PRESENCE

Funded by the European Union

A toolset for hyper-realistic and XR-based human-human and human-machine interactions, PRESENCE

Grant Agreement nº 101135025

HE Call identifier: HORIZON-CL4-2023-HUMAN-01-CNECT
Topic: HORIZON-CL4-2023-HUMAN-01-21

Type of action: HORIZON Research and Innovation Actions

D7.4: Ethics Framework and Data Management Plan II

DISSEMINATION LEVEL

☒	**PU**	**Public**
☐	SEN*	Confidential, only for members of the consortium (including the Commission Services)

Grant Agreement nº:	Project Acronym:	Project title:
101135025	PRESENCE	A toolset for hyper-realistic and XR-based human-human and human-machine interactions

Lead Beneficiary:	Document version:
UHAM	V1.0

Work package:
WP7 – Project Management and Coordination

Deliverable title:
D7.4 Ethics Framework and Data Management Plan II

Start date of the project:	Contractual delivery date:	Actual delivery date:
1st of January 2024	30th of June 2025	30th of June 2025

Editor(s):
Fariba Mostajeran (UHAM)

LIST OF CONTRIBUTORS

PARTNER	CONTRIBUTOR
UHAM	Fariba Mostajeran, Frank Steinicke

LIST OF REVIEWERS

PARTNER	REVIEWER/S
EEA	Joana Porcel

CHANGE HISTORY

VERSION	DATE	PARTNERS	DESCRIPTION/COMMENTS
V0.1	28-04-2024	UHAM	Content transferred from D7.2
V0.2	06-05-2024	UHAM	Draft, data instruction summary sections for partners contributions
V1.0	25-06-2024	UHAM, i2CAT	Format review, addition of the EEA signed report; submitted through the Participant Portal

Executive summary

This document describes the second version of the Ethics Framework and Data Management Plan (DMP) of the PRESENCE project. The DMP provides a common framework to respect the rights and dignity of human participants and encourages adherence to the FAIR data principles for all datasets collected, produced, and processed within the project for management and research purposes. It defines procedures that all project partners have to follow to protect the ethical rights of involved human participants, implement a dignified and safe treatment of participants, preserve their privacy, and ensure trustworthy and secure handling of their personal data.

It also completes the response of the PRESENCE project to the Ethics Requirement OEI nº 1, outlining PRESENCE's compliance with the received recommendation on the obligation to designate an Ethics Advisor.

The current version of this document contains definitions of the ethical and legal background as well as an analysis of the main elements of the data management policy used within the project. This is a living document, and it will face possible changes due to the design of the research activities.

Table of contents

1. Introduction 6
1.1. Purpose of this document 6
1.2. Scope of this document 6
1.3. Status of this document 6
1.4. Relation with other activities in the PRESENCE project 6
2. Data Summary 7
2.1. Summary of generated, collected, or re-used data per partner 7
2.1.1. Data Summary of i2CAT 7
2.1.2. Data Summary of ACTRO 8
2.1.3. Data Summary of UHAM 9
2.1.4. Data Summary of CERTH 10
2.1.5. Data Summary of RAYTRIX 11
2.1.6. Data Summary of SG 11
2.1.7. Data Summary of INTER 12
2.1.8. Data Summary of DIDIMO 12
2.1.9. Data Summary of VECTION 13
2.1.10. Data Summary of UB 13
2.1.11. Data Summary of Capgemini 14
2.1.12. Data Summary of SOUND 14
2.1.13. Data Summary of IMEC 15
2.1.14. Data Summary of JRS 15
2.1.15. Data Summary of SVR 16
2.1.16. Data Summary of ZAUBAR 17
2.1.17. Data Summary of ARTANIM 17
2.2. Summary of generated, collected, or re-used data per study 17
3. FAIR Data Guidelines 19
3.1. Making data findable, including provisions for metadata 20
3.2. Making data accessible 21
3.3. Making data interoperable 22
3.4. Increase data re-usability 22
4. Other Research Outputs 23
4.1. Academic Publications 23
4.2. Software 23

5. Allocation of Resources 23
6. Data Security 24
6.1. Personal Data Security 24
6.2. Logical Data Security 24
6.3. Physical Data Security 25
7. Intellectual Property Rights 25
8. Ethics 26
8.1. Recruitment, information, and consent for data collection 27
8.2. Handling of personal data 28
8.3. Handling of incidental findings 28
9. Other Issues 29
10. Abbreviations and Definitions 29
10.1. Abbreviations 29
10.2. Definitions 30
11. References 30
Annex I: Report of the External Ethics Advisor 31

List of Tables

Table 1: Data summary provided by i2CAT 8
Table 2: Data summary provided by ACTRO 9
Table 3: Data summary provided by UHAM 10
Table 4: Data summary provided by CERTH 11
Table 5: Data summary provided by RAYTRIX 11
Table 6: Data summary provided by SG 12
Table 7: Data summary provided by DIDIMO 13
Table 8: Data summary provided by UB 13
Table 9: Data summary provided by SOUND 15
Table 10: Data summary provided by IMEC 15
Table 11: Data summary provided by JRS 16
Table 12: Data summary provided by SVR 17
Table 13: Data summary provided by ARTANIM 17
Table 14: Summary of user studies 19
Table 15: List of PRESENCE partners with REC 27

1. Introduction

1.1. Purpose of this document

This document serves as the updated iteration of the Ethics Framework and Data Management Plan (DMP) for the PRESENCE research project. It includes definitions of the ethical and legal background and offers guidance on data management policies within the scope of the project. The DMP is based on the project's Grant and Consortium Agreement (GA/CA), adheres to the FAIR (findable, accessible, interoperable, reusable) data principles, and complies with the General Data Protection Regulation (GDPR) of the European Union (EU) 2016/679 [Ref. 1].

1.2. Scope of this document

The DMP covers mainly the following aspects:

- Presenting a comprehensive overview of the data that the PRESENCE project partners will generate, collect, or re-use during the project period.
- Providing guidelines for handling data in compliance with the FAIR data principles, ensuring that the data remains findable, accessible, interoperable, and reusable.
- Establishing guidelines for handling potential Intellectual Property Rights (IPR) on data that have been or will be generated, collected, or re-used during the project period.
- Implementing strict policies for promoting physical and logical data security, especially regarding personal data governed by the GDPR.
- Providing an ethical framework for all data collection activities involving human participants, emphasising the principles of respect, privacy, and informed consent.

1.3. Status of this document

This document presents a second version of the DMP delivered as D7.4 in month 18 of the project. It comprises all information on the origin, nature, and management of data as well as ethical foundations known or defined up to this stage of the project. The DMP is a living document that will become more detailed in upcoming versions and face possible changes due to the design and execution of the planned research activities. The updates will be published in D7.6 in month 36 of the project.

1.4. Relation with other activities in the PRESENCE project

This document guides all activities that involve the generation, collection, and re-use of data within the project and therefore, relates to all WPs. Particularly, it gathers information on data collected during the user studies performed within the evaluation activities described in T1.3, T1.4, T5.3, and T5.4 rooted in WP1 and WP5. It also relates to all other research and dissemination activities of any other WP that involve the generation, collection, and re-use of data.

2. Data Summary

The following sections offer an overview of the data that the PRESENCE project partners have generated, collected, or re-used throughout the project's duration. They provide a summary of user and technical studies that will be or have been conducted within the scope of the project to keep track of all collected data.

2.1. Summary of generated, collected, or re-used data per partner

This section of the DMP presents a table for each PRESENCE partner summarising information about the type and content of the data that each partner plans to generate, collect, or re-use during the course of the project. At this early stage of the project, the reported information will be collected and processed per partner to determine the general specifications and structures of the data and metadata that will be managed within the DMP. In future versions of the DMP, a summarised version of the generated, collected, and re-used data will be provided.

