Participants will be informed about the nature of the study and the procedure for data collection and management. They will be explicitly asked for authorization to collect and store data and to share data for scientific, dissemination, and institutional purposes in an anonymized form. Participants will be also explicitly asked for authorization to record video during experimental sessions that could be shared for scientific, dissemination, and institutional purposes without connecting them to ID information.

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
	be accessible to project collaborators involved in the specific study. UKBA personnel involved in the project received specific training about personal data management, with particular attention to Data Security and Protection. Personal Data will be pseudonymized and stored in a repository system provided by UKBA. All the Data related to the study must be kept stored for a minimum of 10 years after the regular or anticipated conclusion of the study. The mapping table linking the participant's ID information to corresponding codes will be stored in a different and encrypted repository system, accessible only to experimenters. Thereby, all data collected during the experiment for each participant will be processed in connection with the code assigned to participants rather than with their ID information. The studies will follow the national and international codes of conducts and institutional ethical guidelines, including the American Psychological Association guidelines and be reviewed by the UKBA ethical committee. Personal Data will not be kept any longer than the time necessary for the objective they had been processed for (see Article 5, par. 1, point e) of GDPR). The UKBA Data Protection Officer is available at the following e-mail address: dpo@uniba.sk or the post address: Univerzita Komenského v Bratislave, Centrum informačných technológií UK, Šafárikovo námestie 6, P.O. Box 440, 814 99 Bratislava. For the data collected at IIT and only shared with UKBA, the ethical protocol will be the one approved by "Comitato Etico Regione Liguria" for IIT.		For what concerns the attribution of tasks and duties to authorised subjects, the internal staff managing personal data is designated with the instrument of appointment. Communication of written instructions to internal staff authorised for personal data management. A channel for the communication of possible violations of obligations about Personal Data Management is available. Application of measures of de-identification of personal data so that data cannot be ascribed to a specific participant without using additional information. The IIT Data Protection Officer is available at the following e-mail address: dpo@iit.it. With respect to the type of studies and participants planned within the projects the Data Protection Impact Assessment (DPIA) was not evaluated as necessary. An evaluation from the DPO will be provided if requested. All the experiments will be carried out in line with national and international codes of conducts and institutional ethical guidelines. Specifically, they will follow ethical protocols approved by the regional ethical committee "Comitato Etico Regione Liguria".
Other ethical issues (e.g.	*NO*	*NO*	*NO*

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
involving animal subjects)	No data from animal subjects will be included in this dataset.	No data from animal subjects will be included in this dataset.	No data from animal subjects will be included in this dataset.
Intellectual property rights	Members of the research team from all participating institutions involved in the specific studies will have the ability to restrict or grant data access to others. Some parts of datasets will be made available for the public under the open science principles on the open science repository upon the publication of manuscripts relevant to the data.	The repositories will be owned by a group on the GitLab server or on GitHub. The group will consist of members of the research team from all participating institutions who will have the ability to restrict or grant access to a given repository to others.	As written in the Grant Agreement, "results are owned by the beneficiaries that generate them. However, two or more beneficiaries own results jointly if: I) they have jointly generated them and II) it is not possible to: a) establish the respective contribution of each beneficiary, or b) separate them for the purpose of applying for, obtaining, or maintaining their protection. The joint owners must agree — in writing — on the allocation and terms of the exercise of their joint ownership ('joint ownership agreement'), to ensure compliance with their obligations under this Agreement." This dataset stores data that is owned by IIT. Part of them could be also owned by the other beneficiaries if they comply with the abovementioned criteria (I and II). Data will be deposited in the institutional repository and, as soon as possible in compliance with the FAIR principle "As Open as Possible, as Closed as Necessary", will be re-organized and deposited in Zenodo Data Repository together with the related Metadata. Restrictions might arise in case the sharing will be i) against the beneficiary's legitimate interests, including commercial exploitation, or ii) contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement. However, if open access is not provided (to some or all data), it will be specified and justified. Metadata of deposited data will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
			embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata will include persistent identifiers for related publications and other research outputs.
DATA SHARING AND LONG-TERM PRESERVATION			
Data sharing	Specific datasets that will be made available under the open science principles will be uploaded on open science repositories, which will provide a persistent identifier and ensure discoverability by using specific keywords relevant to the dataset, or searching for the specific project. Datasets will also be referenced in respective publications or conference talks and may be reused by other researchers.	The repositories which were available under the open science principles will remain public and be accessible to everyone. The released code will be accessible using git-based tools or via a browser by visiting a public url. Links to relevant repositories will be included in the published papers and in conference presentations. We will also share links to repositories on public platforms (e.g. PapersWithCode.com) and social media. The repositories will be findable using common search engines. Code might also be provided to third-parties upon request when necessary for peer-review or follow-up work even before it is made accessible to everyone. Specific releases of code versions without the associated git metadata will be released and assigned a DOI at repositories such as Zenodo.	Data collected within an experimental study will be stored in the institutional repository (IIT Dataverse). Once the entire process of analysis and publication will be finalised, shareable data (not personal data) will be organised and shared in new datasets, comprehensive of shareable data and metadata, and linked with code datasets. So, after publication, such datasets will be provided with a persistent identifier (Digital Object Identifier) to ensure their findability and accessibility. Another possible repository for such datasets will be Zenodo.
Data repository	**Institutional**	**Institutional**	**Institutional or Zenodo**
	Zenodo	**GitHub, Zenodo**	**IIT Dataverse**
		https://github.com	https://dataverse.iit.it/
Restrictions on sharing	Anyone will be able to use the data shared in public repositories for non-profit research and educational activities.	Anyone will be able to use the data shared in public repositories for non-profit research and educational activities.	Data will be shared only once results of the experimental study they were used for will be published. During the period of access restriction, those who will be able to use the data are the beneficiaries, the other partners of the project according to the Grant

