Project No. 964220

Intelligent digital tools for screening of brain connectivity and dementia risk estimation in people affected by mild cognitive impairment

Deliverable D2.1

Standardisation of available and prospective data collection

WP2 – Data management and features extraction

Authors	Mia Liljeström, Hanna Renvall, Antti Kinnunen, Juha Montonen, Ville Mäntynen (HUH), Christoffer Hattlestad-Hall (Brainsymph)
Lead participant	HUH
Delivery date	28th October, 2021
Dissemination level	Public
Type	Report

Version 1.0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964220

1. Revision History

Author(s)	Description	Date
Mia Liljeström (HUH)	Draft deliverable	07/10/2021
Hanna Renvall (HUH)	Revision 1	10/10/2021
Antti Kinnunen (HUH)	Revision 2	12/10/2021
Ville Mäntynen (HUH)	Revision 3	13/10/2021
Juha Montonen (HUH)	Revision 4	15/10/2021
Christoffer Hattlestad-Hall (Brainsymph)	Revision 5	18/10/2021
Federico Ramírez, Fernando Maestú (UCM)	Revision 6	19/10/2021
Camillo Marra (UCSC), Francesca Miraglia (Neuroconnect), Paolo Maria Rossini (IRCCS), Fabrizio Vecchio (Neuroconnect)	Revision 7	20/10/2021
Ira Haraldsen (OUS)	Revision 8	25/10/2021
Hanna Renvall (HUH)	Revision 9	27/10/2021
Andreia Cruz (accelCH)	Final version	28/10/2021

2. Abbreviations

AD	Alzheimer's disease
AI	Artificial intelligence
APOE	Apolipoprotein E
BIDS	Brain Imaging Data Structure
CANTAB	CAmbridge Neuropsychological Test Automated Battery
CE	Certificate of medical device to comply with the applicable EU regulation
DTA	Data transfer agreement
HPI	Head Position Indicator
ECG	Electrocardiography
EDC	Electronic Data Capture
EDTA	Edetate disodium
EEG	Electroencephalography
EOG	Electro-oculography
MCI	Mild Cognitive Impairment
MDD	Medical Device Directive
MEG	Magnetoencephalography
MRI	Magnetic resonance imaging
NPT	Neuropsychological testing
P-Tau181	Plasma phosphorylated tau181 protein
SOP	Standard Operation Procedures
TSD	Services for sensitive data - Tjenester for sensitive data, University of Oslo
WP	Work Package

3. Partner Short Names

OUS	Oslo University Hospital
AALTO	Aalto University
accelCH	accelopment Schweiz AG
AE	Alzheimer Europe
Brainsymph	BrainSymph AS
DNV	Det Norske Veritas
HUH	Helsinki University Hospital
IRCCS	Scientific Institute for Research, Hospitalization and Healthcare, San Raffaele Roma
Lurtis	Lurtis Rules S.L
Neuroconnect	Neuroconnect Srl
OsloMET	Oslo Metropolitan University
Radboudumc	Radboud University Medical Center
TLU	Tallinn University
UCM	Complutense University of Madrid
UCSC	Università Cattolica del Sacro Cuore

Table of Contents

1 Executive Summary

This deliverable (D2.1) defines the technical requirements for standardised data collection (Electroencephalography (EEG) and magnetoencephalography (MEG), digital cognitive testing, blood sampling, textual data) to be used by the clinical partners of AI-Mind for obtaining comparable and high-quality clinical data.

1.1 Purpose and Scope of the Deliverable

This deliverable describes the shared requirements, technical solutions, and methods for collecting clinical data at the AI-Mind clinical partner sites. For that purpose, we define protocols for using research devices, materials, and clinical data handling practises to be implemented at all prospective participant visits. The clinical data will be collected at five sites in four countries (OUS, Oslo, Norway; HUH, Helsinki, Finland; IRCSS and UCSC, Rome, Italy; UCM, Madrid, Spain). The scope of this deliverable spans from acquiring the EEG/MEG data, digital cognitive testing, blood samples, and textual data from questionnaires and neuropsychological testing (NPT), to managing the (pseudonymised) clinical data at the respective sites. These protocols apply to all the clinical partners; MEG protocols will be implemented only in UCM and HUH.

The AI Mind project handles both prospective (to be collected) and retrospective data (acquired earlier for other clinical and scientific purposes and included in the data transfer agreements (DTA) between the clinical sites). The emphasis of this deliverable is on the handling prospective data, but it will also define the minimum common requirements for the data handling protocols of the retrospective EEG data at the respective sites. Harmonisation of retrospective data will be further discussed in Deliverable D2.3, where also the preprocessing protocols for all the clinical MEG/EEG data are defined.

The AI-Mind standardisation of available and prospective data collection covers the following:

- Description of data collection hardware, software, and equipment for the prospective data;
- Minimum requirements for handling retrospective EEG data;
- Procedures for acquiring EEG/MEG data, digital cognitive testing, and textual data from NPT and background questionnaires;
- Procedures for handling blood samples for genetic testing and protein analysis;
- Requirements for pseudonymisation and handling of data and metadata at the clinical site;
- Measures taken for risk mitigation.

Here is the relation of D2.1 with other AI-Mind deliverables:

- Legal and ethics consideration of AI-Mind data collection and sharing - D1.3 (M9)
- AI-Mind data governance and data management protection framework - D1.4 (M12)
- Standardisation of prospective data and preprocessing procedures - D2.3, D2.5 (M16, M38)
- Secure transfer of clinical data to AI-Mind's shared internal platform - D4.1 (M10)
- Participant recruitment and inclusion/exclusion criteria and procedures - D5.1 (M9)
- Prospective data collection at the clinical sites - D5.2, D5.7 (M24, M60)

1.2 Outcomes

Well-defined procedures in acquiring prospective clinical data within AI-Mind ensure high-quality of the data being collected by multiple clinical partners during the multi-year project. This deliverable provides a solid base for our partners to develop the AI Connector and AI Predictor, thus acting as an essential part of the whole AI Mind Project.

2 Background

2.1 Non-invasive electrophysiological recordings

EEG and MEG provide a non-invasive measure of electrical activity in the brain with millisecond temporal resolution. Neural currents from thousands of synchronously active neurons sum up linearly to form a weak electromagnetic field that is measurable outside the head (for a review, see Hämäläinen *et al*. 1993). The main contribution to these measurable signals arises from postsynaptic currents (Creutzfeldt 1983). Using an array of electrodes placed on the surface of the scalp, EEG measures the electric potential differences on the surface of the head between each electrode and a chosen reference. MEG measures the magnetic field caused by the same neural activity with devices equipped with sensitive SQUID sensors.