The data summary tables contain the following information:

- The purpose of the data and their relations to the objectives of the project
- Re-using the existing data and its purpose
- Data types and formats
- Size of the data
- Origin or provenance of the data
- Use of the collected or generated data outside the project

2.1.1. Data Summary of i2CAT

Expected purpose of the data and their relations to the objectives of the project:
i2CAT has gathered a) RGBD video data, extracted from lightfield cameras provided by RAYTRIX and from other types of cameras (i.e. Kinect4Azure, Orbecc FemtoBolt), and b) system performance data from different active sessions of the Holoportation SDK being implemented in WP2. Both data types are being used in the frame of WP2 to 1) research in digital human reconstruction; 2) research in volumetric video compression; and 3) research on the scalability of holographic communications. So far, anonymized user opinions in the form of questionnaires have been also gathered to analyse the subjective perception of end-users when applying different compression parameters.
Expected re-use of existing data and its purpose:
No previous data is expected to be used from i2CAT in PRESENCE.
Expected data types and formats:
RGBD data (i.e., videos with RGB and depth textures obtained from cameras). Point cloud data (i.e., PLY or proprietary formats obtained after calibration and reconstruction processes). QoS data (i.e., gathering different parameters generated by the real-time volumetric video pipelines and related media processing components).
Expected size of the data:

Our multi-camera capture pipeline has recently been optimized to reduce data recording sizes. At the maximum resolution of 1920×1080, each RGB-D frame from a single camera occupies approximately 1.4 MB. We support configurations ranging from 3 to 9 cameras, operating at 30 frames per second. This results in an estimated data rate of 2.5 GB per camera per minute. Consequently, recording one minute of data with a 9-camera setup yields approximately 22.5 GB of data. On the other hand QoS data and other metrics can be neglected in terms of data size. Also, our multi-camera capture process enables direct server-side recording, which reduces data volume by pre-encoding each frame using H.264 or H.265. At the maximum resolution of 1920x1080, operating at 30 frames per second with a GOP of 15, the estimated data rate is 0.3 GB per camera per minute. Therefore, a 9-camera configuration generates approximately 2.7 GB of data per minute. For enhanced Quality of Service (QoS) and granular data management during server-side recordings, each frame is meticulously stored across three distinct files: one for metadata, another for depth information, and a third for color data. Additionally, the relevant audio track associated with the user is also preserved. All these files are securely stored on the server, allowing for direct transmission of these representations from the server. Furthermore, the activation and deactivation of these server-side recordings can be precisely controlled via a REST API. This capability provides maximum optimization, allowing us to record exactly what's needed, precisely when it's needed. At this point of the project, the size of the data is unknown and will depend on the amount of user bodies and the length of the associated video recordings. RGBD and PLY data representing dynamic scenes (i.e., moving bodies) will be recorded and the associated dataset size will be, most likely, between 1 GB and 10 GB. QoS data consists of basic parameters. A specific dataset for that purpose is not intended since it is related to the internal development of WP2 tools (i.e., the data has value in terms of research publication, but a potential dataset would not be usable by other research teams inside or outside the PRESENCE consortium).
Expected origin/provenance of the data:
The origin of the data is either Raytrix or other depth cameras or the result of the reconstruction processes implemented in WP2.
Expected use of the data outside the project:
Researchers and institutions working in the digital human reconstruction domain or the volumetric video compression domain.

Table 1: Data summary provided by i2CAT

2.1.2.Data Summary of ACTRO[1]

Expected purpose of the data and their relations to the objectives of the project:
Data coming from ACTRO can be used for targeting purposes (e.g., mapping of a skeleton model of the user for haptic interaction rendering) but no data will be collected or stored anywhere.
Expected re-use of existing data and its purpose:

[1] The partner (currently in a termination process) has not provided update. The information is the one in D7.2

ACTRO has no intention of re-using existing data.
Expected data types and formats:
The file format is not yet defined, most probably it will be a human-readable file format like JSON or generic TXT. Data will only be used during runtime.
Expected size of the data:
No re-usable data will be stored.
Expected origin/provenance of the data:
Data is coming from hardware devices, but no re-usable data will be stored.
Expected use of the data outside the project:
No re-usable data will be stored.

Table 2: Data summary provided by ACTRO

2.1.3.Data Summary of UHAM

Expected purpose of the data and their relations to the objectives of the project:
UHAM has generated data by conducting user studies. The data have been used to evaluate the technical solutions and user experience aspects of WP4 and consist of participant information and VR-related psychometric measures (e.g., realism, co-presence, and plausibility of intelligent virtual agents). Hence, the data include demographics (e.g., age, gender, experience), subjective questionnaires (e.g., IPQ, SUS, TAM2), and objective interaction performance measures (e.g., required time to complete tasks). In addition, it might be necessary to perform audio or video recordings from interactions during user studies used for post-hoc analysis. While interacting with the intelligent virtual agents (IVAs) we will also record the conversation that the users have with the LLMs. The latency and token consumption data by AI services have been collected. During the software development process realizing the technical solutions of WP4, text-based scripts, audio files, and 3D data have been generated and re-used. In addition, some of the generated data have been used for producing different deliverables or research outputs like scientific publications.
Expected re-use of existing data and its purpose:
Priorly collected experimental data have been re-used for producing different deliverables or research outputs like scientific publications.3D models, plug-ins, and assets have been re-used for different software developments.
Expected data types and formats:
We have and plan to continue collecting or re-using data in the following formats: Text: TXT, RTF, MD, DOCX, TEX Tables: CSV JSON, XLSX Graphics: PNG, JPG, SVG, EPS 3D Data: FBX, OBJ Audio: WAV, MP3 Video: MP4 Others: PDF, CS

Expected size of the data:
Collected experimental data is usually text-based and will not exceed a total size of 1GB. For the purpose of recording experiments, it might also be necessary to record audio, video, or movement data usually leading to a larger amount of data with a total size of up to 1 TB. Software development repositories containing 3D data can grow to considerable sizes of up to a total of 1 TB when containing sample applications.
Expected origin/provenance of the data:
Regardless of its purpose, the generated or re-used data will originate from several sources, which include internal or external preexisting data, data from the scientific literature, public data repositories, and data collected during user studies.
Expected use of the data outside the project:
Our generated anonymized data originating from user studies might be useful to other researchers also performing user studies related to our domain or beyond. It can serve as comparative or benchmark data and increase the reproducibility of our results. Software developed within WP4 that will be published might be useful for other researchers working in the domain of intelligent virtual humans or could be of interest to any persons integrating intelligent virtual humans into their applications.

Table 3: Data summary provided by UHAM

2.1.4.Data Summary of CERTH

Expected purpose of the data and their relations to the objectives of the project:
CERTH has been generating data representing the physical appearance of human subjects as volumetric data. The data were used to accurately represent humans in an XR environment. Moreover calibration related datasets are being produced for the proper spatial alignment of the light-field cameras During the ongoing development of the solutions, data are generated and re-used to further improve and evaluate the tools.
Expected re-use of existing data and its purpose:
Existing public datasets are being used to develop the volumetric capturing and calibration modules.
Expected data types and formats:
Captured image frames: PNG, JPG, PGM Generated volumetric data: OBJ, PLY, GLTF Calibration data: JSON
Expected size of the data:
Public datasets that are re-used are sized between 700 MB and 5 TB. Generated datasets are mostly comprised of RGBD images and are typically 1-2 GB for HD resolutions.
Expected origin/provenance of the data:
Data from public datasets as well as data generated within the consortium for evaluation, and finally data generated by the users of the developed tool.
Expected use of the data outside the project:

Some of the datasets which are produced during the development of the camera calibration and volumetric capturing modules would be made public and potentially be used by the research community to further extend the knowledge within the field of human 3D reconstruction and multi-view spatial calibration.