DATA SUMMARY	DATASET 1	DATASET 2	DATASET 3
			Agreement, and other partners of the network if the data have been collected in a collaboration project and, in any case, in compliance with the Grant Agreement.
Data curation	In accordance with the GDPR, personal data shall be stored in an encrypted storage for a period of no more than 5 years.	No data will be destroyed. Earlier versions of the codebased will be accessible via the git versioning system.	
Requirements for reusability	We will prioritise the use of non-proprietary data formats which could be accessed using non-proprietary tools. Some of the published data may require proprietary software, mostly related to the hardware used. We may also provide some software tools to work with the published data (see dataset Code).	Use of the git versioning software is recommended to access shared code, but is not required. In order to run the code publicly available compilers and interpreters will be required along with publicly available libraries. Some parts of the code may require proprietary software such as Matlab.	As far as possible, data will be shared using non-proprietary formats to ensure reusability. Information about devices, tools, software, versions, and libraries to collect and analyse data will be provided to ensure reproducibility.
Licensing	*CC BY - Attribution*	*CC BY - Attribution*	*CC BY – Attribution*
DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES			
Roles and responsibilities	Data management of this subset will be the responsibility of Xenia Daniela Poslon (xenia.poslon@fmph.uniba.sk)	Data management of the codebase will be the responsibility of Viktor Kocur (viktor.kocur@fmph.uniba.sk)	Once a new dataset will be created, roles and responsibilities for that specific dataset will be defined and indicated in updated versions of the DMP. Coordination of DMP will be assigned to the coordinator partner of the project, but responsibilities about datasets will depend on its beneficiary or beneficiaries. Regular updates of the DMP will be reported every time a new dataset will be needed.
Resourcing			

Table 3. *Presentation of the extended information on the* ***TERAIS datasets 4-6***

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
Main dataset (name)	TERAIS_Dataset_IIT_Codes	TERAIS_Dataset_UHAM_GesturesEstimation	TERAIS_Dataset_UHAM_Code
Sub-dataset (name)	Codes	Pointing Gestures - Addressee Estimation	Code Dataset
Responsible partner	**IIT**	**UHAM**	**UHAM**
Other partners involved	University of Hamburg, Comenius University Bratislava	IIT	UKBA, IIT
Goal	This Dataset will be used for Work Package 3 (Achievement of Scientific Excellence) see Grant Agreement - Annex 1, Part B, par. 1.2.4 "Supporting scientific excellence and international collaboration - RESEARCH and NETWORKING"	Design and develop HRI scenarios to improve intention understanding	This Dataset will be used for Work Package 3 (Achievement of Scientific Excellence) see Grant Agreement - Annex 1, Part B, par. 1.2.4 "Supporting scientific excellence and international collaboration - RESEARCH and NETWORKING".
Data origin	*Generated within the project*	*Generated within the project*	*Generated within the project*
	Please justify below the need of new data to be generated	Please justify below the need of new data to be generated	*Please justify below the need of new data to be generated*
	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project, would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable, and trustworthy robots.	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project, would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable and trustworthy robots.	Achieving scientific excellence is a core objective of the project. To this aim, it is crucial for research to grow and progress, based on the acquisition of novel empirical data that, in the case of this project would be necessary to investigate humans in HRI scenarios and develop human-aware, explainable, and trustworthy robots
Data collection	**Codes**	The data will be collected using the robot's camera. **Videos of its interaction with objects and its surroundings will be collected**. Apart from the experimenters, no persons are visible.	The data will be produced by project participants who will write code. It will be organised using the git versioning system. The code will be written in specific languages such as Python or C++ and follow common standards for the relevant languages.
Dataset type		*Experimental data*	*Tools*
			Code
File formats	According to the programming language, files will be stored, for instance, in the following formats:	video files (ex: .mp4), csv data, and text data	Most files will be stored in standard files containing text with relevant extensions for code and scripts

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
	• Code: .cpp, .h, .py, .r, .m • Quantitative tabular data: .csv, .txt, .omv • Documentation and scripts: .txt, .htm, .pdf • Qualitative and textual data: .xml, .txt, .json		(.py, .c, .h, .cpp, .hpp, .sh, .bat). Some configuration files may be stored as .json, .csv or .xml files or other human readable formats. Other data might be stored in text files such as .txt, .md. Documentation might be stored in .pdf or .html format. Some code might be stored in the form of compiled binary files.
Expected volume of data	**100 MB**	This depends on the experiment's length. **5 GB** is only a mere estimation of the entire accumulated recordings.	**100 MB**
Expected time of release	We will release datasets after publication of papers that reference it. In any case, data will be released one year after the end of the project, to ensure having time for publishing results on such data.	By the end of the project, only if the related papers are published.	We will release parts of the developed codebase in individual repositories upon publication of papers that reference it. We will also release code in cases where it would be beneficial for users of our other published datasets as a tool to facilitate a better use of the published datasets. We may postpone the release of some parts of the code until all of the papers relevant for a given study are finished.
DOCUMENTATION AND DATA QUALITY			
Metadata and documentation	Beyond giving as much information as possible in published papers to ensure reproducibility of the studies, metadata will be incorporated into the dataset to help others identify, discover, and reuse data. We will use the git versioning system, which keeps track of the metadata for the repository. When released, the code will be hosted in public git repositories in our institutional GitLab server. • Data will be divided in folders according to their nature and format. • Readme text and, if possible, graphic representation of the dataset will be released • Links to papers or additional materials explaining the procedure of the study will be provided	At least the following information will be provided: - data source and used methods - data format and units - when and where was the data collected As to the metadata standards, it is possible to use the recommendations of Australian National Data, UK Digital Curation Centre (DCC), or the Research Data Alliance (RDA). Please refer to https://www.fdm.uni-hamburg.de/en/fdm/metadaten.html for more information	We will use the git versioning system, which keeps track of the metadata for the repository. Codebase documentation will also be provided for the scientific community to be able to utilise the code for their own research. When released the code will be hosted in public git repositories either hosted on third-party services such as GitHub or on GitLab servers hosted by one of the participating institutions. Specific releases of code versions without the associated git metadata will be released and assigned a DOI.