In the AI-Mind project, the neurophysiological data will consist of EEG resting-state recordings measured with high-resolution 128-electrode devices, and in a subset of the participants, of concurrent MEG recordings measured with 306-channel devices (Helsinki and Madrid). Resting-state EEG/MEG recordings particularly provide a measure of ongoing oscillatory neural activity and their pathological modulations in neurological disorders. Computational methods will allow for estimation of the measured signals' neural sources and interactions between brain regions at specific frequencies.

2.2 Overview of data collection

After inclusion into the study is confirmed in a separate screening procedure (see D5.1), participants will follow the study protocol for 2 years, during which they will take part in four measurement sessions (M0, M8, M16, M24), as specified in D5.1 (see Table 1 for an overview).

The participant will complete an EEG recording and digital cognitive testing during each measurement session using CANTAB (CAmbridge Neuropsychological Test Automated Battery; Cambridge Cognition Ltd.). A subset of the participants will complete a MEG recording concurrently with their EEG recording on their first and last visit (M0, M24) in Helsinki and Madrid. Co-registration data for alignment of EEG (and MEG) data with template MRIs will be obtained at each measurement.

In addition, blood sampling will be performed in all participants at the first visit (M0) for genetic testing (APOE allele) and protein biomarker analyses (P-tau181). Textual data (*e.g.*, patient demographics, background medical information, as specified in Deliverable 5.1), will be collected during the first and last visit (M0, M24) using digitised web forms when possible. Structural MRIs, if available, may be provided for a subset of the participants for optimising M/EEG source reconstruction.

Table 1. Overview of prospective data collection

		M0	M8	M16	M24
Inclusion/consent to participate	X				
EEG		X	X	X	X
MEG		*			*
Coregistration data		X	X	X	X
CANTAB testing		X	X	X	X
Blood sample		X			
NPT		X			X
Textual data		X			X
MRI		**			
Technical and clinical logs		X	X	X	X

*MEG is collected from a subset of the participants at two clinical sites (Madrid, Helsinki)
**Structural MRIs are provided for a subset of the participants if available

2.3 BIDS standard

The EEG and MEG data collection procedures in the AI-Mind project have been developed to ensure compliance with the Brain Imaging Data Structure *(BIDS) standard*. The BIDS (https://bids.neuroimaging.io), originally developed for MRI images, and later expanded to cover EEG and MEG data, is a research community standard for organising and sharing brain imaging data (Gorgolewski *et al*. 2016, Niso *et al*. 2018, Pernet *et al*. 2019). The BIDS standard follows the FAIR principles of findability, accessibility, interoperability and reusability, by providing rich metadata in sidecar files, and by using standard data formats. The BIDS standard merges data from different brain imaging modalities into the same general framework and stipulates naming conventions for structuring brain imaging data and metadata. Importantly, the structured BIDS data enables fully automated data analysis workflows on multiple imaging modalities.

3 Data collection hardware, software and equipment

3.1 EEG

The EEG data collection hardware and software of AI Mind have been acquired in a joint call for tender, to ensure that every site has a similar operational system. The crucial requirements for the product were *i*) CE certified medical device, *ii*) possibility for high-density EEG measurements (128 channels), *iii*) and MEG compatibility.

3.1.1 EEG Hardware and Equipment

EEG measurements are conducted with an ***eego™*** system (eemagine Medical Imaging Solutions GmbH), provided by ***ANT neuro***. The system provides instrumentation for data collection, visualisation, and archiving. *eego™* is CE marked as a class IIa medical device according to the Medical Device Directive (MDD).

At all four clinical sites, a cascaded 128-channel system with two 64-channel amplifiers and ***waveguard™*** 10/5 layout EEG caps are used (see **Appendix A** for specifications). The EEG caps

additionally include 1 electro-oculogram (EOG) electrode, 1 electrocardiography (ECG) electrode and a ground (GND) electrode. Shielded cables include removable (for MEG-compatibility) ferrite cores. An additional Sensebox with adapter for four auxiliary and two passive sensors is included for, *e.g.*, the possibility of additional electromyography (EMG) recordings. A trigger input to ***eego™*** will be used for synchronising MEG and EEG measurements; see below 3.2.2.

3.1.2 EEG measurement environments

The EEG measurement environments at the clinical sites are non-identical, but the effort is made for harmonising them for general lighting and background noise level and visual surroundings.

In *Helsinki*, the measurements will be conducted in two magnetically shielded rooms. Both rooms are of low noise and with adjustable lighting. In *Rome,* the measurements will be conducted in a quiet room with low noise and adjustable lighting. In *Oslo*, the EEG measurements are acquired in two close-to-identical rooms without shielding (i.e., normal clinical EEG rooms). The lighting is adjustable, and measures are taken to minimise background noise. In *Madrid*, measurements will take place in several locations (Center for Biomedical Technology for simultaneous EEG and MEG data, and in clinical sites for EEG alone). The light level is adjustable in the first location and will be monitored in the remaining ones.

During the measurements, the participants will sit comfortably in a chair, and a visual fixation cross mounted 1.2 m away from the participant's eyes will be implemented at all sites to minimise eye movements. To reduce variation between-participant regarding task comprehension, written instructions will be given to the participants in their native language before the measurements.

3.1.3 EEG acquisition software and settings

The ***eego™*** system uses dedicated acquisition software, which operates on Microsoft Windows and is USB connected to the ***eego™*** amplifier. Software version 1.9.2 is available at all sites at the time of this deliverable. No other applications will run in parallel to the ***eego™*** software.

A fixed measurement setup will be used at all clinical sites to assure similar data collection. The setup will consist of *i*) the amplifier settings (all 128 electrodes enabled), *ii*) the corresponding montage setting, and *iii*) sampling frequency (with corresponding anti-aliasing frequency). Note that the montage setting does not affect the raw data collection. The EEG measurement settings are depicted in **Table 2**.

Table 2. EEG measurement settings

EEG settings	
Number of EEG electrodes	128
Number of other electrodes (EOG, ECG, GND)	3
Sampling rate	2000 Hz
Filters	0.01-660 Hz
Range	150mV
Reference	CPz

3.1.4 EEG digitisation

Information on the spatial locations of the EEG electrodes with respect to individual anatomical landmarks will be acquired using a 3D scanner device before the measurements. From the 3D model, the electrode locations will be identified utilising custom-made software.