Table 4: Data summary provided by CERTH

2.1.5.Data Summary of RAYTRIX

Expected purpose of the data and their relations to the objectives of the project:
RAYTRIX will capture data to develop and demonstrate the volumetric capture of humans exhibiting natural movements at high resolution and low latencies for an immersive experience. This data will be captured at various stages of processing including 1) light-field data from the cameras, 2) RGBD/RGBXYZ video generated from the light-field data, and 3) volumetric video generated from the RGBD videos. Additionally, temperature data is collected by the image sensor to estimate and compensate for the influence of thermal expansion on the optical system and calibration data is collected during manufacturing and provided with the camera. This includes the specifications of the image sensor and microlens array, which are required to convert light fields into RGBD data.
Expected re-use of existing data and its purpose:
Prior data on the effects of temperature and thermal expansion is used for thermal compensation.
Expected data types and formats:
Light field raw image in RAY format and light field raw video in RAY format. The SRGBD/RGBDXYZ format needs to be determined.
Expected size of the data:
We require 12,5 Gbps per camera in uncompressed format.
Expected origin/provenance of the data:
Volumetric capture setups are located at test facilities with i2CAT in Barcelona, Spain, CERTH in Thessaloniki, Greece, and Raytrix in Kiel, Germany.
Expected use of the data outside the project:
The use of data is mostly internal. However, the partners involved in the Horizon Europe PRESENCE and 6GXR projects have reached an agreement to exchange data and findings. This collaboration aims to minimize redundant development efforts and enhance the overall quality of the projects. Datasets for dissemination will be created separately and of interest to any researcher working in the domain.

Table 5: Data summary provided by RAYTRIX

2.1.6.Data Summary of SG

Expected purpose of the data and their relations to the objectives of the project:
We will create hand movement data of users during the user tests and use this data to map over a skeleton model of the user to render haptic interactions. This data will not be stored after the session unless the session is captured by a partner performing the user tests, who will share the data only for internal use.
Expected re-use of existing data and its purpose:

SG has no intention of re-using existing data.
Expected data types and formats:
The file format is not yet defined, most probably it will be a human-readable file format like JSON or generic TXT. Data will only be used during runtime and will not be stored permanently as explained above.
Expected size of the data:
The expected size of the data shared internally will be less than 1 GB.
Expected origin/provenance of the data:
The captured sensor information is provided by the SG hardware devices.
Expected use of the data outside the project:
No data will be stored longer than the runtime of a user test, except when the data is shared internally.

Table 6: Data summary provided by SG

2.1.7.Data Summary of INTER

As it was declared in D7.2, partner INTER indicates that they currently have no intention to generate, collect, or re-use data within the PRESENCE project.

2.1.8.Data Summary of DIDIMO

Expected purpose of the data and their relations to the objectives of the project:
DIDIMO has two use cases. In the first use case, when avatars are randomly generated, no data needs to be collected. In the second case, if avatars represent a person (e.g., generated from a selfie), then we will collect the photograph in our API. The data will be used to accurately represent humans. During the development of the project's solutions, data will be generated and re-used to further improve and evaluate the tools. If our API is used in order to generate avatars from a selfie, then it will be governed by our terms and conditions [Ref. 2].
Expected re-use of existing data and its purpose:
We have internal training datasets that we use to train our machine-learning models.
Expected data types and formats:
Captured image frames in PNG and JPG format.
Expected size of the data:
Images tend to be a few MB, so it will depend on the number of avatars generated from images.
Expected origin/provenance of the data:
Preexisting data originates from an own private dataset, an open research dataset (non-EU), an open research dataset (EU), and other files under IPR. Data generated during the project will originate within the

consortium during evaluation and ultimately by the users of the developed tool.
Expected use of the data outside the project:
The possible dataset produced during the project could be used by the research community to further the knowledge within the field of 3D human reconstruction.

Table 7: Data summary provided by DIDIMO

2.1.9. Data Summary of VECTION

Same than declared in D7.2, the partner VECTION indicates that they currently have no intention to generate, collect, or re-use data within the PRESENCE project.

2.1.10. Data Summary of UB

Expected purpose of the data and their relations to the objectives of the project:
The data will be generated solely for the purposes of the project. It will be used to assess presence in relation to the configuration of technical developments within the project.
Expected re-use of existing data and its purpose:
No data will be re-used.
Expected data types and formats:
The data will consist of questionnaire data (demographics and responses to the virtual environments). It will also consist of sequences of choices made by participants by selecting different configurations of the environment that they are in, to optimize their illusion of presence. Multi-subject EEG data (2-4 persons) will be acquired using multi-channel (32 electrodes) recording. This data will be associated with the VR scenarios created for the evaluation of the illusion of presence and correlated with behavioural and other measures (eye-tracking and pupillometry).
Expected size of the data:
The size of the data will be relatively small consisting of spreadsheets with approximately 50 rows once, one for each participant. For EEG recording, a large amount of data (estimated 1TB) will be acquired. The size of the recordings depends on the number of subjects x number of electrodes x time (minutes recorded) x digitalization rate (500 Hz). Depending on the timing, we plan to record in between 30 to 60 minutes.
Expected origin/provenance of the data:
The origin of the data will be experimental studies carried out during the project.
Expected use of the data outside the project:
The data itself might be useful for other researchers. Some aspects of it (anonymized and devoid of any identifying details) might be made available as part of a publication (since most journals demand this now). The EEG database could be made available for researchers who might explore it using different data mining strategies.

Table 8: Data summary provided by UB

2.1.11. Data Summary of Capgemini

Expected purpose of the data and their relations to the objectives of the project:
Internally for Centralised Code Hosting in Gitlab and for Ticket Management we collected consortium partner user data and provided necessary privilege.
Expected re-use of existing data and its purpose:
Initially access is provided for a certain period (3,6,12 months) while extending access the same date will be reused.
Expected data types and formats:
User name and email addresses.
Expected size of the data:
n.a.
Expected origin/provenance of the data:
The data originated from personal information.
Expected use of the data outside the project:
Not Planned to use outside.

2.1.12. Data Summary of SOUND

Expected purpose of the data and their relations to the objectives of the project:
The data we collect is personal data to track, connect, and create a vivid community. It is used to keep people actively engaged in Joint Business Clinics and other business events and connect with them on business opportunities relating to PRESENCE.
Expected re-use of existing data and its purpose:
SOUND has no intention of re-using existing data.
Expected data types and formats:
We have personal information in data sheets that will not be shared outside the consortium.
Expected size of the data:
The size of the datasheets is expected to be less than 100 MB.
Expected origin/provenance of the data:
The data originates from personal information.

Expected use of the data outside the project:
No data will be published.

Table 9: Data summary provided by SOUND

2.1.13. Data Summary of IMEC

Expected purpose of the data and their relations to the objectives of the project:
As part of WP1, IMEC will conduct several user/human-centred activities (such as workshops, focus groups, interviews, etc.) with consortium partners and external users of the PRESENCE technological products/experiences. The activities carried out will be recorded (audio/video) for further analysis combined with notes taken during the activities. All the recordings and notes will be anonymized and will not be available to external organizations. The combination of desk research, interviews, and other human-centred activities will shape user requirements and expectations.
Expected re-use of existing data and its purpose:
Potentially information from the scientific literature.
Expected data types and formats:
During WP1, different data types and formats will be gathered. Personal data will never be available for or shared with third parties and will be anonymized when used for publications. The data formats that will be generated are DOCX, PDF, MP4, MP3, JPG, CSV, and XLSX. These data will be collected through interviews, focus groups, surveys, etc. Data for consortium partners will be accessible on the PRESENCE Google Drive and/or Slack. Anonymized data that has been prepared for analysis will be stored on a separate Drive folder accessible to consortium partners UHAM and IMEC. Raw data is stored with access to IMEC only.
Expected size of the data:
Up to 1 GB for text-based data and up to 1 TB for recorded audio/video data.
Expected origin/provenance of the data:
Re-used data will originate from scientific literature and generated data will be collected during user studies.
Expected use of the data outside the project:
If possible, open data sets and other information found/gathered in WP1 can be made publicly available to a certain extent, as the data is anonymised and the identity of participants is protected. Reports, deliverables, and other information for public distribution can be made available through public repositories or the Presence communication channels. Raw data (such as recordings, notes, user data, etc.) will not be available to third parties.