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
Keywords	Human-Robot Interaction; Human-awareness; Trust; Explainability; Cognitive Robotics; Social Robotics. Once a new dataset will be created, specific keywords for that dataset will be provided. According to the type of dataset and research, tools such as MeSH on Demand will be used to ensure the findability of the dataset.	Pointing gestures, addressee estimation, intention understanding, HRI, object recognition	Human-Robot Interaction; Human-awareness; Trust; Explainability; Large Language Models; Memory; Gesture Recognition
Data quality	Codes will be stored in datasets only after accurate control of correct running, storing data and safety for human participants (robot's code) both in pilot studies and experiments.	The camera and mechanical parts of the robot will be calibrated before the experiment. The recording will involve researchers who will operate and interact with the robot and control the immediate environment of the robot.	Code will be verified by its execution on relevant hardware. Code segments may be reviewed by multiple people. Some of the functionality might be verified using unit tests.
STORAGE AND BACKUP DURING THE RESEARCH PROCESS			
Storage and backup solutions	Computer data and configuration will be backed-up on an encrypted external storage system at least every week and software installed on computers will be updated regularly. Data will be tested regularly to ensure their usability in case of necessity.	The data will be stored safely and securely on the premises of the University of Hamburg. The data is backed up on a regular basis. More information here: https://www.fdm.uni-hamburg.de/en/fdm/daten-speichern.html	The git versioning system supports the codebase to be stored both in remote storage (GitLab, GitHub) and locally on the devices used by researchers. The GitHub remote storage is hosted by a third-party which ensures data security. Specific releases of code versions without the associated git metadata will be released on platforms such as Zenodo.
Data security and protection	Data will be processed with a secure LAN network. Every system will be accessible only via authentication with credentials assigned to every person responsible for data processing. Passwords will consist of at least 8 characters and will be modified by the person on first use and at least every three months. All systems will be monitored. Data will be protected following the technical and organisational safety measures to ensure protection from unauthorised access, data theft and/or data leakage. For this reason, data will be encrypted to be unreadable in case of data breach.	The security and protection is cared for as part of the multi-step data backup at the department of informatics, University of Hamburg. More information here: https://www.inf.uni-hamburg.de/en/inst/irz/it-services/backup.html and https://www.fdm.uni-hamburg.de/en/fdm/daten-speichern.html	The codebase will not contain sensitive data therefore we will use standard practices. The private repositories will be accessible by researchers working on the project via their account protected with passwords or authorization tokens/keys. The public repositories will be visible to anyone, but only researchers working on the project will have the ability to modify them in the same manner as described for private repositories.

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT			
Personal/ special data categories	*NO*	*NO*	*NO*
Protection of personal/ special data categories	No human subjects' data will be included in this dataset.	No personal or sensitive data will be included in this dataset.	No personal or sensitive data will be included in this dataset.
Ethical issues issues (research data involving experiments with humans)	*NO* No human subjects' data will be included in this dataset.	*NO* No human subjects' data will be included in this dataset.	*NO* No human subjects' data will be included in this dataset.
Other ethical issues (e.g. involving animal subjects)	*NO* No data from animal subjects will be included in this dataset.	NO No data from animal subjects will be included in this dataset.	*NO* No data from animal subjects will be included in this dataset.
Intellectual property rights	As written in the Grant Agreement, results are owned by the beneficiaries that generate them. However, two or more beneficiaries own results jointly if: I) they have jointly generated them and II) it is not possible to: a) establish the respective contribution of each beneficiary, or b) separate them for the purpose of applying for, obtaining, or maintaining their protection. The joint owners must agree — in writing — on the allocation and terms of the exercise of their joint ownership ('joint ownership agreement'), to ensure compliance with their obligations under this Agreement. This dataset stores data that is owned by IIT. Part of them could be also owned by the other beneficiaries if they comply with the abovementioned criteria (I and II). Data will be deposited in the institutional repository and, as soon as possible in compliance with the	As written in the Grant Agreement, results are owned by the beneficiaries that generate them. However, two or more beneficiaries own results jointly if: I) they have jointly generated them and II) it is not possible to: a) establish the respective contribution of each beneficiary, or b) separate them for the purpose of applying for, obtaining, or maintaining their protection. The joint owners must agree — in writing — on the allocation and terms of the exercise of their joint ownership ('joint ownership agreement'), to ensure compliance with their obligations under this Agreement. This dataset stores data that is owned by UHAM. Part of them could be also owned by the other beneficiaries if they comply with the abovementioned criteria (I and II). Data will be deposited in the institutional repository and, as soon as possible in compliance	The repositories will be owned by a group on the GitLab server or on GitHub. The group will consist of members of the research team from all participating institutions who will have the ability to restrict or grant access to a given repository to others.