3.1.5 Retrospective EEG data

During the project, the retrospective EEG and MEG data sets (available in corresponding patients at three clinical partner sites) will be evaluated for compatibility with the prospective data. The retrospective data have been collected with slightly varying protocols at the respective sites. The data sets that will be accepted to be used in the project will be harmonised with prospective data definitions/guidelines to the extent possible and managed by the same principles (see D2.3 for standardisation and preprocessing of the data).

Minimal common requirements for the retrospective EEG data to be accepted for further inspection will be *i*) at least 19 EEG electrodes, *ii)* at least 4 minutes of eyes closed data. Requirements for data quality (*e.g.*, signal-to-noise ratio, SNR) and for exact clinical and neuropsychological inclusion/exclusion criteria will be determined during the process (D5.1 patient recruitment and inclusion/exclusion criteria, D2.3 standardisation and preprocessing of the data).

1. 3.1.6 Retrospective neuropsychological data

Retrospective neuropsychological data will be harmonised between all clinical sites. Since cognitive domains have been examined with different tasks among the sites, the neuropsychological evaluations will be harmonised on the basis of standardised scores per task. According to the performance in any cognitive domain, irrespectively of the test used, an equivalent score will be assigned according to standardised national MCI criteria.

3.2 MEG

MEG is available at two of the clinical sites, Helsinki and Madrid. At these sites, simultaneous M/EEG recordings will be conducted at two visits for a subset of the patients (see Table 1). The EEG recordings will be conducted with the MEG-compatible versions of the ***eego***^{TM} system (see above). For details on the MEG compatible study setup, see **Appendix D**.

3.2.1 MEG Hardware and Equipment

MEG is recorded with a 306-channel system (***Triux*** in Helsinki, ***VectorView*** in Madrid) with 102 triple-sensor elements based on SQUID-sensor technology, each element containing two gradiometers and one magnetometer. The systems are produced by the same manufacturer, their sensor array configurations and layouts are similar, but they may differ in their inherent noise-level of the sensors. Continuous head position tracking is available for offline compensation of head movements at both sites. The main measurement-related specifications of the systems are depicted in **Appendix B**.

3.2.2 Synchronisation between MEG and EEG recordings

The synchronisation of the simultaneous MEG and EEG measurements will be conducted using trigger pulses generated by the MEG systems at given time intervals (e.g., at 1 s, 10 s and 60 s intervals on three trigger channels). The pulses will be collected at the trigger channels of the EEG data. This

procedure will allow for synchronising the EEG and the MEG data to a time precision sufficient for the project goal.

3.2.3 MEG measurement environment

Due to its high sensitivity to external disturbances, MEG data is measured in a magnetically shielded room. All the measurement-related equipment must be non-magnetic (*cf.* MEG-compatible EEG electronics). During the recordings, the participant is monitored using a camera, and a microphone and intercom system are available for contact with the participant. The lighting in the room is adjustable and will be adjusted to the recommendations of the EEG measurements in the project.

3.2.4 MEG acquisition software and settings

The MEG systems use dedicated acquisition software that operates on Unix. The acquisition software (versions rel-dacq-6.0.4-bin-1610211219 in Helsinki and 4.8.5 in Madrid) allows for continuous visualisation of raw data, adjustable sampling rate and filtering, monitoring of external trigger inputs and generation of internal triggers to be sent to the EEG system (see 3.2.2). A fixed measurement setup will be used at both sites to assure similar data collection. The MEG measurement settings are depicted in **Table 3**.

Table 3. MEG measurement settings

MEG settings	
Number of MEG sensors	306
Sampling rate	2000 Hz
Filters	0.01-660 Hz
Number of EOG channels	1
Number of ECG channels	1
Number of HPI coils	5 (Helsinki)/4 (Madrid)
cHPI	ON
Trigger channels (STI)	3 (for MEG-EEG synchronization)
Internal stimulus generation	ON

3.3 Neuropsychological testing and other textual data

3.3.1 CANTAB hardware and software

CANTAB is a cognitive assessment system comprising a collection of language-independent, computerised cognitive tests of neuropsychological performance. The cloud-based CANTAB Connect platform is designed to collect the cognitive assessment data in a single electronic data capture system. The CANTAB Connect software will be used to collect data for a predefined clinical assessment battery (see D5.1 for details) with instant scoring and synchronised data transfer.

People with Mild Cognitive Impairment (MCI) compose the clinical target group of AI Mind project. These people are characterised by cognitive decline greater than expected based on the individual's age and education level, and MCI is considered a major risk factor for the development of dementia, especially Alzheimer's disease (AD; Winblad *et al.* 2004). In the present clinical practice, cognitive impairment demonstrated via neuropsychological testing forms the core of memory disorder diagnostics. Computerised CANTAB test batteries have been successfully used to screen AD-typical memory impairment (see, *e.g.*, Junkkila *et al*. 2012).

CANTAB has been validated for use on iPads, and three iPads will be available per clinical site. The CANTAB software is two-fold and includes an *Administration platform* for setting up and monitoring the data collection procedures and a *Rater* application for executing the data collection. The former includes *Study* design, user role configuration, and data export. Fixed measurement setups (*Study* in CANTAB terminology) will be used at all sites by creating five equivalent studies to support the different languages (Finnish, Swedish, Norwegian, Italian, Spanish). All participants will be predefined in the CANTAB system, via the *Administration platform* for the four measurement time points (M0, M8, M16, M24) in pseudonymised form. The iPad *limited access* option will be used to secure the data collection from accidental participant-related mistakes in the software use.

The study reports are generated from each individual *Study* in .csv format. The API integration will automatically pull the data from the CANTAB server into the Electronic Data Capture (EDC) system designated for the AI-Mind project. All data is transmitted in an encrypted format and securely stored within a HIPAA/GDPR and SOC II certified data centre. The locally acquired data is transmitted to the CANTAB server immediately after collection when Wi-Fi is available. The automatic null request from the project's EDC system will be executed in fixed intervals (*e.g.,* once a day). In all the stages of data collection and transmission, study data is only accessible to authorised users per clinical site.