Table 10: Data summary provided by IMEC

2.1.14. Data Summary of JRS

Expected purpose of the data and their relations to the objectives of the project:
The data we generate synthetically, or re-use will be primarily used for the development (training, evaluation) of the AI technology for predicting and classifying the actions of humans. The data will consist of the media

files (mainly videos, images, and sensor information) and the corresponding annotations/ground truth.
Expected re-use of existing data and its purpose:
Existing public datasets will be used which are commonly employed for training and evaluation of AI algorithms for a certain task, e.g., the Kinect dataset for action recognition.
Expected data types and formats:
Input media will be mainly RGB videos or images, optionally with depth information (RGBD) and sensor data in various formats. Commonly employed compression standards are employed, like MPEG-4 or JPG/PNG. Annotations are usually in some human-readable text format like TXT, JSON or XML.
Expected size of the data:
Video content takes up much space, even with compression. We expect up to several hundreds of GB of space in total.
Expected origin/provenance of the data:
Internal or external preexisting data, data from the scientific literature, public data repositories, and data collected during user studies by other partners.
Expected use of the data outside the project:
Data could be of interest to researchers in the fields of human action recognition.

Table 11: Data summary provided by JRS

2.1.15. Data Summary of SVR

Expected purpose of the data and their relations to the objectives of the project:
The primary purpose of data generation is to enhance the effectiveness of SyncVR's software applications aimed at pain reduction. The data collected will include the duration of VR sessions, interactions such as clicks with the controller, head movements, and input from haptic gloves and vests. This data will help refine the VR experiences to ensure they are more effective at reducing pain and improving user comfort.
Expected re-use of existing data and its purpose:
No data will be re-used.
Expected data types and formats:
The data will primarily be collected in CSV format, which facilitates easy analysis and transformation. This format is suitable for capturing a range of data types, from numerical inputs like session duration to categorical data such as button clicks.
Expected size of the data:
The data size will be less than 10 GB.
Expected origin/provenance of the data:
The data will be generated from several hardware devices integral to the VR setup, including the VR

headset, controllers, haptic vests, and gloves. These devices will capture user interactions and physiological responses during VR sessions.
Expected use of the data outside the project:
The data could be of significant interest to researchers in fields related to VR, pain management, and human-computer interaction, offering insights into user behaviour and the effectiveness of different VR interventions.

Table 12: Data summary provided by SVR

2.1.16. Data Summary of ZAUBAR

As indicated in D7.2, the partner ZAUBAR declares that they currently have to have no intention to generate, collect, or re-use data within the PRESENCE project.

2.1.17. Data Summary of ARTANIM

Expected purpose of the data and their relations to the objectives of the project:
ARTANIM will collect motion data (tracking position) and questionnaire data to evaluate virtual reality experiences within WP1. This evaluation will be done with general volunteers including informed consent. We will also collect motion data from actors (or researchers standing in as actors) to be used as animation data used to animate characters, and to be used as reference movements to train character controllers.
Expected re-use of existing data and its purpose:
We do not plan to re-use data.
Expected data types and formats:
Animation data in FBX format and personal information in CSV format.
Expected size of the data:
Our data will have less than 10 GB.
Expected origin/provenance of the data:
The data will originate from VR hardware, motion capture systems, and online questionnaires.
Expected use of the data outside the project:
The data will be made available in anonymized form together with scientific publications to allow other researchers to validate and reproduce our results.

Table 13: Data summary provided by ARTANIM

2.2. Summary of generated, collected, or re-used data per study

The following section serves the purpose of summarising the data that is planned to be or has been collected during any studies performed within the scope of the PRESENCE project. These studies aim to 1) evaluate each utilised technology independently to understand its impact on presence and

other user experience-related measures; 2) evaluate each developed component to provide scientific and technical feedback in order to facilitate their use in each project iteration as well as beyond the project scope; and 3) evaluate the ultimate overall user experience of the project demonstrators that have been developed following a user-centred design process, as further described in the GA. Hence, each partner is required to provide a summary of the data that each user study collects under their responsibility.

Table 14 summarises the user studies that have been conducted so far.

The data summary will contain the following information for each study:

- Title of the study
- Leading partner for the study
- Responsible research ethics committee (REC) for the study when applicable
- Other partners involved in the study
- New data to be collected or generated within the study
- Existing data to be re-used in the study
- Data types and formats of the data
- Origin/provenance of the data
- Use of the data outside of the project
- Potential IPRs that apply to the data

Nº	**Short Title**	**Lead** (collaborating partners)	**Data Collected**	**Use outside of the Project**
1	IVH SDK v0.1 evaluation	UHAM	Questionnaires (data types: text and numbers): SUS, AttrakDiff, TAM2, open-ended questions	Publications and presentations
2	IVH SDK v0.2 evaluation	UHAM	Questionnaires (data types: text and numbers): SUS, AttrakDiff, TAM2, open-ended questions	Publications and presentations
3	IVH app evaluation at IEEE VR 25	UHAM	Questionnaires (data types: text and numbers): Godspeed, social presence questionnaire, TAM2	Publications and presentations
4	IVH experiment: The role of IVA on Emotional Contagion, Conformity, and Opinion Shaping	UHAM (ARTANIM, UB)	Questionnaires (data types: text and numbers): SSQ, IPQ, Conformity Scale, BFI-AGR, open-ended questions	Publications and presentations
5	IVH experiment: Interrupting IVA	UHAM	Questionnaires (data types: text and numbers): UEQ, Godspeed, open-ended questions	Publications and presentations
6	IVH experiment: Nuance in Non-Verbal Communication with IVA	UHAM	Questionnaires (data types: text and numbers): Godspeed, social presence questionnaire, mood, open-ended questions	Publications and presentations

Nº	Short Title	Lead (collaborating partners)	Data Collected	Use outside of the Project
7	IVH experiment: Deictic Interactions with a voice-only IVA	UHAM	Performance data (the time to complete a task, the number of prompts needed to complete a task, the number of correct and incorrect answers), Questionnaires (data types: text and numbers): NASA-TLX, UEQ, open-ended questions	Publications and presentations
8	IVH experiment: Comparing IVA and Static Picture Instructions for exercising	UHAM	Performance data (motion tracking data from the VR headset to assess head position accuracy during the exercises), Questionnaires (data types: text and numbers): NASA-TLX, UEQ, social presence questionnaire, open-ended questions	Publications and presentations
9	PRESENCE application evaluation	UHAM (IMEC)	Questionnaires (data types: text and numbers): SUS, AttrakDiff, TAM2, social presence questionnaire, IPQ, open-ended questions	Publications and presentations
10	Evaluation of methods for movement synthesis in characters approaching VR users	ARTANIM	Presence Questionnaires (data types: text and numbers), Motion data	Publications and presentations
11	Impact of physics-based animation methods on interpersonal behavior in IVH	ARTANIM & UHAM	Presence Questionnaires, Motion Data, eye tracking data, reaction times	Publications and presentations
12	Holographic Communications in Social VR: Subjective Experience Evaluation	i2CAT	USer Questionnaires (Presence, Visual Quality) and System Performance data (latency, bitrate, FPS)	Publication and presentations
13	Brain synchrony and eye-tracking to study co-presence in VR	UB	EEG, eye-tracking, Questionnaires (Presence), Sentiment analysis	Publications and presentations
14	Common Framework Experiment – The Gallery	UB	Decisions (3M Method), questionnaires	Publications and presentations'

Table 14: Summary of user studies

3. FAIR Data Guidelines

The FAIR data guidelines presented in the next sections aim to improve the findability, accessibility, interoperability, and re-usability of digital data generated, collected, and processed within the PRESENCE project. These principles are mandatory for each partner to follow to make publishable data and other research output publicly accessible and generally usable without technical, legal, or semantic barriers and thus maximise the impact and benefits of project investments. FAIR data not

only supports the re-use of data by individuals such as scientists but is also intended to enhance data collaboration across different fields and support computer systems like large language models to find and use the data automatically. Within the project, each partner generating or collecting data is responsible for publishing data following the principle of “as open as possible and as closed as necessary” while adhering to the FAIR data guidelines.