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
	FAIR principle "As Open as Possible, as Closed as Necessary", will be re-organized and deposited in Zenodo Data Repository together with the related Metadata. Restrictions might arise in case the sharing will be i) against the beneficiary's legitimate interests, including commercial exploitation, or ii) contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement. However, if open access is not provided (to some or all data), it will be specified and justified. Metadata of deposited data will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata will include persistent identifiers for related publications and other research outputs.	with the FAIR principle "As Open as Possible, as Closed as Necessary", will be re-organized and deposited in Zenodo Data Repository together with the related Metadata. Restrictions might arise in case the sharing will be i) against the beneficiary's legitimate interests, including commercial exploitation, or ii) contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement. However, if open access is not provided (to some or all data), it will be specified and justified. Metadata of deposited data will be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: datasets (description, date of deposit, author(s), venue and embargo); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata will include persistent identifiers for related publications and other research outputs.	
DATA SHARING AND LONG-TERM PRESERVATION			
Data sharing	Data (code) employed for an experimental study will be stored in the institutional repository of the beneficiaries of the data. Once the entire process of analysis and publication will be finalised, shareable data will be organised and shared in new datasets, comprehensive of sharable code and metadata, and linked with experimental data. So, after publication, such datasets will be	Data (code) employed for an experimental study will be stored in the institutional repository of the beneficiaries of the data. Once the entire process of analysis and publication will be finalised, shareable data will be organised and shared in new datasets, comprehensive of sharable code and metadata, and linked with experimental data. So, after publication, such datasets will be	The repositories which were available under the open science principles will remain public and be accessible to everyone. The released code will be accessible using git-based tools or via a browser by visiting a public url. Links to relevant repositories will be included in the published papers and in conference presentations. We will also share links to repositories on public platforms

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
	uploaded to a repository for Open Science, such as Zenodo, and will be provided with a persistent identifier (Digital Object Identifier) to ensure their findability and accessibility.	uploaded to a repository for Open Science, such as Zenodo, and will be provided with a persistent identifier (Digital Object Identifier) to ensure their findability and accessibility.	(e.g. PapersWithCode.com) and social media. The repositories will be findable using common search engines. Code might also be provided to third-parties upon request when necessary for peer-review or follow-up work even before it is made accessible to everyone. Specific releases of code versions without the associated git metadata will be released and assigned a DOI at repositories such as Zenodo.
Data repository	**Institutional GitLab platform that will be integrated with Institutional Data Repository (see below) or Zenodo, to link the code with a DOI**	**Institutional**	**Institutional**
	IIT Dataverse		**GitHub, Zenodo**
	https://dataverse.iit.it/ https://gitlab.iit.it		https://github.com
Restrictions on sharing	Data will be shared only once results of the experimental study they were used for will be published. During the period of access restriction, those who will be able to use the data are the beneficiaries, the other partners of the project according to the Grant Agreement, and other partners of the network if the data have been collected in a collaboration project and, in any case, in compliance with the Grant Agreement.	Data will be shared only once results of the experimental study they were used for will be published. During the period of access restriction, those who will be able to use the data are the beneficiaries, the other partners of the project according to the Grant Agreement, and other partners of the network if the data have been collected in a collaboration project and, in any case, in compliance with the Grant Agreement.	Anyone will be able to use the data shared in public repositories for non-profit research and educational activities.
Data curation			No data will be destroyed. Earlier versions of the codebased will be accessible via the git versioning system.
Requirements for reusability	As far as possible, data will be shared using non-proprietary formats to ensure reusability. Information about devices, tools, software, versions, and libraries to run codes will be provided to ensure reproducibility.	Since the stored data follows known formats, people who have access to it should be able to view it using any open-source software.	Use of the git versioning software is recommended to access shared code, but is not required. In order to run the code publicly available compilers and interpreters will be required along with publicly available libraries. Some parts of the code may require proprietary software such as Matlab.
Licensing	*GNU GPLv3* or *MIT License*, depending on the specific dataset we will release	*GNU GPLv3* or *MIT License* or *CC BY - Attribution*, depending on the specific dataset we will release	*CC BY - Attribution*

DATA SUMMARY	DATASET 4	DATASET 5	DATASET 6
DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES			
Roles and responsibilities	Once a new dataset will be created, roles and responsibilities for that specific dataset will be defined and indicated in updated versions of the DMP. Coordination of DMP will be assigned to the coordinator partner of the project, but responsibilities about datasets will depend on its beneficiary or beneficiaries. Regular updates of the DMP will be reported every time a new dataset will be needed.	Data management of the codebase will be the responsibility of Hassan Ali (hassan.ali@uni-hamburg.de)	Data management of the codebase will be the responsibility of Hassan Ali (hassan.ali@uni-hamburg.de)
Resourcing			

4. FAIR data

4.1. Making data findable, including provisions for metadata

In order to make each public dataset findable and citable, a Digital Object Identifier (DOI) was assigned to each dataset. Final datasets in the TERAIS project have been uploaded and publicly shared through ZENODO, IIT DATAVERSE, Github and GitLab, which provide DOIs to all publicly available uploads. The DOI of each dataset was added to the datasets' tables reported in Section 3.2 in all updates of the document.

To facilitate datasets' findability and reusability, file-naming conventions agreed among partners was used and clearly explained in associated "readme.txt" files. Filenames and their structure will follow conventions agreed upon by the project partners with the goal of making the purpose of individual files and their relations to each other clear.

