VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease

Public deliverable report

M24 deliverable 3.5: Full-cohort MEG time series in BIDS share and analysis pipeline ready.

Date	17 May 2021
Authors	Isabel Suárez, Gianluca Susi, Fernando Maestú (**UCM**) Ehtasham Javed, Satu Palva, Matias Palva (**UH**) Petra Ritter (**CHARITE**)
Dissemination level:	**PUBLIC**
Website	www.VirtualBrainCloud-2020.eu

This project has received funding from the **European Union´s Horizon 2020** research and innovation programme under **grant agreement No 826421**

Table of content

1. Background 3
2. Abstract 3
2.1 FDMC data overview 3
3 Deliverable preparation 5
3.1 Partners involved 5
3.2 Group meetings relevant to this deliverable 5
4 Description of the data included in the FDMC 6
4.1 Description of the sample 6
4.1.1 Cross-sectional and longitudinal data 7
4.2 Data acquisition 7
4.2.1 MEG recordings 7
4.2.2 MRI scans 8
4.2.3 Additional data 8
5 Description of work performed 11
5.1 Database implementation 11
5.1.1 Data storage 18
5.1.2 Data protection and measures adopted to prevent misuse 18
5.2 MEG pipelines 20
5.2.1 UCM pipeline 20
5.2.2 UH Pipeline 22
5.3 MRI processing 24
6 Conclusions 25
7 Glossary 25
8 References 27
ANNEX 1: *Joint Controllership Agreement* 28

1. Background

The overarching goal of The VirtualBrainCloud (TVB-Cloud) is personalized prevention and treatment of dementia. To achieve generalizable results that help individual patients, the VirtualBrainCloud integrates the data of large cohorts of patients and healthy controls through multi-scale brain simulation using The Virtual Brain (or TVB) simulator. There is a need for infrastructures for sharing and processing health data at a large scale that comply with the EU general data protection regulations (or GDPR). The VirtualBrainCloud consortium closes this gap, making health data actionable. Elaborated data protection concepts minimize the risks for data subjects and allow scientists to use sensitive data for research.

2. Abstract

Data Processing, Standardization, and FAIRification are key elements to the Virtual Brain Cloud (TVB-Cloud) project as they contribute to establishing a link between the clinical-related and the modeling WPs. In this context, this public deliverable (**D3.5**) addresses the release of the full dataset of the Madrid cohort (henceforth FDMC) and the pipelines used for its extraction. The FDMC consists of magnetoencephalographic (MEG) data arranged following the *Brain Imaging Data Structure* (BIDS) standard to foster interoperability and tackle the heterogeneity of data organization. It includes data of four different diagnosis groups in the preclinical, prodromal, and clinical stages of the Alzheimer's disease (AD) continuum: healthy controls (HC), subjective cognitive decline (SCD), mild cognitive impairment (MCI), and AD patients.

The present public report is divided into two main sections:

- Presentation of the **FDMC**. Description of the MEG data and metadata provided for a total of N_{TOT} = 364 participants (HC (119), SCD (88), MCI (142), AD (15)). A set of the MCI cohort includes additional longitudinal recordings for N_{LONG} = 40 participants that were scanned twice with a 2.5 years inter-scan interval (longitudinal subjects). During this period, half of the participants converted to AD (*progressive MCI*), and half remained as MCI (*stable MCI*). For the most of subjects, MEG data is complemented by structural brain data; neuropsychological scores and genetic information are also available for the longitudinal subjects. Whenever brain structural data is available, it is appropriately formatted to match The Virtual Brain (TVB) required formats. The FDMC is safely stored in a dedicated database, residing on a UCM server, with security measures corresponding to the maximum risk level.

- Presentation of the **analysis pipelines** used to process the MEG data into the outputs, metadata, and relevant annotations that constitute the FDMC.

2.1 FDMC data overview

The **FDMC** includes relevant data for older adults pertaining to four different diagnosis groups: HC, SCD, MCI, and AD (see **Table 1**). MEG data and metadata are formatted following the BIDS standard, appropriately modified to support MEG derivatives including source-space aggregated time-series. TVB-compliant structural, genetic, demographic, and neuropsychological data is also provided for a part of the cohort. All data is presented in pseudonymized form.

Table 1. Summary description for the FDMC.