Following Wilkinson et al., 2016 [Ref. 3] and the GA, we consider the following aspects to be the critical requirements for implementing the FAIR data principles in the PRESENCE project.

Requirements for data **to be findable:**

F1. (Meta)data need to be assigned to a globally unique and eternally persistent identifier
F2. Data need to be described with rich metadata
F3. (Meta)data need to be registered or indexed in a searchable resource
F4. Metadata needs to specify the data identifier

Requirements for data to **be accessible:**

A1. (Meta)data need to be retrievable by their identifier using a standardised communications protocol
A1.1. The protocol needs to be open, free, and universally implementable.
A1.2. The protocol needs to allow for an authentication and authorisation procedure where necessary
A2. Metadata needs to be accessible, even when the data are no longer available

Requirements for data to **be interoperable:**

I1. (Meta)data needs to use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2. (Meta)data needs to use vocabularies that follow FAIR principles
I3. (Meta)data needs to include qualified references to other (meta)data

Requirements for data to **be re-usable:**

R1. (Meta)data needs to have a plurality of accurate and relevant attributes
R1.1. (Meta)data needs to be released with a clear and accessible data usage license
R1.2. (Meta)data needs to be associated with their provenance
R1.3. (Meta)data needs to meet domain-relevant community standards

3.1. Making data findable, including provisions for metadata

Data generated and processed within the PRESENCE project should be easily locatable by both humans and computers. To fulfil this, each partner generating or collecting data is responsible for ensuring that all appropriate data will be made openly available through an institutional (e.g., FDM for UHAM) or open data repository (e.g., Zenodo, EOSC). To ensure that the data is findable easily, rapidly, and identically, exact and standard parameters need to be used to identify the datasets. Each published dataset is required to receive a globally unique and persistent identifier (e.g., DOI or Handle) and each contributing individual must have a personal identifier (e.g., ORCID) to make both easily findable.

In addition, all datasets need to receive rich and indexable metadata, ideally according to a commonly used metadata schema (e.g., Dublin Core), to make them easy to discover. The metadata needs to provide information at least about datasets (description, date of deposit, author(s), venue, and embargo), Horizon Europe 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 must include persistent identifiers for related publications and other research outputs. The added metadata should further contain information about methods used to capture the data, the measuring instruments used, the individual who performed the collection or generation, the data formats and units, and the date and location when the data was created. Additionally, the metadata should include appropriate keywords describing the content of the dataset and offer information on how it can be re-used and how the included data needs to be interpreted.

3.2. Making data accessible

Once the data published by the PRESENCE project has been found, it should be easily accessible. Therefore, the datasets need to be stored openly available in an institutional (e.g., FDM for UHAM) or open data repository (e.g., Zenodo, EOSC), where they can be retrieved using open, free, and universally implementable protocols (e.g. HTTPS) provided by web services that also allow downloading the datasets. The chosen repository must ensure that uploaded datasets receive a globally unique and persistent identifier (e.g., DOI or Handle) by which they can be accessed. If additional software is needed to access the data, documentation on how the data can be accessed should be included in the dataset.

In general, all datasets associated with research output will be made accessible as default unless there is a specific reason not to publish the data. Other data may be made available on a case-by-case basis if it is relevant for third parties. The following data will not be made publicly available:

- Data that contains personal or sensitive information.
- Data was obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available.
- Data that discloses the identity of a manufacturer.
- Data that has been declared confidential by the CA.
- Data for which a partner lodges an objection to publication within a 15-day objection period.
- Data that compromises the protection of a partner(s) IPR or goes against their legitimate interests or other constraints as per the GA.

If public access to the data needs to be limited for any legitimate reason, the dataset needs to be listed under restricted access but will not be made available directly. Third parties and partners of PRESENCE may request access by contacting the issuing partner. In this case, the access rights are managed via the data repository. In addition, data may need an embargo period to allow time to publish the research output, protect intellectual property, or adhere to publisher conditions. For this reason, an embargo period of a maximum of two years can be applied to retain the data before it must be published.

The metadata of the datasets must be made available openly and licensed using the Creative Common Public Domain Dedication (CC 0) license or equivalent that allows for redistribution. Information about the accessibility of the dataset needs to be included in the metadata. All published

data and metadata will remain available as long as the repository remains active. The metadata should remain available if the data needs to be removed for any specific reason.

3.3. Making data interoperable

The PRESENCE project aims to make published data as interoperable as possible. However, as the exact nature of all data has not been completely defined at this project stage, it is not yet feasible to derive a comprehensive set of guidelines that fully guarantees data interoperability.

In general, all partners should aim to publish their data using common interoperable data formats like TXT, MD, CSV, JSON, XML, PNG, JPG, or PDF. If it becomes apparent that either proprietary or non-interoperable data formats need to be used, it must be described how interoperability can be achieved. For data where no agreed metadata standards exist, we aim to ensure interoperability by providing appropriate documentation on how to handle the data, including suggestions for compatible open-source software to handle the data. When naming and describing data, PRESENCE project partners are encouraged to use controlled vocabulary as far as available, adhere to the common terminology used in the respective field, and avoid ontologies that are not commonly used. If non-common vocabulary or ontologies will be used, they must be described in the data. To implement controlled vocabulary, it is recommended to consult the various controlled vocabulary entries on the Publications Office of the European Union website [Ref. 4]. Furthermore, it is recommended that research output identifiers be linked to published datasets and vice versa.

3.4. Increase data re-usability

The PRESENCE project aims to promote the re-usability of data that has been made publicly accessible. For this reason, the published datasets must contain comprehensive interoperable documentation addressing the data structure, the definition of variables, and the units of measurement. Ideally, the documentation is included in the corresponding research output. Where possible, datasets should be made accessible as early as possible after production following the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a license with equivalent rights. Experimental and test data might only become available after the publication of the results. We expect most of the data generated to be made available without restrictions and only datasets subject to IPR and other confidentiality issues will be restricted. Where this will be the case, agreements will be made based on the individual datasets. Requests for using the data by externals will be approved by the partner issuing the dataset.

To further support data re-usability, the PRESENCE project partners must ensure high data quality. To this end, the quality control of the data needs to happen at various stages during the quality assurance process. As the quality of the data collection methods used strongly influences data quality, an initial quality check must be carried out at the beginning of the data collection by the partners who were responsible for planning and implementing the measures. The initial check should include control of the data integrity and validation of the collected measures. Further controls should be performed by other involved project partners when available. A final quality check regarding the integrity of the dataset and metadata needs to be performed by the issuing partner before publishing the data.

4. Other Research Outputs

The PRESENCE project aims to uphold the FAIR data principles not only for data generated, collected, or re-used during performed studies but also for any other generated research output. Therefore, the following sections define the data handling procedures concerning any other expected research outputs.

4.1. Academic Publications

Scientific publications generated within the context of the PRESENCE project are considered data that need to be covered by the FAIR data principles. All partners must guarantee open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that a machine-readable electronic copy of the published version, or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications no later than at the time of publication. Immediate open access needs to be provided to the deposited publication via the repository under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a license with equivalent rights. For monographs and other long-text formats, the license may exclude commercial uses and derivative works (e.g., CC BY-NC or CC BY-ND). Further information needs to be given via the repository about any research output or any other tools and instruments that are needed to validate the conclusions of the scientific publication. To this end, partners and authors must retain sufficient IPR over their publications to ensure compliance with the access requirements. Before publishing any academic output, the partners involved need to contact the technical coordinator for revision/validation.