Meaningful search keywords have been provided in the datasets' metadata to optimise the possibility for discovery and potential re-use. Search keywords were chosen according to a standard nomenclature or vocabulary, such as MeSH vocabulary or the ACM Computing Classification System (CCS).

Zenodo is compliant with the EU definition of "trusted repository". Regarding metadata, Zenodo allows extensive citation of metadata, including DOIs, authors, contributors, keywords, funding, related or alternate identifiers, and references to scientific articles or other types of publication. Zenodo is indexed in OpenAIRE Explore and registered in re3data.org and FAIRsharing.org. Zenodo is compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a widely used protocol for harvesting metadata. Following OAI-PMH, the available standard formats for metadata harvesting in Zenodo are DataCite and DublinCore.

UKBA institutional repository. During the pre-publication phase of confidential access, the datasets were hosted on the institutional local servers of UKBA, institutional OneDrive, institutional Google Drive, with protected access. After the publication, open access was created by publishing the descriptions of the data along with any relevant metadata on the official departmental website (https://dai.fmph.uniba.sk/w/Project:TERAIS/en), where the actual downloadable datasets were also linked. The website and the download links have been accessible publicly and could be found by search engines.

IIT Dataverse (https://dataverse.iit.it/) is the institutional research data repository of the Istituto Italiano di Tecnologia, for both preservation and sharing of research datasets. IIT Dataverse is compliant with the EU definition of "trusted repository". It is based on the Dataverse software, developed at Harvard University (www.dataverse.org). Dataverse assigns persistent DOIs to all data uploads for findability and it is accessible through standard HTTPS protocol. IIT Dataverse is indexed in OpenAIRE Explore and registered in re3data.org. Moreover, it provides APIs to search and access datasets, including a SWORD API. Each data upload includes 1) citation metadata, 2) optional and customizable domain-specific metadata (e.g., for Life Sciences), and 3) file-level metadata. Metadata can be exported in different standard formats (DataCite, OpenAIRE, JSON, JSON-LD, OAI, etc.) for maximal interoperability. Dataverse ensures reusability of datasets, by supporting open licences, like Creative Commons licences, and offers the possibility to customise specific data usage agreements. IIT Research Data Management service oversees dataset publication, by providing basic data curation to ensure dataset quality and FAIRness.

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338

Research Data Repository (RDR) (https://www.fdm.uni-hamburg.de/en/fdm.html) is a central service for data archiving at the University of Hamburg. It can be used by researchers to store data securely at several locations. It also provides other features, such as versioning and a unique Digital Object Identifier (DOI) for easier citation. The data should be in a final state before storing since the repository is not intended for daily work on the data.

GitHub and GitLab is an Internet hosting service for software development and version control using Git. It provides the distributed version control of Git plus access control, bug tracking, software feature requests, task management, continuous integration, and wikis for each project. GitHub is commonly used to share open-source software by various research communities. Projects can be found by using the search feature on GitHub or via other search engines. GitHub supports adding keywords for each repository and a markdown README file which is displayed on the title page for each project. GitLab is a software package enabling similar functionality to GitHub, but can be hosted on an institutional server.

4.2. Making data accessible

The project complies with the Open Science and Research Data Management requirements about openness and accessibility of research data, metadata, and other outputs resulting from HE grants, as detailed in the TERAIS Grant Agreement (art. 17), and described in the HE Annotated Model Grant Agreement (Annex 5) and the HE Programme Guide. Therefore, as early as possible, research data generated during the project, which include raw and processed data as described above in Section 3.2, have been made open, with an exception of datasets that supported unfinished peer-reviewed publications, patent applications, or information that cannot legally be made openly accessible (e.g. personal or sensitive data, following “as open as possible, as closed as necessary”). The last column of Table 1 summarises the accessibility level foreseen for each produced dataset. **Datasets that support peer-reviewed scientific articles have been made open at the time of publication with no exceptions. Datasets (data and metadata) have been accessible in trusted repositories (see 4.1) through standardised and freely accessible HTTPS protocol.**

Along the project duration, whenever possible, project documents and research data have been shared within the consortium through internal sharing service (e.g., cloud tool); user authentication has been mandated to keep confidentiality of data until required. Applied restrictions may include embargo, user authentication, or explicit acceptance of a custom data usage agreement. Restricted access has been required for specific datasets if the data sharing would negatively impact Technology Transfer and patenting. However, in such cases the DMP has been coherently updated to provide more comprehensive motivations. An embargo of one year after the end of the project will be applied to those datasets that will not be related to any publication, in order to ensure their beneficiaries publication. In case of personal data, data has been shared only after pseudonymization and the pseudonymization mapping scheme is available only by experimenters with user authentication.

Appropriate and comprehensive documentation (e.g., extensive and complete *readme.txt* files), together with relevant metadata (e.g., reported in a structured and machine-actionable format, such as .json), has been prepared and attached to the data before sharing. Metadata have followed DCMI, Datacite, DDI standards or similar vocabularies. Software codes needed to access / read / visualise the data have been made openly accessible through dedicated GitHub / GitLab repositories, which provide open access and long-term storage of source codes.

Noteworthy, **GitHub and GitLab repositories can be linked in Zenodo and in IIT Dataverse** and code releases can be assigned DOIs for findability and greater reproducibility. All metadata in Zenodo is licensed under Creative Commons Zero, while the

data files may be either open access and subject to a licence described in the metadata. Zenodo metadata contain references to related materials and tools (e.g., codes) by citing and linking DOIs. By default, data and metadata in Dataverse are licensed under Creative Commons Zero. Specific licences for data are clearly specified in the metadata (e.g., CC-BY instead of CC0). Dataverse metadata contains references to related materials and tools (e.g., codes) by citing and linking DOIs. Code released on GitHub or GitLab contains a licence file within the code repository (CC-BY or CC0). All **data stored in Zenodo** remain accessible for the lifetime of the repository, which is currently warrantied for a minimum of 20 years. Metadata will remain available also after data is no longer available.