	HC	SCD	MCI	AD
N	119	88	142	15
Gender (females)	79	70	93	7
Age (years)	70.29 ± 4.38	72.34 ± 5.21	73.40 ± 5.44	76.73 ± 5.20
MEG	119	88	142	15
T1-MRI	119	88	142	15
dw-MRI	107	80	122	11

Table 2. Summary numbers for the longitudinal set of the FDMC.

	***Stable* MCI** [MCI → MCI]	***Progressive* MCI** [MCI → AD]
N	20	20
Gender (females)	12	12
Age (years) *	71.05 ± 4.96	74.15 ± 3.66
MEG	20	20
MRI *	20	20

* at the first scan.

For a *full participant* (i.e., a participant for whom every piece of data is available) the following data is present in the FDMC:

- **MEG time series and derivative data**. MEG source-space time-series aggregated in ROIs and metadata as obtained from the pre-processing and source-reconstruction pipelines (including information from the recording setup, artifact removal, independent component analysis (ICA), data segmentation, headmodel, leadfield, and beamforming filters).
- **TVB-ready data**. Additional data required to run TVB. Provided for those participants with dw-MRI data available. The TVB-ready dataset includes cortical surfaces, region mappings, structural connectivity and tract lengths.
- **Neuropsychological scores** and **genetic** information (for the longitudinal patients), obtained through relevant cognitive tests.

Upon request, the following *additional data* can also be provided for the most of participants:

- **Volumetric data**, including medial temporal lobe volumes and the overall intracranial volume.
- **Neuropsychological scores** and **genetic** information for the *non-longitudinal* patients.

3 Deliverable preparation

3.1 Partners involved

UNIVERSIDAD COMPLUTENSE DE MADRID (UCM)
UNIVERSITY OF HELSINKI (UH)
CHARITE UNIVERSITY MEDICINE (CHARITE)

3.2 Group meetings relevant to this deliverable

- TVB-Cloud *Electrophysiology Group* meeting | May 8th, 2019 (Helsinki). Participants: UH, UCM, UNIGE.
- TVB-Cloud *Electrophysiology Group* meeting | September 5th, 2019 (Barcelona). Participants: UH, UCM, UNIGE.
- Video conference to discuss case-specific legal aspects of data sharing | January 17th, 2020 (online). Participants: UNIVIE, CHARITE, UH, UCM, UNIGE.
- TVB-Cloud General Assembly meeting | February 14th, 2020 (online). Participants: AE, AMU, CHARITÉ, CODEBOX, CODEMART, Eodyne, Fraunhofer, FZJ, IBEC, ICM, INRIA, TP21, UCM, UH, UNIGE, UNIVIE, UOXF.
- Video conference to discuss case-specific aspects of data sharing | April 27th, 2020 (online). Participants: UNIVIE, UH, UCM, UNIGE.
- Video conference to discuss scientific aspects of the *Electrophysiology Group* | May 14th, 2020 (online). Participants: UH, UCM, UNIGE.
- TVB-Cloud *Electrophysiology Group* + CHARITÉ virtual meeting | May 20th, 2020 (online). Participants: CHARITÉ, UH, UCM, UNIGE.
- TVB-Cloud General Assembly meeting | Oct 6th, 2020 (online). Participants: AE, AMU, CHARITÉ, CODEBOX, CODEMART, Eodyne, Fraunhofer, FZJ, IBEC, ICM, INRIA, TP21, UCM, UH, UNIGE, UNIVIE, UOXF, INDOC.

Figure 1. Schematic illustration of the data included in the FDMC. **a)** Resting-state MEG recordings. **b)** T1-MRI + dw-MRI. **c)** Clinical data and neuropsychological scores. **d)** Genetic data.

4 Description of the data included in the FDMC

4.1 Description of the sample

The FDMC includes data from older adults pertaining to four different diagnosis groups: HC, SCD, MCI, and AD. Raw MEG, MRI, clinical, neuropsychological, and genetic data for this sample were acquired between the years 2011-2016 in the Laboratory of Cognitive and Computational Neuroscience (LNCyC) (Center for Biomedical Technology, CTB, Madrid, Spain) and the Hospital Universitario Clínico San Carlos (Madrid, Spain). Participants were recruited from the Center for Prevention of Cognitive Impairment (Madrid, Spain), the Faculty of Psychology of the Universidad Complutense de Madrid (Madrid, Spain), the Neurology and Geriatrics Departments of the Hospital Universitario Clínico San Carlos, and the Seniors Center at Chamartín District (Madrid, Spain). FDMC data have been pooled from the databases of two projects supported by the Spanish Ministry of Economy and Competitiveness (PSI2009-14415-C03-01, PSI2012-38375-C03-01). All participants were between 65 and 80 years old, right-handed, and native Spanish. Data was collected with the informed and explicit consent of each participant, after being provided with all the information required to ensure transparency and respect for the rights of the interested parties.