Table 4. CANTAB test selection for AI Mind

Test	Full name	Cognitive domain
MOT	Motor Screening Task	Attention and psychomotor speed
DMS	Delayed Matching to Sample	Memory
OTS	One-Touch Stockings of Cambridge	Executive function
PAL	Paired Associates Learning	Memory
PRM	Pattern Recognition Memory	Memory
RTI	Reaction Time	Attention and psychomotor speed
RVP	Rapid Visual Information Processing	Attention and psychomotor speed
SWM	Spatial Working Memory	Executive function
MTS	Match to Sample Visual Search	Attention and psychomotor speed

3.3.2 Other NPT or textual data equipment

Textual data, including demographic information (*e.g.*, age, sex, education) and clinical variables (*e.g.*, medication, comorbidities), will be collected in a standardised form implemented in the *Nettskjema* service provided by the University of Oslo. The *Nettskjema* service securely transfers the entered information directly to separate site-specific directories for storage in TSD. The data are stored pseudonymised in incremental text files.

Scores obtained from the classical neuropsychological tests (see D5.1) conducted at the clinical sites with paper notes will subsequently be digitised using a standardised *Nettskjema* form.

For each participant visit, two *Nettskjema*-based log forms will be filled in: *i*) a technical log detailing EEG and CANTAB-related issues, and *ii*) a clinical status log, used for registering any change in the participant's clinical status.

3.4 Blood sampling for APOE genotyping and P-tau181

3.4.1 Overview

The AI-Mind project will measure genetic risk variants for AD (APOE genotypes) and plasma biomarker (*e.g.*, P-tau181, or other phosphorylated forms of tau protein), to be taken into account in the AI-based risk prediction. Carrying the APOE ε4 allele belongs to the best-known genetic risk factors of AD (Farrer *et al.* 1997) and is associated with increased risk of cerebral amyloid angiopathy (Corder *et al*. 1997); P-tau181 has recently been associated with subsequent development of AD in both cognitively unimpaired and in patients at MCI state (Janelidze *et al.* 2020).

Blood sampling of all participants will be performed locally, at their first visit (M0), according to the institutional routine clinical procedures. Samples will be processed immediately after collection, stored locally and shipped in batches to the study biobank at Oslo University Hospital.

3.4.2 Sample management and shipment

Samples are processed locally according to the AI-Mind Standard Operation Procedures (SOP) defined in **Appendix E.** Briefly, plasma (EDTA) tubes are handled according to the SOP and centrifuged at 2000 g for 10 minutes at +4C. After centrifugation, 1.0 ml of plasma is aliquoted into each of five 1.5ml polypropylene tubes, and stored (within 30–60 min of collection) at -80°C. Whole blood (EDTA) tubes for DNA extraction (not centrifuged) are stored in cryoboxes at -20°C or lower.

At each stage of the process, the local staff ensures that the unique ID number of the participant, as well as relevant information in the sample data worksheet, are stored with the samples. Samples are shipped in batches packed in dry ice and labelled appropriately to ensure proper handling by the courier. The receiving party at Oslo University Hospital must be notified of the shipping.

The staff who process the samples at each site are responsible for ensuring that samples are processed, labelled, and stored according to the AI-Mind SOP and health and safety guidelines..

4 Data collection procedure

4.1 Harmonized EEG and MEG data collection procedure

The harmonised EEG/MEG data collection protocol includes SOPs for the recordings, including procedures for the patient pseudonymisation process and the naming convention of all data files (Section 4.2).

Before the recording, the 128-electrode EEG cap and the needed additional electrodes (ground, vertical EOG, ECG) will be attached to the participant according to the SOP (**Appendix C**). Good contact between the surface of the skin and the electrodes is ensured through standard preparation procedures, with the requirement of impedance < 25 kOhm in 90% of the electrodes and no more than four adjacent electrodes with impedance exceeding this cut-off value. Impedance values at the beginning and end of each recording are saved with the data as default.

Standardised fixed settings in the acquisition software (**Tables 2 and 3**) and standardised montage settings will be used across sites. Preparation for the combined EEG/MEG recordings will follow the same procedures, but in addition, five head position indicator (HPI) coils, one EOG bipolar electrode, one ECG electrode and the separate ground will be used.

Before the recording, a 3D scan of the participant's head will be performed to enable coregistration with and adaptation of MRI templates, in which all electrodes and anatomical landmarks (ears, nose) are visible and marked. A paper mask that covers the face partially may be held up in front of the face to improve pseudonymisation of the 3D scan.

In MEG, an additional digitisation of the HPI coils, anatomical landmarks (nasion/nose, preauricular points/ears), and a subset of the EEG electrodes (20 electrodes included in the 10-20 system) is conducted using a Polhemus digitisation stylus, and stored within the raw MEG data file.

The recording procedure and task will be explained to the participants with written instructions standardised across sites. During the recordings, the participant is seated comfortably in a chair, and instructed to rest and let their thoughts wander, but not to fall asleep. During the eyes-open recordings the participant is instructed to fixate on the cross.

An initial 1-min artefact calibration session is performed during which the participant is asked to perform eye blinks, saccades, and facial muscle movements in a predefined and annotated sequence. The EEG (and concurrent MEG) recordings consist of 3-4 resting-state recordings (5 minutes with eyes open, 5 minutes with eyes closed, 5 minutes with eyes open, and possibly extra 5 minutes of eyes closed based on pilot results, **Table 5**). The final length of the recording will be decided based on the SNR and stability estimates from the pilot recordings across sites.

Table 5. Tasks and order in which they should be performed

EEG tasks	
Artefact calibration recording	1 min
Eyes open	5 min
Eyes closed	5 min
Eyes open	5 min
Eyes closed (to be decided)	(5 min)

4.2 Pseudonymisation process

All participant data will undergo pseudonymisation at the clinical sites, prior to transfer to TSD. The participant numbering system will be based on the EAN-8 system, enabling the generation of barcodes associated with specific participant numbers. The whole 8-digit number is named the *Session ID*, which is a unique combination of three sub-IDs: the *Site ID* (2 digits), the *Participant ID* (3 digits), and the *Visit ID* (2 digits). The final digit is a check digit, used to verify the structure of the *Session ID* stem. *Session ID* numbers will be generated centrally and sent out to the clinical sites. Once a *Session ID* number has been assigned to a site-participant-visit combination, all data will be stored using this ID number: Each dataset from the same measurement day will be assigned the same ID. Strong identifiers, such as the participant's name or birthday, will never be entered into the acquisition software or in any of the test forms.

As outlined in D1.3, the principal investigator at each clinical site is responsible for storing the mapping between participant identity and the assigned ID number, in accordance with the ethical approvals obtained at each site.

The data collector performs a double-check that the *Session ID* is correct when saving any file. To ease the procedure, *Session ID* stickers will be provided to the participant during the data collection visits.