Research data stored in IIT Dataverse will remain accessible online after the end of the project with no specific deadline and until it is required. Metadata will remain available also after data is no longer available or transferred to an offline storage location (e.g., tapes). **UKBA Institutional Repositories** will remain accessible online after the end of the project with no specific deadline and until it is required. Code stored in the locally hosted **GitLab servers at UKBA** will be available after the project ends with no specific deadline. Code available on GitHub will remain available unless the operator of GitHub changes their terms of service. The data stored in the **Research Data Repository (RDR) of Universität Hamburg** is guaranteed to be stored for a minimum of 10 years. It is also possible to store the data for longer periods if chosen by the researchers. After upload, the data will not be altered nor migrated to newer formats. Upon the end of the storage duration, an extension is possible and the researchers would be contacted before deletion. However, the metadata will always remain in the repository. More information about the terms of the licence use for RDR at UHAM can be found at: https://www.fdr.uni-hamburg.de/record/11057#.ZBsDs4DMIdw.

4.3. Making data interoperable

Raw data has been acquired in file formats associated with the instrument the data had been generated on, which allows for the data to be usually read out by anyone who is in possession of the dedicated acquisition software. However, to facilitate exchange and re-use of data, datasets have been exported and stored in formats that are commonly accessible (e.g., .txt, .csv, .json), including the associated metadata and additional comments and descriptive text to aid the interpretation of the data.

To ensure the interoperability of data and metadata, standard or community-endorsed vocabularies (DCMI, Datacite, DDI standards or similar vocabularies) have been used when applicable. In case new vocabularies are generated or uncommon vocabularies are used, mappings between custom and community-endorsed vocabularies have been associated to the dataset for interdisciplinary interoperability. Abbreviations, codes, and variables' names have always been clarified at first use or defined in "readme.txt" files.

4.4. Increase data re-use

As mandated in art. 17 of the Grant Agreement, the digital (or physical) access to the results needed to validate the conclusions of scientific publications has been provided, including access to all the information about the research outputs/tools/instruments needed to validate publications and enable the re-use of data. The access to research data has been provided through deposition in trusted repositories, as detailed in the previous paragraphs. Whenever possible, research datasets have been available under the **Creative Commons Attribution 4.0 International (CC BY 4.0)** licence, allowing third parties to share and adapt data with no restrictions as long as attribution (i.e., citation) has been provided. All Consortium partners have ensured that proper licences were attached to the deposited data to define all conditions under which the work had been provided and could be reused, also in the case when the data have not been released under an open licence.

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338

Along with data, any information about methods, protocols, models, software, algorithms, workflows, simulations, electronic lab notebooks' records, etc. have be also been provided and linked in the repository through the use of persistent identifiers (e.g., DOIs) (see also "Other research outputs"). For instance, source codes have been version-controlled, deposited in GitHub / GitLab and made available in open access with suitable licences (GNU GPL, MIT, etc.). Whenever useful, the information about data cleaning, data quality assurance procedures, methodology, as well as variables' definitions, units of measurements, software dependencies, and, in general, data structures have been included in "readme.txt" or "readme.md" files, deposited together with data.

For data quality assurance, experiments have been conducted based on recent literature and previous pilot studies exploring problems in experimental settings, data collection, and early data analysis. When possible, video or image data has been recorded by more cameras (e.g., external cameras, robot's cameras). Tools for data collection, such as to collect human gaze and kinematic data, were calibrated before recording data. We have followed standard and documented methods for data collection and analysis and, in case of novel methodologies, we have also linked the analysis with traditional ones.

5. Other outputs

In addition to the management of data, the TERAIS project partners also considered and planned for the management of other outputs that have been generated or re-used throughout their projects. **The TERAIS project, as the Horizon Europe coordination and support action aimed at building capacities of a university from a Widening country, involves a number of specific actions (e.g., conferences, workshops, webinars) and qualitative data collection.** Therefore, it was necessary to consider management, protection and use of data in this context as well. Moreover, the questions pertaining to FAIR data above were also applied to the management of these project outputs. Table 5 provides an overview of the details on how these other outputs have been managed and shared, or made available for reuse, in line with the FAIR principles.

Table 5. *Summary of other TERAIS project outputs with implications for data management*