4.2. Software

Software developed within the context of the PRESENCE project should also adhere to the FAIR data principles and be made publicly available under the conditions of open-source software licenses (e.g., MIT, GNU, Apache) using either an open software repository (e.g., GitHub) or institutional solutions (e.g., GitLab). To this end, each partner acting as a licensor of software developed as part of the project, which is not subject to IPR, is responsible for making the software publicly available. However, it can be expected that parts of the software will be subject to IPR preventing it from being published openly. Software being distributed within the project should be stored using institutional software repositories that permit access only to authorised project partners, as further outlined in the section 6 Data Security.

5. Allocation of Resources

As per the GA, the project partner i2CAT is responsible for overseeing the overall data management within the PRESENCE project consortium in its position as task leader of T7.3 (Project Ethics, IPR, and Data Management) and as project coordinator. However, partners who generate or collect data are responsible for ensuring that the respective data adheres to the FAIR data principles. No further resources have been explicitly allocated within the GA for the purpose of managing the data of the project.

In general, the consortium aims to make all data accessible using either institutional or publicly available data repositories. These are free of charge to use and, to our knowledge, will remain so

also in the future. Accordingly, we expect that the costs for making data accessible will remain free of charge, ensuring long-term preservation. Regarding making scientific publications open access, publishers that offer open access free of charge should be favoured. Otherwise, it is up to the respective partners to decide whether and how they can cover the costs. Indirect workload incurred because of the effort required to comply with the FAIR data principles needs to be reported within the corresponding WPs. Hence, it is not expected that the PRESENCE project will have additional costs to implement the FAIR data principles. Any changes to this policy or exceptions will be included in future versions of the DMP.

6. Data Security

All data management activities within the PRESENCE project need to adhere to the data security policies described in this section. To this end, each partner of the consortium is responsible for implementing the policies, including the logical security, physical security, network security, and security of computer systems, to ensure the security of data and prevention of unauthorised access, changes to data, disclosure, or destruction of data.

6.1. Personal Data Security

The PRESENCE project partners must fully comply with the GDPR to ensure lawful and ethical use and avoid data breaches of personal data. For this purpose, the ethical handling of personal data beyond the general policies for logical and physical data security is further specified in the section about ethics of this document.

6.2. Logical Data Security

To protect unpublished data generated, collected, or re-used within the PRESENCE project from unauthorised access, secure user authentication methods need to be applied to restrict the access of data. To this end, each member of the project must use unique credentials to access project data, including consistently strong passwords and two-factor authentication (2FA) where possible. Project partners are not allowed to store any confidential data, such as those containing personal information, on servers or computers connected to an external or free network. An exception to this is the sharing of data between partners of the PRESENCE consortium, which must be realised through secure institutional solutions within the individual WPs and adhere to the here-described data security policies. In this case, strict access levels must be assigned based on the user's role within the project. For instance, project or technical coordinators or task leads should have full access to all data, while executing partners should have restricted access based on their tasks. A version control system (e.g.,Git) should be used to manage changes to data and ensure that accurate records of modifications are maintained.

To maintain network security, firewall protection and security-related upgrades and patches for the operating systems of PCs and servers must be put in place to prevent attacks through breaches, viruses, and malicious code. Data transmitted between systems using an internet connection need to be encrypted using safe transport layer security protocols (e.g., HTTPS) to protect against interception and eavesdropping. Automated full backups of all data should be conducted weekly, with incremental backups performed daily, also including PCs and workstations used by project members. Backup copies should be stored in secure off-site locations to protect against data loss

due to hardware failure or other disasters as further described in the next section about physical data security. In the event of data loss or corruption, the responsible partner needs to initiate data recovery procedures, aiming to restore data within 24 hours.

6.3. Physical Data Security

To assure physical data security, all PRESENCE partners need to implement highly restricted access to the buildings where data, computers, or servers are held. Transporting of data storage media is restricted to exceptional circumstances (i.e., repair of physical systems). Ideally, the servers are protected against data loss and downtime due to power failures by surge protection systems and uninterruptible power supply systems.

To ensure the security of local computer systems and files, all members of the PRESENCE partners must always lock their computer systems with a strong password when currently not in use. To avoid breaches or loss of data in case of theft or loss of an end device, data should be stored in encrypted form (e.g., AES-256) on the hard drives of the end devices and backups should be performed regularly as described above.

7. Intellectual Property Rights

The general rules regarding IPR within the PRESENCE project are governed by the GA and further detailed in the CA. To ensure compliance with IPR and to reinforce the success of the project, all project partners must adhere strictly to the established rules. The following section summarises how these rules affect the data management policy of the project and formulate guidelines ensuring that all data is managed in full compliance with the IPR provisions, safeguarding the interests of all partners, and facilitating the effective and lawful use of the project outcomes.

Data generated, collected, or re-used within the PRESENCE project must strictly adhere to the rules for IPR specified in the GA. To this end, data considered as background, such as pre-existing know-how or information critical for the project's implementation or exploitation of results, must be accessible to other partners on a royalty-free basis to fulfil specific project obligations and ensure collaborative efficiency, unless restricted by external third-party rights or specific rules in the GA or CA. This background data should be clearly identified and documented to facilitate its appropriate use and access within the project.

Newly generated data, such as results from user studies or datasets, remain the property of the partner who generated them. If two or more partners have jointly generated the data and it is not possible to establish the respective contribution of each partner or separate them for the purpose of applying, obtaining, or maintaining their protection, the data will be subject to a joint ownership agreement as further outlined in the GA. Generated data should be protected through suitable means, such as patents or copyrights if deemed commercially viable and justified. Furthermore, partners must employ their best efforts to exploit the results, either directly or through licensing, ensuring that the data serves its intended purpose beyond the project duration. If the results are not exploited within one year after the project ends, partners should use the Horizon Results Platform to find interested parties for further exploitation. For data re-use, it is critical that partners ensure compliance with any third-party rights. This means that all necessary licenses and authorisations

need to be obtained before incorporating third-party data into the project. This diligence helps to avoid any legal disputes or breaches of IPR during and after the project.

To protect the ownership of the data, proper documentation and certification is essential. For this reason, the DMP maintains a list of the planned and conducted user studies in the data summary section, documenting the ownership of the data and potentially existing IPR. All data that is subject to IPR, whether newly generated or re-used from prior work, must be clearly attributed to its original owner and handled in a way that preserves its integrity and confidentiality unless explicit authorisation from the data owner is obtained. Adhering to these practices not only ensures compliance with IPR agreements but also promotes a collaborative and transparent research environment.

In the event of a partner planning to disseminate project results, a minimum of 15 days advance notice must be provided to the other involved partners, along with detailed information regarding the intended results for dissemination, unless otherwise stipulated. Should any other party have concerns about potential harm to their IPR with regard to the results or background information within 15 days of receiving the notification, they reserve the right to object. Under such circumstances, dissemination of the results will be suspended unless measures are taken to address and protect those interests.

Except from existing IPR, the granting authority has specific rights to use non-sensitive project-related data, such as summaries, deliverable documents, and multimedia materials, for purposes such as policy making, information dissemination, and publicity. This usage right is granted under a royalty-free, non-exclusive, and irrevocable license. It is important for partners to ensure that any data shared with the granting authority is non-sensitive and complies with the project's data use and dissemination policies.

8. Ethics

The PRESENCE project consortium aims to maintain the highest ethical standards recommended by professional bodies, institutions, and governments both during user studies and when generating, collecting, or re-using data. In general, all planned user studies need to be performed in accordance with the Declaration of Helsinki. Particular attention shall be paid to:

- The principle of proportionality.
- The right to privacy.
- The right to the protection of personal data.
- The right to the physical and mental integrity of a person.
- The right to non-discrimination.
- The need to ensure the protection of the environment.
- The need to ensure high levels of human health protection.