The information contained in the coregistration data (obtained with a 3D camera) and structural MR images could potentially be used to reconstruct a participant's face. To hinder such a possibility, the data will be pseudonymised locally at each clinical site using established defacing algorithms to remove/blur the participant's face. Due to its importance for automated coregistration algorithms, the nose of the participant will be retained in the image when possible.

Session ID stickers are provided to the laboratory or provider that is collecting the blood samples to ensure that the same pseudonymisation procedure is followed. Otherwise, forms provided by the analytics laboratory (OUS) that will conduct the analyses are used to ensure that we comply with their data handling procedures.

4.3 Local data management and quality assurance

EEG files are exported in their native format as .cnt (64-bit) to ensure that no metadata is lost and in Brainvision .eeg format to be compliant with the BIDS standard (Section 2.3). Data and metadata will be stored in native format, and converted to a BIDS-compliant format in the TSD server, using one of many available BIDS converters (e.g. MNE-BIDS, https://mne.tools/mne-bids/stable/index.html). In alignment with the BIDS standard, the naming convention includes tags for imaging modality,

participant, session, task, and run. To ensure that data is compliant with the BIDS format and no data is missing, the BIDS validator (https://bids-standard.github.io/bids-validator/) can run on each dataset.

Local data management includes:

- Quality assurance script and pseudonymisation check at the time of each measurement;
- Defacing of 3D camera scan and MR images (if such are available);
- A python script for packaging the data file and metadata into a zip file;
- Local data storage with backup;
- Upload to TSD (designated Team member, preferably immediately after each measurement).

The local quality assurance script (in Python) for EEG data includes checks for missing metadata, bad channels, bad data segments, inconsistencies/failures in pseudonymisation, and missing data files. If the quality assurance script detects errors in the acquired data, the Data collector should take mitigating actions, such as re-acquiring the EEG after removing the source of the errors.

To ensure data security, access to data at each clinical site will be restricted to the core research team, in alignment with the local ethical approval..

4.4 Data upload

Data upload is the responsibility of the clinical site data steward. Raw data from EEG/MEG, metadata, and the 3D camera scan will be merged in a .zip file using a Python tool designed for the purpose. Data will be uploaded using the TSD Data Portal data transfer tool. Each data collection site will use a unique URL for uploading data to their designated TSD target directory. The solution ensures that data is transferred in an encrypted format. To ensure that the transmission unalters source data files from the local site to TSD, a verification system, for example using checksums, will be implemented.

Data will be uploaded, preferably immediately after the measurement but latest within a week, into the local staging area at the TSD of each site (for details, see Deliverable 4.1).

5 Team composition

The team composition varies between the clinical sites. To ensure high-quality and harmonised data collection across sites, specific tasks will be assigned to team members (**Table 6**).

Table 6. Roles and responsibilities at each site

Role	Responsibilities
Clinical site data steward	Managing and maintaining prospective data, uploading data into TSD, offline pseudonymisation (defacing), offline quality assurance.
Data collector(s)	Collecting EEG, MEG, CANTAB and textual data, online pseudonymisation (entering IDs), online quality assurance.
Source data owner	Responsibility for source data until uploaded to TSD, responsibility for storing pseudonymisation key.

6 Piloting strategy

A finalised data collection protocol will be based on pilot recordings conducted at all clinical sites. Piloting will cover data collection (M/EEG, CANTAB), pseudonymisation, data upload procedures, and blood sample transportation procedures. Minimal requirements for piloting at each site are depicted here, but additional testing will be performed at each site to ensure data quality in local circumstances.

6.1 EEG data collection training sessions

ANT neuro has provided training sessions at each site.

6.2 EEG/MEG data collection and preliminary analysis

During EEG/MEG piloting, a more extended data collection protocol (altogether 10 minutes eyes open, 10 minutes eyes closed) will be used to ensure the stability of spectral measures and enable a preliminary evaluation of connectivity measures and other extractable features.

At least 5-6 successful pilot recordings will be performed at each site. At least one participant will be measured twice at each site, to ensure the stability of the results. An artefact calibration procedure will be developed in which the participants will be asked to perform a predefined sequence of eye blinks, saccades, and jaw/facial muscle movements. Power spectra (eyes open, eyes closed) will be calculated based on the data from the different sites, and the following quality measures will be assessed: 50-Hz power line noise and 1/f characteristics at all sites, SNR based on occipital alpha (~10 Hz) peak variation (eyes closed *vs*. eyes open) with respect to other frequency bands, and the distribution and number of bad channels.

For MEG, empty-room recordings without a participant will be performed in Helsinki and Madrid, to compare the noise levels at the two recording sites.

6.3 Data pseudonymisation and upload

The data pseudonymisation and data upload processes will be piloted to ensure that the participants are assigned with the same IDs across all data collection procedures (EEG, CANTAB, textual data, etc.), and that no names are ever entered into any acquisition software. For this purpose, a phantom subject will be created at each site, and appropriately pseudonymised data from EEG, MEG, 3D camera and metadata sources will be merged. Data upload procedures for the merged (phantom) data are verified.

6.4 Blood sample shipping

All stages of the blood sample transportation procedures are verified through a phantom data transport.

7 Risk management

Risk factors related to the data collection procedures at the local sites have been identified, and measures to manage the risks or mitigate their impact have been suggested and implemented, as outlined in **Table 7**.

Table 7. Risk factors and mitigation measures

Risk factors	Description	Mitigation measure
Pseudonymisation failure	The identity of the participants is entered into acquisition software, or IDs are mixed up between participants.	Standardised data collection procedures (Appendix C, D), a centralised procedure for generating IDs and participant ID stickers (Section 4.2). EAN-8 system with a check digit will be employed.
Storage hardware failures or accidental deletion	Data is lost due to hardware failure at local site, or due to deletion of data files before upload.	Recommendations for local raw data storage with backup, and weekly (minimum) upload to TSD. Backup at TSD. Restricted permissions to access raw data files (Section 4.3, Deliverable D4.1).
Lack of documentation & metadata	Data cannot be trusted due to mislabeling or lack of documentation.	Standardised data collection procedures, imaging metadata fulfills BIDS standard, python scripts for checking quality and merging data (Section 4.3).
Missing files	Data cannot be used due to missing part of the files.	Standardised data collection procedures, check-list for data collection, weekly upload to TSD.
Poor data quality	EEG data can not be used due to too much noise, bad channels, or movement artefacts.	Limit for acceptable impedance values, annotations during the recording, assessment of EEG data quality immediately after collection, providing the opportunity to re-record.
Lack of information about data quality	Data is not informative due to lack of knowledge of its quality.	Online annotations of EEG data included in data files, data quality assessment results included in metadata.
File format obsolescence, or data loss during format conversion	Data cannot be accessed due to lack of software for reading the file format, metadata is lost when converting to another file format.	EEG data exported in both original .cnt format, and in the BIDS-standard compliant .eeg format. Both data formats are stored in line with BIDS recommendations. A library provided by ANT neuro for reading .cnt files into python will be tested and used. MEG data exported in .fif format, which adheres to BIDS standards. Textual data converted to and stored in .csv or other standard format.
Freezer failure or power cuts affecting stored blood samples	Blood samples are lost before shipping due to freezer failure.	The local freezer should be controlled by an alarm-system. Power must be provided by an emergency back-up electrical supply in case of a power cut.
Blood samples are processed wrong	Blood samples are lost due to problems with processing.	Standard operating procedures provided by the analytics laboratory (Appendix E).