During the initial screening session, the following domains were exhaustively assessed for every participant: (1) memory, (2) language, (3) executive function, (4) cognitive status, (5) subjective memory, and (6) functional capacity and mood. Memory was assessed with the Digit Span Test (forward and backward) of the Wechsler Memory Scale-III-R (Spanish version), and the Texts of Verbal Memory and the Word List of the Wechsler Memory Scale-III. Language function was assessed with the Boston Naming Test and the Phonemic and Semantic Fluency Tests (Controlled Oral Word Association Test). Executive function was assessed with the Trail Making Test parts A and B. General cognitive status was assessed with the Mini-Mental State Examination. Memory in everyday life was assessed with the Rivermead Behavioral Memory Test, and mood was assessed with the Geriatric Depression Scale-Short Form.

Jessen criteria[1] were applied to determine whether candidates were eligible for the SCD group. Cognitive concerns were self-reported by the candidates during a structured interview with an expert clinician. Definite inclusion in the SCD group was agreed on by multidisciplinary consensus (neuropsychologists, psychiatrists, and neurologists) after discarding possible cofounders of SCD (e.g., medication and psycho-affective disorders). Following the recommendations made by the *SCD Initiative Working Group*, all candidates were older than 60 at the onset of SCD, having it occurred within the last 5 years. MCI diagnosis was established according to the *National Institute on Aging-Alzheimer's Association* (NIA-AA) criteria[2]: (1) self- or informant-reported cognitive complaints; (2) objective evidence of impairment in one or more cognitive domains; (3) preserved independence in functional abilities; and (4) not demented[3]. All participants classified as AD fulfilled the *National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association* (NINCDS-ADRDA) criteria of probable AD[3,4]. This requires patients to meet the clinical criteria for all-cause dementia along with insidious onset; clear history of worsening of cognition by report or observation; and initial and most prominent cognitive deficits that include amnestic presentation and/or deficits in language presentation, visuospatial presentation, and executive function. Inclusion and exclusion criteria are detailed in D10.1: H – Requirement No. 1.

4.1.1 Cross-sectional and longitudinal data

The FDMC includes *cross-sectional data* from N_{TOT} = 364 participants. Importantly, a set of MCI patients underwent a clinical and neuropsychological follow-up every 6 months (N_{LONG} = 40), with a second MEG recording within a ~2.5 years inter-scan interval. Considering the clinical outcomes at the time of the second MEG session, participants were split into two different groups: *stable* MCI (the participants remained as MCI) and *progressive* MCI (the participants had developed AD). The later condition was true for half of the participants at the time of the second MEG session. An exhaustive description of the characteristics of the sample and the study design can be found in [5]. The general numbers for the cross-sectional and longitudinal data are summarized in **Tables 1 and 2**.

4.2 Data acquisition

4.2.1 MEG recordings

MEG data were acquired using a 306-channel (102 magnetometers and 204 planar gradiometers) Vectorview MEG system (Elekta AB, Stockholm, Sweden) placed inside a magnetically shielded room (VacuumSchmelze GmbH, Hanau, Germany) located at the UCM-UPM LNCyC in the CTB (Madrid, Spain).

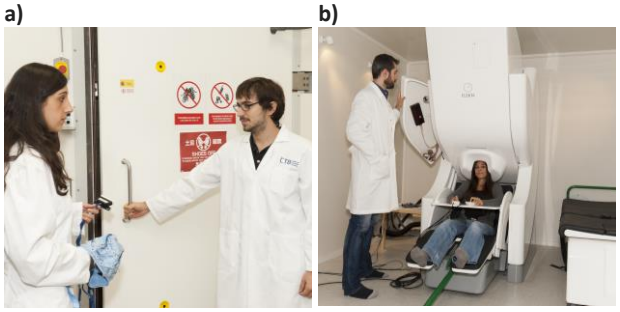

Figure 2. MEG system at the LNCyC. **a)** Magnetically shielded room. **b)** Patient seated on the Vectorview MEG system.