Before conducting a study involving human participants, participating partners should critically scrutinise whether their planed evaluation activities will contradict the above-mentioned principles in any way. When needed, an ethics self-assessment form will be developed in consultation with the external ethics advisor (EEA) and will be appended to future versions of this document. In case the

critical self-assessment raises ethical concerns, it is imperative that studies involving human participants are not conducted unless approved by the research ethics committee (REC) of a participating partner. Partners who do not have access to a REC need to cooperate with a partner who has access to a REC or obtain external ethics approval. A list of PRESENCE partners with access to a REC can be found in Table 15.

For example, all studies conducted by UHAM have assessed the initial ethics criteria by filling out the basic questionnaires[2]. According to the guidelines of the local ethics committee at UHAM, if all questions are answered with "no", we do not need to consult the ethics commission. However, in case one or more questions are answered with "yes", we need to send the completed basic questionnaire and also the additional questionnaire "Ethics Assessment Form" as well as other relevant information (e.g. questionnaires, informed consent document) on our study to the Ethics Commission to ask for ethics approval. Partners conducting user studies without official ethical approval must be aware that publishing scientific results and data in high-impact publication targets often requires the approval of a REC or at least proof of compliance with local ethical guidelines. To avoid data breaches leading to ethical objections, any generated, collected, or re-used data, particularly data originating from user studies, must always strictly adhere to the guidelines for data security described above.

Partner	Research Ethics Committee (REC)
UHAM	Ethics Commission from the Department of Computer Science at Universität Hamburg (UHH)
UB	Research Ethics Committee at University Barcelona and Research Ethics Committee with medicines (CRE/CREm) at Institut d'Investigació Biomèdica de Bellvitge (IDIBEL)
IMEC	Ethics Committee for Human Sciences (ECHS) at Vrije Universiteit Brussel (VUB)
ARTANIM	Research Ethics Committee (CUREG) at Université de Genève (UG)

Table 15: List of PRESENCE partners with REC

8.1. Recruitment, information, and consent for data collection

When recruiting participants for user studies as part of the PRESENCE project, particular care must be taken to ensure a diverse selection in terms of gender, ethnicity, identification, age, orientation, education, beliefs, and more. Recruitment of adult healthy volunteers should be favored whenever possible, with exceptions requiring ethical approval. All human participants who intend to take part in a user study performed by any project partners must be informed about the purpose of the study and the collected data, the procedure they are undergoing, and the potential risks related to it (notably using VR). Additionally, they must be informed about the procedures for handling incidental findings, especially when collected data may contain clinical findings. Participants must explicitly agree to participate in the study and confirm that they have been comprehensively informed about it. Where studies might involve some level of discomfort, participants must be explicitly informed of this point in the written information and verbal briefing before participation. All participants will be free to withdraw at any stage without reason. They may receive compensation for their participation depending on the guidelines and best practices of the institution concerned. Data from user studies

[2] Available for consultation under demand

with human participants may only be made publicly accessible if the participants consent to the publication of their data and the data is properly depersonalised. Sample documents for information and consent are provided in D1.1, which includes a comprehensive description of the foundations for conducting user studies involving human participants within the project.

8.2. Handling of personal data

To ensure the proper handling of personal data in the PRESENCE project, the General Data Protection Regulation (GDPR) requirements will be implemented. According to GDPR, personal data refers to any information relating to a natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. As a first baseline, the project will work along the seven principles: Notice, Purpose, Consent, Security, Disclosure, Access, and Accountability as stated in the GDPR on the protection of personal data.

Therefore, all personal data of participants will be treated confidentially by only storing and sharing it through repositories with restricted access to authorised users of the partners involved in the data collection. Participants must be informed with which institutions their personal data will be shared, and the deletion of personal data must be guaranteed at any time upon request. As soon as possible, personal data must be depersonalised by means of anonymisation, pseudonymisation, or coding list procedures. To this end, anonymisation defines the process by which personal data is irreversibly altered in such a way that the natural person can no longer be identified directly or indirectly, either by a project partner alone or in collaboration with any other party. Pseudonymization, on the other hand, is a well-known deidentification process where the name or another identifying feature is replaced by a pseudonym, usually a code consisting of a combination of letters or numbers, in order to rule out or make it significantly more difficult to establish the identity of the natural person represented by the code. A coding list is a detailed mapping of data elements to specific codes with the purpose of replacing identifiable information with anonymised or pseudonymised data. When using coding lists, it needs to be assured that the list and data will be stored safely in different locations to preserve privacy.

8.3. Handling of incidental findings

In the event of any unanticipated incidental finding throughout the execution of the PRESENCE project that raises any ethical concerns, the incidental findings need to be immediately forwarded to the project and technical coordinators to assess their ethical implications and decide on the next steps. The data protection officer is only involved if it appears necessary. The following rules will govern any incidental findings:

- In case the impact of the incidental findings can be neglected, the project and technical coordinators will consider the deletion of the findings.
- In case of incidental findings that include recording an illegal activity, the consortium will comply with all relevant local and international laws.
- In case of incidental findings that include any information of public interest, the project and technical coordinators would decide about the need, means, and timing of their communication to relevant stakeholders.

- In case of incidental findings raising ethical concerns through the execution of the project, it will be referred to the PRESENCE responsible partner for the study and the project management team with the view to evaluate the implications and to reach a decision on further action. If there is danger to life and limb, any necessary action must be taken immediately. The relevant authorities need to be informed only if considered necessary. In addition, confidentiality needs to be maintained in line with GDPR and local laws and the responsible REC or ethical advisor should be contacted for advice on informing the participant. If required, clear information needs to be provided to the participant and further medical evaluation is recommended.

9. Other Issues

According to the current state of the PRESENCE project, there are no other issues that need to be addressed in the DMP.

10. Abbreviations and Definitions

10.1. Abbreviations

2FA	Two-Factor Authentication
AES	Advanced Encryption Standard
CA	Consortium Agreement
CC BY	Creative Commons Attribution International Public Licence
DMP	Data Management Plan
EEA	External Ethics Advisor
EU	European Union
GA	Grant Agreement
GDPR	General Data Protection Regulation
HTTPS	Hypertext Transfer Protocol Secure
IPR	Intellectual Property Rights
PC	Personal Computer
REC	Research Ethics Committee
WP	Work Package

10.2. Definitions

FAIR data	FAIR data refers to data that meets principles of being Findable, Accessible, Interoperable, and Reusable, ensuring it can be easily shared and utilized for research and other purposes.
Consortium Agreement	Means the agreement concluded amongst PRESENCE partners for the implementation of the action. Such an agreement shall not affect the partners obligations to the community and/or to one another arising from the contract.
Open Access	Open Access is the practice of providing unrestricted online access to scholarly research, allowing anyone to read, download, and use the content without financial, legal, or technical barriers.
Declaration of Helsinki	The Declaration of Helsinki is a set of ethical principles developed by the World Medical Association to guide researchers in conducting research involving human subjects, ensuring respect for individuals and prioritizing their well-being.
Intellectual Property Rights	Intellectual Property Rights are legal protections that grant creators exclusive rights to use and distribute their inventions, literary and artistic works, and symbols, names, and images used in commerce.
Research Ethics Committee	A Research Ethics Committee is a group of independent experts tasked with reviewing and overseeing research proposals to ensure that they comply with ethical standards and protect the rights, dignity, and well-being of participants.