4. References

Corder, E.H., Saunders, A.M., Risch, N.J., *et al*. Protective effect of apolipoprotein E type 2 allele for late onset
Alzheimer disease. *Nat Genet* **7**: 180–184 (1994).

Creutzfeldt OD. Cortex Cerebri, 1983. Berlin: Springer-Verlag.

Farrer, L.A., Cupples, L.A., Haines, J.L., *et al*. Effects of age, sex and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *JAMA* **278**: 1349-1356 (1997).

Gorgolewski, K., Auer, T., Calhoun, V. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* 3, 160044 (2016). https://doi.org/10.1038/sdata.2016.44

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J and Lounasmaa OV. Magnetoencephalography – theory, instrumentation, and applications to non-invasive studies of the working human brain. *Rev Mod Phys* **65,** 413–497 (1993).

Janelidze, S., Mattsson, N., Palmqvist, S. *et al.* Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nat Med* **26**, 379–386 (2020). https://doi.org/10.1038/s41591-020-0755-1

Niso, G., Gorgolewski, K., Bock, E. *et al.* MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Sci Data* **5**, 180110 (2018). https://doi.org/10.1038/sdata.2018.110

Pernet, C.R., Appelhoff, S., Gorgolewski, K.J. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci Data* **6,** 103 (2019). https://doi.org/10.1038/s41597-019-0104-8

Winblad, B., Palmer, K., Kivipelto, M., *et al.* Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Int Medicine* **256***,* 240-246 (2004).

5. Appendix A. *eego*™ amplifier and *waveguard*™ cap technical specifications

Amplifier models	EE-224 and EE-225; cascaded setup EE-228
Number of referential channels	128 actively shielded inputs; separate reference and patient ground
Number of bipolar channels	2 or 6 (MEG-compatible devices), included in EE-225 amplifier
Referential input signal range	150 – 1000 mVPP (programmable gain)
Referential input impedance	> 1 GOhm
Bipolar maximum sampling rate across all referential channels	16384 Hz per referential channel
Resolution	24 bit
Trigger input	8 bit TTL
USB output interface	USB 2.0 compatible, electrically separated
Battery	Integrated rechargeable Li-batteries
Operational time	Up to 5h, charging time 2h

Cap electrodes	128 integrated and wired silver-silver chloride (Ag/AgCl) electrodes, ground (GND) and CPz reference, 3 ring electrodes for GND, VEOGL (vertical eye movements), and CLAV (ECG)
Cap sizes	S = 47-51 cm, M = 51-56 cm, L= 56-61 cm

6. Appendix B. Related technical specifications of *Triux* (MEGIN) and *Vectorview* (Elekta Neuromag) systems

Sensor units	102 identical units with two orthogonal planar gradiometer flux transformers, one magnetometer flux transformer, and three dc-SQUIDs (Superconducting Quantum Interference Devices)
Number of bipolar channels	12 bioamplifier channels; optionally 32, 64, or 128 EEG channels
Number of auxiliary channels	10
Head Position Indicator (HPI)	5 (Helsinki) or 4 (Madrid) marker coils are attached to the head of the participant; detection of signals by MEG sensor array. Determination of position of the marker coils with respect to a participant frame of reference obtained with a 3D digitiser
3D digitiser	Polhemus Fastrak electromagnetic digitiser
MEG channel electronics	DC coupled Resolution 24 bits/sample Adjustable sampling rate (1-5 kHz), adjustable high-pass cutoff at DC-10 Hz, adjustable low-pass cutoff (predefined values 330 Hz — 3,3 kHz)
Dewar	Helmet-shaped
Trigger input	16 TTL

7. Appendix C. Standard Operating Procedures for EEG recordings

Getting started

- Start up the ***eego™*** acquisition software
- Take the amplifiers from the chargers and bring them into the measurement room
- Connect amplifiers according to Figure 1
- Always check at this point from the battery indicators in the software that the amplifiers are charged, and if not, recharge them while preparing the participant
- Start the acquisition workflow by pressing Acquire in the upper left corner
- Add five (5) New Subjects for each participant (to ensure BIDS-compatible naming conventions):
    - Enter *Session ID* as First Name. Double-check that the ID number corresponds to the participant's ID sticker.
    - Enter different Last Names for each run:
      Artefact, for artefact calibration recording; *EO-01* and *EO-02* for eyes open conditions (two runs), *EC-01* and *EC-02* for eyes closed (two runs)
    - Never enter the participant's real name, nor additional personal information into the acquisition software.
- Use the predefined measurement setup, and check that your Recording Parameters are:
    - Sampling Rate: 2 000 Hz
    - Amplifier Setup: AIMind-standard-EEG128-BIP24-CA214, Setup code:228
    - Montage Setup: AI-Mind-standard-10-5-referential
    - Annotations and Online Review are turned on
- Keep the amplifiers turned off while starting to prepare the participant in order to save battery

In Figure 1 (below):

- The EEG data acquisition laptop computer is connected with a USB cable to a local fibre optic extender (Local Extender, LEX).
- The LEX is connected with an orange optic cable to a remote fibre optic extender (Remote Extender, REX).
- The REX provides USB cable connections to the amplifiers situated next to the participant.
- The LEX is powered by the computer via the USB cable used also for data transfer, and the REX is powered by a rechargeable power bank (Powerbank).
- Each of the amplifiers must be connected to the REX separately with a USB cable.
- For recording bipolar channels, the Sensebox needs to be connected to the amplifier with a labelled input for it (*i.e.*, the amplifier with Product Code EE-225). The Sensebox is powered via a dedicated USB-type cable, either from the Powerbank or from a separate rechargeable battery (recommended by the manufacturer).
- Each of the four labelled cables (labels A1, A2, B1, B2) of a 128-channel EEG cap needs to be connected to a dedicated connector in the amplifiers.