OUTPUT NAME	PARTNER	ORIGIN	TYPE	ACCESS TYPE	WP & TASK	ACCESSIBILITY[2]	DATA RETENTION
Output 1: **Documentation on the participants in the events organised by UKBA**	UKBA	Generated by the project	**Document** (.docx, .xlsx, .pdf) **Digital image data** (.tif, .jpg, .pdf, .bmp, .png) **Digital audiovisual data** (wav, .mp3, .mp4, .ogv)	**PHYSICAL** (list of participants for in person events) or **DIGITAL** (list of participants for online events, pictures/ videos from the events)	**WP2:** Tasks 2.1-2.3 **WP7:** Tasks 7.1-7.3	**Confidential to Consortium including EC services** (Consortium only)	Regular period for keeping the records will be **5 years after the final payment** - in line with section 6 of the Data Sheet in the Grant Contract
Output 2: **Documentation on the participants in the events organised by IIT**	IIT	Generated by the project	**Document** (.docx, .xlsx, .pdf) **Digital image data** (.tif, .jpg, .pdf, .bmp, .png) **Digital audiovisual data** (wav, .mp3, .mp4, .ogv)	**PHYSICAL** (list of participants for in person events) or **DIGITAL** (list of participants for online events, pictures/ videos from the events)	**WP2:** Tasks 2.1-2.3 **WP7:** Tasks 7.1-7.3	**Confidential to Consortium including EC services** (Consortium only)	Regular period for keeping the records will be **5 years after the final payment** - in line with section 6 of the Data Sheet in the Grant Contract
Output 3: **Documentation on the participants in the events organised by UHAM**	UHAM	Generated by the project	**Document** (.docx, .xlsx, .pdf) **Digital image data** (.tif, .jpg, .pdf, .bmp, .png) **Digital audiovisual data** (wav, .mp3, .mp4, .ogv)	**PHYSICAL** (list of participants for in person events) or **DIGITAL** (list of participants for online events, pictures/ videos from the events)	**WP2:** Tasks 2.1-2.3 **WP7:** Tasks 7.1-7.3	**Confidential to Consortium including EC services** (Consortium only)	Regular period for keeping the records will be **5 years after the final payment** - in line with section 6 of the Data Sheet in the Grant Contract
Output 4: **Recordings of the focus group discussions with UKBA staff and students**	UKBA	Generated by the project	**Digital audio** (.waw, .mp3)	**DIGITAL** (audio recordings)	**WP4:** Task 4.1 **WP5:** Task 5.1	**Beneficiary institution access**	The data will be permanently deleted from the institutional Google drive after submission of deliverables D4.1 and D5.1 in **04/2023**

[2] *In case of patenting or potential commercial exploitation, different levels of confidentiality can be foreseen (in case of Collaborative Projects):*

- ***Beneficiary institution access:*** *The disclosure of information is not provided at any level and/or any time. This option is applied when, regardless of the long-term value and scope for wider use, the dataset contains data that would lose their value if disclosed (e.g. experimental results and validation). The beneficiary who chooses to protect this information from any external access aims at patenting/exploiting it or at protecting trade secrets and must ensure confidentiality beyond the clauses agreed in the Consortium Agreement;*
- ***Confidential to partner:*** *This option is applied when, regardless of the long-term value and scope for wider use, the dataset contains personal data that cannot be protected once disclosed. These may include among others videos and images collected during the project tests;*
- ***Confidential to Consortium including EC services (Consortium only):*** *This option is applied for data containing confidential information (e.g. exploitable results) requiring IP protection, aimed at possible exploitation. In certain cases to be defined along with the development of the project, the owners may allow Open Access upon dissemination of the associated results in peer-reviewed scientific publications;*
- ***Open Access:*** *This option is applied when data has no IP restrictions and will be openly available and re-usable.*

6. Allocation of resources

At this concluding stage of the project implementation, the costs foreseen for data management have been related to:

- the working time needed to set up and perform the data collection, including synchronisation of devices and analysis activities;
- the working time required to setup local and shared data collection devices/servers;
- the working time needed to write documentation, metadata, etc.

In particular, all TERAIS project partners indicated that one person month for each project year has been dedicated specifically to make the data FAIR and for data management. Expenses for cloud storage used to share data and documents within the consortium have been covered by host institutions. The repository chosen for long-term data sharing and preservation (e.g., Zenodo) offers free data archiving up to 50 GB *per* dataset, which is enough given the expected data volume.

The project coordinator (UKBA) is in charge of the DMP from both the scientific and technical perspectives, as well as the release of the first version and the regular updates – in close collaboration with the partners (IIT and UHAM) who have responsibility for scientific and networking lead. The partners responsible for the generation of the research data in WP3 have also been in charge of the validation and registration of datasets and metadata, as well as backing up of the data for sharing through open access repositories, while quality control of these data has been the responsibility of the relevant WP leader, supported by the Project Coordinator. Each partner should respect the policies set out in this DMP.

7. Data security

According to the information security risk level associated with the datasets (i.e., low and/or medium), **only IT assets approved for the corresponding risk level have been used, as detailed in the Information Security policy of each partner institution.** Laptops, workstations, and servers have been managed by the institutional ICT Service departments and all procedures in place for storage and backup comply with the institutional Information security policy of the project partners (UKBA, IIT, UHAM). Where necessary, authentication with institutional login was required to protect data confidentiality and integrity.

UKBA has an Information Security policy which summarises common principles of IT Security at Comenius University Bratislava as well as the user's obligation. The policy defines security as ensuring the required availability, confidentiality and integrity of information and data, but also hardware and software. Assets are all classified into four security levels (protected, standard, unprotected, and special) based on the extent of the impact of a possible security breach. Guidelines, recommendations, and processes are specified for the classification of assets, for maintaining security in general, and for responding to security incidents. The policy has been published as the UKBA Internal regulation nr. 12/2009, and the Information security guidelines detailing its implementation can be found in the Internal regulation nr. 2/2011.

IIT has an Information Security policy which classifies information and data into three levels of risk (High, Medium, Low) in relation to the protection of Confidentiality, Integrity and Availability (Data Risk classification). Consequently, the policy indicates which ICT services and systems are approved for the different levels of data risk, according to the required

security standards. Therefore, to decide which ICT systems can be safely used in IIT to store, backup, and share research data, scientists should first of all determine which level of risk their data refer to. According to the policy, data are classified at high-risk if at least one of the following conditions are true: (1) The loss of Confidentiality, Integrity or Availability of the information can have a significant adverse impact on Intellectual Property or on the mission or reputation of the Foundation, or a significant economic or legal impact; (2) The protection of the information is required by Law; (3) The protection of the information is required by contracts or certifications of national or international standards (such as the ISO standards).