11. References

Ref. 1 General Data Protection Regulation (GDPR) of the European Union (EU) 2016/679 https://eur-lex.europa.eu/eli/reg/2016/679/oj

Ref. 2 https://privacy.didimo.co/api-terms-of-use-license-agreement/

Ref. 3 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9. DOI: https://doi.org/10.1038/sdata.2016.18

Ref. 4 Publications Office of the European Union website, controlled vocabulary: https://op.europa.eu/en/web/eu-vocabularies/controlled-vocabularies

Annex I: Report of the External Ethics Advisor

ETHICS ADVISOR REPORT

PRESENCE - A toolset for hyper-realistic and XR-based human-human and human-machine interactions (Grant Agreement n. 101135025)

Period: 1st July 2024 – 30th June 2025
(Note: the first version of this report was presented in February 2025 with an analysis of D4.1. The current version adds an analysis of D1.2 and D7.4)

APPOINTMENT OF AN EXTERNAL ETHICS ADVISOR

The PRESENCE consortium proposed the appointment of an External Ethics Advisor (EEA) at the time of the proposal. In May 2024, Joana Porcel was appointed as the project's EEA to assess the ethical aspects of the work carried out in the project and to provide independent recommendations. In particular, the following deliverables will be review and reports will be prepared (June 2024, June 2025, December 2026):

Year 1:
D1.1- Human Centred Development Phase I - Foundations, Requirements and Initial Planning
D7.2- Ethics Framework and Data Management Plan I

Year 2:
D1.2- Human centred Development Phase II - Intermediate User Testing, Presence Evaluation, Ethics, Trust & Privacy
D4.1- Virtual humans technologies report I
D7.4- Ethics Framework and Data Management Plan II

Year 3:
D1.3- Human centred Development Phase III – Final User Testing, Presence Evaluation, Ethics, Trust & Privacy
D4.2- Virtual humans technologies report II
D7.6- Ethics Framework and Data Management Plan III

Additional input in the set-up of the experiments, e.g., the content of the informed consent, the nature of the requested participation, data management practices or provided incentives is also foreseen.

MEETINGS and OTHER CONSULTATIONS

During the Period, the beneficiaries and the EEA have met on:

DATE	ATTENDEES	OBSERVATIONS
July 10th, 2024	- PRESENCE Consortium - Joana Porcel (ISGlobal). EEA	The EEA attended the consortium meeting on July 2024. The presentation is attached to this report.
February 5th, 2025	- Louise Hallström, Researcher VUB - Joana Porcel (ISGlobal). EEA	Consultation if obtaining ethical approval from an ethical board is required for the study where the general public (students from the Vrije Universiteit Brussel) will try the First Playable app, a VR application developed within the project. This test session will last approximately 30

		minutes and is designed to gather feedback on user experience and usability, using questionnaires prepared by the partner UHAM. The researcher is advised that in my role as EEA, I cannot give ethics approval. Approval, if required, must come from an accredited committee. I recommend to contact the university committee that reviews social science or technology projects to check whether this study requires approval under local regulations. Also check if the involvement of students requires specific approval from the university. The informed consent form was also reviewed. It is very well prepared. The only advice I gave was about the possibility of anonymising the data in case participants do not need to be contacted again, including the removal of the voice recordings.

DOCUMENTS REVISED

D1.2- Human centred Development Phase II - Intermediate User Testing, Presence Evaluation, Ethics, Trust & Privacy (version 1.0, 20-06-2025)

The document, provides an intermediate report on the human-centred design and evaluation activities within the project. The specific aims are to report on i) the intermediate phase of usability and User Experience (UX) testing; ii) the first iteration of presence evaluation testing; iii) the validation of the user requirements and use case definitions; and iv) the ethical, trust, and privacy-related issues.

The deliverable describes the activities involving human participants to gather feedback and evaluate the developed technologies (summarised at the Table. Summary of UX (and sense of presence) studies).

Several activities have been conducted under Task 1.5 - "System Ethics, Trust and Privacy", which define the ethics, trust and privacy strategy, describing the gender, ethical, trust- and privacy-related issues to guide the further development of the PRESENCE technologies. During the playable app testing, users provided feedback on potential ethic issues. The main themes that were discussed were related to accessibility and inclusivity, privacy and data transparency, cultural and historical integrity, realism and functionality, and equal representation of people.
A series of workshop for the consortium's member are being planned for analysing and discussing the ethics issues related to the use cases and the use of AI. The results of the workshops will also inform the AI trustworthiness assessment, in relation to the AI act and in the context of the use cases.

As recommended in the first report, it is advisable that young researchers (pre-doctoral and post-doctoral) and technicians also contribute to the discussions to be held in these workshops.

D4.1- Virtual humans technologies report I (Version 0.5; 18-02-2025)

This deliverable describes the objectives, key performance indicators (KPIs), and essential tasks concerning the technological pillars of smart avatars (SAs) and intelligent virtual agents (IVAs).

The document includes a section on the Ethical Considerations, including privacy issues and the legal basis for personal data processing (see Privacy policy of Didimo; https://privacy.didimo.co/privacy-policy/). An analysis of the different tools, platforms used in these activities in terms of privacy is also included.
The researchers inform that certain content of hate speech, violence, harassment, etc. may occur when using ChatGPT. As well as informing participants that this may happen, the team is advised to monitor closely and take steps to minimise its occurrence.
Elements of age, gender and ethnicity diversity in the design of the 3D human models are considered.
The following manuscript is recommended: Gerlich, Michael. 2024. Societal Perceptions and Acceptance of Virtual Humans: Trust and Ethics across Different Contexts. Social Sciences 13: 516. https://www.mdpi.com/2076-0760/13/10/516

D7.4- Ethics Framework and Data Management Plan II (Version 1.0; 20-06-2025)

This is an updated version of the data management plan. This version incorporates updated and more detailed information on data management practices, specific partner contributions, and explicitly detailing compliance with ethics recommendations. The document includes a detailed section on ethics, ensuring that all user studies and data handling activities will adhere to the highest ethical standards, including the declaration of Helsinki. The list of research ethics committees is included. Specific sections describe the i) recruitment, information, and consent for data collection, ii) handling of personal data, and iii) handling of incidental findings.

The recommendations from the first report of the EEA have been addresses as follows:

- It is advisable to develop project-specific protocols 1) to manage personal data incidents, including data breaches, and 2) to address the rights of participants regarding the processing of personal data. These protocols could be part of an updated version of the DMP and should be shared with all members involved in the project.
The current version does not include a description of these protocols. It is advisable for all consortium members to receive training on how to manage potential personal data-related incidents and participants' requests for their rights. To this end, it is important that consortium members have access to their institutions' DPO details.

- The contact details of the DPOs could also be included in the DMP so that all partners can easily contact the DPOs if necessary.
The contact details of the partners' DPOs are not included at the document. However, this is not needed, as far as their contact details are available internally to all members of the consortium.

- A project-specific policy for dealing with potential incidental findings should be developed, as the current description is rather general. This is particularly relevant in the case of potential incidental findings that may affect participants' health. The policy needs to be communicated to participants during the informed consent process.

This has been confirmed at section 8.3. Recruitment, information, and consent for data collection: “Additionally, they must be informed about the procedures for handling incidental findings, especially when collected data may contain clinically relevant findings (notably using EEG).”.

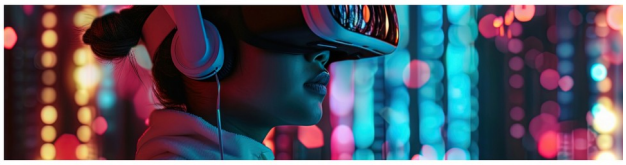
- Regarding the privacy of participants, it is mentioned that the data will be anonymised, but it also appears that the key to the personal identifiers will be retained. Where possible, the consortium is encouraged to use anonymised data rather than pseudonymised data.
It is indicated that while pseudonymisation is a default approach with the key to personal identifiers securely separated, the document explicitly encourages and aims for the use of anonymised data whenever feasible to enhance participant privacy. This shows that the recommendation is acknowledged and integrated into the data management strategy of D7.4.

ETHICS ISSUES ON EXTENDED REALITY

The consortium is recommended to explore the training modules of the iRECs project and to recommend young researchers and other relevant project staff to undergo this targeted training.

iRECs - Training modules on Extended Reality

Welcome To Extended Reality: Technology Basics.
https://classroom.eneri.eu/node/259



Welcome To Extended Reality: Ethics Issues.
https://classroom.eneri.eu/node/266

Signed:

PORCEL CARBONELL JUANA ALICIA

Firmado digitalmente por PORCEL CARBONELL JUANA ALICIA
Fecha: 2025.06.30 13:38:44 +02'00'