- For a 128-channel EEG recording, the two amplifiers (with Product Codes EE-225 and EE-224) must be connected together with the *Cascading trigger adapter*, which can be powered from the REX via a USB cable.

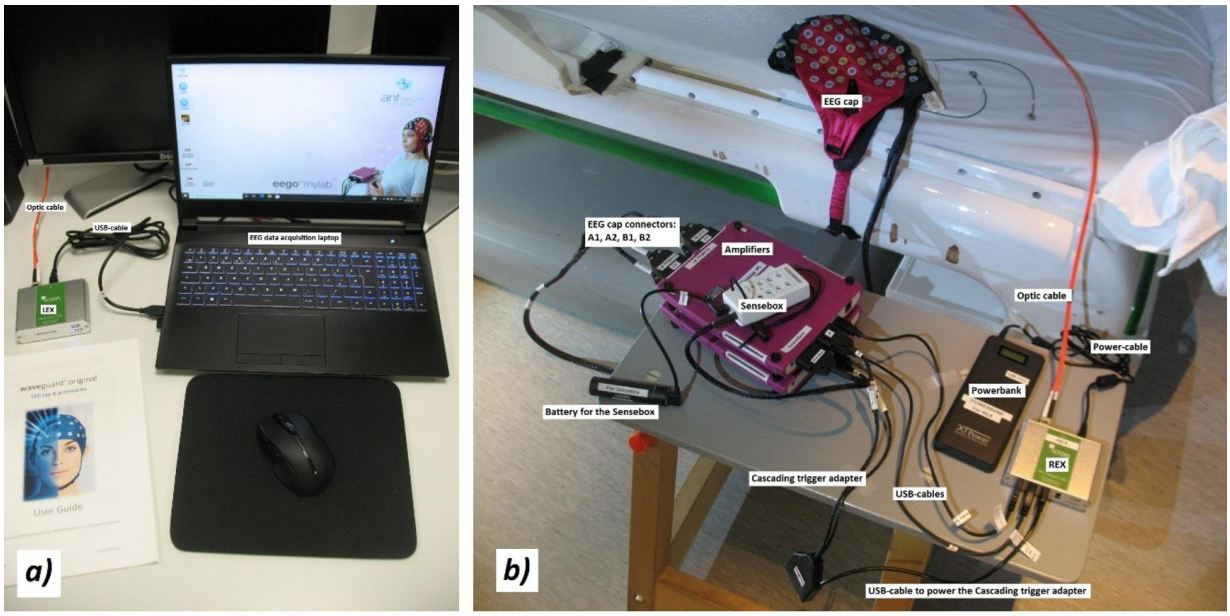

Figure 1. a) The EEG data acquisition laptop. b) Amplifier connections.

Preparing the participant

- Measure the head circumference and find the correct cap size for the participant (see Appendix A for the cap sizes).
- Measure D=nasion-inion, and place the cap so that the distance from nasion to Fpz is 10% of D. Ensure that the cap fits smugly.
- Use abrasive gel and a blunt wooden stick to (gently) scrape the surface of the skin and move hair aside from underneath each electrode (optional, but recommended).
- Fill each electrode with gel using a blunt syringe, avoid excessive gel usage so that a conductive bridge between neighbouring electrodes is avoided.
- Attach the non-cap electrodes: clean the skin gently with alcohol, apply gel to the electrode and use a piece of skin sensitive tape to attach the electrode to the skin.
  Ground electrode: attach the electrode to the left mastoid.
  ECG electrode: attach the electrode to the clavicle bone (right).
  Vertical EOG electrode: place the electrode below the left eye, directly beneath electrode AF7.
- Attach the cap to the amplifiers, then turn on the amplifiers for impedance measurements.
- Check electrode impedances. Re-gel/scrape gently if necessary.
- Electrode impedance is acceptable when below 25 kOhms.
- The overall quality is acceptable, if up to 13 (10%) of the electrodes are above 25 kOhms, and up to 4 of those electrodes are adjacent to each other.

3D camera scan

- Perform a 3D camera scan of the cap with the 3D camera (Occipital Structure Core).
- Let the participant hold up a paper mask in front of their eyes, making sure that the nose is visible.
- Move around the participant in a circle in steps, and take pictures at each step
- Check the resulting 3D image.

Recordings

- Instruct the participant before the recording with the standardised AI-Mind instructions.
- Adjust the lighting in the room, and give/place earplugs on the participant if you measure in a noisy environment.
- Each run (see above) should be recorded in a separate file; the correct naming conventions are obtained by utilising the predefined New Subjects. The tasks are recorded in the following order:
    - 1-min artefact calibration task
    - 5-min eyes open with a fixation cross
    - 5-min eyes closed
    - 5-min eyes open with a fixation cross
    - (5-min eyes closed)
    - If data quality does not pass quality check: an additional 5-min eyes open/eyes closed recording should be conducted
- During the recording, online annotations can be marked for the following events:
    - EO / Eyes Open
    - EC / Eyes Closed
    - Head Movement
    - Movement
    - Swallow
    - Cough
    - Speech
    - Artifact
- For each recording:
    - Instruct the participant
    - Choose a New Subject for each run, using the predefined Subject definitions
    - Turn on recording
    - Make annotations
    - Stop recording after each 5-min recording
- Saving data
    - Export data as native .cnt (64-bit) data, as well as Brainvision .eeg data
    - Do **not** turn on post-processing (filtering) when exporting the data
- After the recording
    - Tell the patient that the recording is over after you have done the quality check
    - Help the participant to remove the cap
    - Turn off amplifiers and connect them to the chargers
    - Turn off/unplug power banks and connect them to the chargers

- Clean the cap and hang it to dry in a well-ventilated place.

8. Appendix D. Solutions for concurrent EEG/MEG recordings

To ensure MEG-compatibility, two systems (Helsinki, Madrid) with a custom-made solution for the *eego* amplifier and the *waveguard* cap were provided by ANT neuro. The solution includes a custom-made magnetically shielded amplifier box using three layers of mu metal foil as shielding, a trigger system for synchronising between MEG and EEG, special adapter cables from the amplifier box to the cap, and a possibility to remove the ferrites from the EEG cap for the MEG measurement.