UHAM has an Information Security policy which is based on three main pillars: availability, integrity and confidentiality. Availability refers to having the data accessible and usable for authorised users at the desired times. Integrity means that the information needs to be complete and correct at any given time. Confidentiality is having the information in a form that is not accessible or disclosed to unauthorised third parties. More information about the data protection and information security policy that governs research data at the University of Hamburg can be found here: https://www.fdm.uni-hamburg.de/en/fdm/datenschutz.html.

The project is not expected to generate high-risk data. The major expected risk is connected with collecting personal data that, however, have undergone a process of pseudonymization. The pseudonymization mapping scheme is not part of shared datasets and is accessible only to experimenters through the authentication process. In case any dataset with high-risk data was planned, the DMP has been coherently updated.

8. Legal and ethics

8.1. Protection of personal/special categories of data

The TERAIS project consortium fully complies with the data protection principles of lawfulness, fairness and transparency in data processing, as well as purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and fully agrees that the protection of personal data is a priority, especially in a project comprising organisation of a number of events (online/offline) as well as involving human participants.

The personal data collected in the documentation in relation to the organised events **has been fully in compliance with the provisions of the General Data Protection Regulation (GDPR)**, article 6 par. 1 point a), personal data of participants has been processed only after receiving their signed informed consent.

For the purpose of experiments involving human participants, the **TERAIS project consortium** defined **data protection roles and responsibilities** (Data Controllers, Joint Controllers, Data Processors) and provided information and informed consent for data processing. In particular, researchers have requested informed consent to disseminate such data for scientific and educational purposes in scientific publications or conferences. The experimental data that include personal data such as human faces will never be used or made public in open access environments unless each concerned participant has given a specific explicit approval for that in a dedicated optional opt-in rubric of the informed consent form.

Furthermore, as a general rule, researchers and representatives of all partners have strived to anonymize personal research data before making them openly available, thus fulfilling both the open research data and data protection rules. Nevertheless, complete anonymization may not always be possible, especially in case of raw data such as images and videos. In these cases, researchers and representatives of the project partners

have proceeded with pseudonymisation of the data (when possible) and carried out technical and organisational security measures to ensure the data confidentiality. The measures to put in place have been data/database encryption, physical security measures to prevent unauthorised data accesses, etc. Finally, all the tasks involving data processing have been documented by researchers and representatives of the project partners, to guarantee an adequate level of transparency. More detailed information on the ethics-related considerations is provided in the **Ethics Reports** (updated deliverables D1.2 and D1.3 submitted in June 2024).

8.2 Ethical issues

8.2.1 Ethical issues (research data involving experiments with humans)

The TERAIS project consortium acknowledges that the governance of the ethical issues in projects involving experiments with humans had to be taken into great account. In particular, the following issues have been considered as priorities:

- Formalising procedures to identify/recruit all kinds of research participants (including children and/or minors in specific cases).
- Providing information and informed consent to the research participants (including all the relevant information regarding the protection of personal data as well as information related to the incidental/unexpected findings policy).
- Adopting adequate technical and organisational measures to safeguard the rights and freedoms of the data subjects/research participants (e.g. data processing, storage and retention, access procedures to the data, written instructions to the authorised persons to process personal data, security measures such as anonymisation, encryption or pseudonymisation etc.).

Furthermore, in the case of children and/or minors participation in the project, the TERAIS project consortium (namely IIT) has provided a governance enforcement, with reference to:

- Requiring a specific informed consent/assent for the children involved in the project, and defining the manner how parental/legal carer consent has been ensured for the participation of the infants and minors in the project.
- Verifying the necessity to obtain a prior clearance from the relevant authorities according to the national law before starting any research involving children.

Despite that, there wasn't any involvement of children nor minors in the experiment of the projects.

8.2.2 Ethical issues (research data involving experiments with animals)

No experiments with animals have been conducted.

9. Other issues

No use of other national/funder/sectorial/departmental procedures for data management is foreseen in the TERAIS project.

10. Abbreviations & keywords

Table 6. *Abbreviations and keywords.*

Abbreviation/ Keyword	Definition
AI	Artificial Intelligence
API	Application Programming Interface
CC BY 4.0	Creative Commons Attribution 4.0 International licence
CCS	Computing Classification System
DCC	Digital Curation Centre
DCMI	The Dublin Core Metadata Initiative
DDI	The Data Documentation Initiative
DMP	Data Management Plan
DOI	Digital Object Identifier
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DQF	Dataset Questionnaire Form
EC	European Commission
FAIR	Findable, Accessible, Interoperable, and Reusable data
GDPR	General Data Protection Regulation
HE	Horizon Europe
HRI	Human Robot Interaction
HTTPS	Hypertext Transfer Protocol Secure
ICT	Information and Communication Technology
IIT	Italian Institute of Technology
IP	Intellectual Property
IPR	Intellectual Property Rights
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
RAID	Redundant Array of Inexpensive Disks
RDA	Research Data Alliance
RDR	Research Data Repository
SWORD	Simple Web-service Offering Repository Deposit
TERAIS	Towards Excellent Robotics and Artificial Intelligence at a Slovak university
UKBA	Comenius University Bratislava (Univerzita Komenského v Bratislave)
UHAM	University of Hamburg
WP	Work Package (in a project)

TERAIS has received funding from the European Union's Horizon Europe Research and Innovation Funding Programme under Grant Agreement No. 101079338