Setting up EEG for a concurrent MEG/EEG recording:

- Take the amplifier box from the chargers, and bring it into the measurement room. Place it safely as far from the MEG sensors as possible.
- Connect the orange optic cable from the local USB extender (LEX) to the data communication port on the amplifier box.
- Connect three (3) black fibre optic cables to trigger channels 1, 2, and 3.
- Starting up the amplifier box: Turn on the power banks (push *activate* button), connect the power banks with the remote USB extender (REX) with the 1/0 -switch at the back, and turn on both amplifiers (A and B: round white buttons at the back). Check that the green light is on in both amplifiers.
- In case there are ferrites around any EEG cap cables, remove the ferrites from the EEG cap cables using a specific plastic tool. *Suggestion*: If possible, always use the same EEG caps in the MEG environment, to minimise the need to plug-in and remove the ferrites from the EEG cap cables. This minimises the risk of breaking the plastic parts of the ferrites.
- **Always use special adapter cables** when connecting the EEG cap to the amplifier box in order to detach the active shielding in the cap, thus avoiding harming the MEG device.

Concurrent EEG/MEG recording:

- Preparation specific to MEG:
  - Attach an extra bipolar EOG channel: place one electrode below the left eye, the other above the right eye
  - Attach an extra ECG channel
  - Attach 5 (or 4) head position indicator coils (HPI) on top of the cap
  - Perform digitisation (in this order) of: anatomical landmarks (nasion, preauricular points), HPI coils, a subset of the EEG electrodes (20 electrodes included in the 10-20 system), and a few extra points including the shape of the nose
- Recording:
  - Turn on the EEG recording before turning on internal stimulus generation from the MEG device: the first trigger will mark the start of the recording
- Charging the amplifiers and the power banks:
  - Charging must always be performed outside the MEG measurement room.
  - Connect an amplifier charger to the 12V DC input, and connect a USB charger to the 5 V DC input (for the power banks).

9. Appendix E. Standard Operating Procedure for Collection of Blood, Processing and Shipment

Version history:
12.5.2021 original draft by Mathias Toft , MD, PhD; Head Research Laboratory Neurological Unit; Oslo University Hospital; University of Oslo; and Ira Haraldsen, MD, PhD; Coordinator, PI, AI-Mind.EU; Cognitive Health Research Group; Division of Clinical Neuroscience; Oslo University Hospital.

Background

The AI-Mind project includes analysis of genetic risk variants for Alzheimer's disease (APOE genotype) and plasma biomarker P-tau181. Blood sampling of all study participants will be performed locally. Samples will be preprocessed immediately after collection, stored locally, and shipped in batches to the study biobank at Oslo University Hospital for further analysis.

Blood collection

Blood samples will be collected from the participants at their first visit in a non-fasting state.

The following blood specimens are collected using vacutainers:

Sample no.	*Specimen*	*Blood volume (ml)*	*BD Diagnostics Tubes*
B01	Plasma (EDTA)	10	Vacutainer K_2EDTA 10 ml
B02	Plasma (EDTA)	10	Vacutainer K_2EDTA 10 ml
B03	Whole blood (for DNA extraction)	3	Vacutainer K_2EDTA 3 ml

- Ensure that the vacutainer tubes are appropriately coded with the study ID number.
- Collect the blood samples according to the institutional routine clinical procedures.
- Tick off the collected samples on the Blood Worksheet (**Appendix F**).
- Record the time of collection on the Blood Worksheet (**Appendix F**).

Processing of blood specimens

Processing of EDTA Plasma tubes (samples B01, B02):

The plasma tubes should be processed directly after being taken.
Invert the tubes gently (8 to 10 times) and then place them in the centrifuge.

- Centrifuge the tubes for 10 minutes at 2000 g (rcf) at +4°C.
- Record the time of centrifugation on the Blood Worksheet (**Appendix F**).
- Label 5 x 1.5 ml polypropylene (PP) cryovials with study ID number.
- Following centrifugation, plasma from both tubes are transferred into one 50 ml polypropylene tube and mixed.
- 1ml of plasma should be aliquoted into each 1.5ml polypropylene tube and stored in a labeled cryobox at -80°C within 30–60 min of collection.
- Record the time of freezing on the Blood Worksheet (**Appendix F**).
- Record the ID of the boxes in which the specimens are stored on the Blood Worksheet (**Appendix F**).

Processing of EDTA Whole Blood tube for DNA extraction (sample B03):

- Do not centrifuge the sample!
- Store the EDTA whole blood tube for DNA extraction (B03) into corresponding cryoboxes at -20°C or lower.
- Record the ID of the boxes in which the specimens are stored on the Blood Worksheet (**Appendix F**).

Sample shipment

- Samples should always be shipped together with the relevant sample data worksheet.
- To ensure safe transit, cryvials and tubes should be placed in appropriate containers.
- Samples should be shipped in suitable (insulated foam) containers, packed in sufficient dry ice to keep them frozen for the duration of the trip (2-3 days)

Samples shipped in dry ice require specific labelling and documentation to be completed. Please check with the courier for details, but as a minimum:

- The shipping container should be marked clearly with:
  - the words 'Carbon Dioxide, ‘solid' or 'Dry Ice'
  - the words 'UN 1845'
  - full name and address of the shipper and the consignee
  - the net quantity of dry ice in the package
- A Class 9 label should be affixed to the container
- All irrelevant labels and marks should be removed

ALWAYS give notice of shipment date to the receiving party before sending samples.

Notes:

- *The staff who processes the samples is responsible for:*
    i. *ensuring that the samples are processed, labeled, documented and stored as per this protocol.*
    ii. *ensuring that the health and safety guidelines pertaining to that particular workplace are adhered to.*
- *The freezer used for storage needs to be in a well-ventilated or air-conditioned room.*
- *Power needs to be provided by the "emergency back-up" electrical supply in case of a power cut.*
- *The freezer should be controlled by an alarm system.*

10. Appendix F. Blood worksheet

BLOOD WORKSHEET

Study ID: _ _ *.____

Year of birth ____

Barcode

Date of collection _ _ - _ _ - _ _ _ _

Time of collection _ _. _ _ **hours**

Name of collector _____

Tick if collected	**Cryovials/cap**	**ID storagebox**
□ **sample B01-2:** Plasma 1 ml	**5** vials	_____
□ **sample B03:** K_2E K_2EDTA vacutainer, 3 ml	**1** tube	_____

Start time of centrifugation samples B01, B02: _ _. _ _ **hours**

Freezing date and time samples: _ _ - _ _ - _ _ _ _ / _ _._ _ **hours**

Samples processed by (name):