MEG signals were recorded while the participants rested awake, comfortably seated with their eyes closed. All participants were asked to remain relaxed and still, and to keep their eyes closed for the duration of the recording (typically 10 minutes). Whatever participants had been doing before the MEG session varied depending on the project, and could constitute a potential source of unmodelled variance (contributing to e.g., tiredness and nervousness). However, if a participant reported feeling sleepy during the session, enough time was given to them to feel more awake before resuming. Before the MEG recording, the head shape of the participants was digitized using a 3D Fastrak digitizer (Polhemus, Colchester, Vermont). Specifically, three fiducial points were registered (nasion, right preauricular, left preauricular), as well as an outline of approximately 400 scalp points. Four head position indicator (HPI) coils were placed on the participant's scalp (two in the mastoids, and two in the forehead). The position of the HPI coils was also digitized and used to continuously monitor the position and movements of the head during the MEG recordings. To capture undesired ocular activity and blinks, two electrooculogram (EOG) electrodes were positioned above and below the left eye. To capture undesired cardiac activity, two electrocardiogram (EKG) electrodes were set across the chest forming a diagonal in a bipolar montage. MEG data were acquired with a sampling rate of 1000 Hz and an online anti-alias band-pass filter between 0.1 and 330 Hz. Afterward, the recordings were processed offline using a tempo-spatial filtering algorithm (tSSS) (Maxfilter Software v2.2, correlation limit of 0.9 and correlation window of 10 s) to eliminate magnetic noises and compensate for head movements during the recording.

4.2.2 MRI scans

MEG recordings were complemented by MRI scans acquired within a month after the MEG session. MRIs were T1-weighted, recorded in a General Electric 1.5 Tesla magnetic resonance scanner, using a high-resolution antenna and a homogenization PURE filter (fast spoiled gradient echo sequence, with parameters: repetition time/echo time/inversion time = 11.2/4.2/450 ms; flip angle = 12°; slice thickness = 1 mm, 256×256 matrix, and field of view = 256 mm).

Diffusion-weighted images (dw-MRI) were acquired with a single-shot echo-planar imaging sequence with the parameters: echo time/repetition time = 96.1/12,000 ms; NEX 3 for increasing the SNR; slice thickness = 2.4 mm, 128×128 matrix, and field of view = 30.7 cm yielding an isotropic voxel of 2.4 mm; 1 image with no diffusion sensitization (i.e., T2-weighted b0 images) and 25 dw-MRI (b = 900 s/mm2).

4.2.3 Additional data

Genetic data

Episodic memory symptoms seem to be associated with genetic factors such as being a carrier of the APOE-4 allele. In fact, about 30% of the late-onset AD cases have been attributed to the APOE-4 allele in chromosome 19, which increases the person's susceptibility to AD, although it is not itself a sufficient cause of the condition[6]. In this regard, it is of interest to evaluate the impact of APOE-4 and other genes that had already demonstrated a strong relationship with the conversion from MCI to AD.

Table 3. Distribution of the apoE ε4 alleles (34 or 44) in the FDMC for the participants with available genetic information.

	HC	SCD	MCI	AD
N	104	70	118	13
Gender (females)	71	56	74	6
APOE 34 or 44	20	15	46	4

Demographic and neuropsychological data

Demographic variables include age, sex, and years of education. The clinical variables and neuropsychological scores are described in **Table 4**.

Table 4. Clinical and neuropsychological variables present in the FDMC.

Test	Description/Reference
Boston Naming Test	Kaplan *et al.*, 1973.
Boston Naming Test (cue)	Kaplan *et al.*, 1973.
Imitation of postures	Subjects are asked to imitate 8 hand gestures performed by the examiner. The score is the number of correct gestures achieved by the subjects.
Digit Span Test (forward)	Wechsler Memory Scale-III-R (Spanish version; Wechsler, 2004).
Digits Span Test (backward)	Wechsler Memory Scale-III-R (Spanish version; Wechsler, 2004).
Logic. Mem. Imm. Units	Subjects are read 2 short texts and asked to pay attention since they will afterward be required to repeat the information giving as much detail as possible. The second text is read twice. For the Thematic scores (Imm. Thematic and Delay. Thematic) a checklist of 7 general ideas is used. Subjects get 1 point for each of these ideas that they are able to remember from the texts. The final score is a total of 7 points per text (7 x 3 in the Immediate since the second text is read twice, 7 x 2 in Delayed). For the Units scores (Imm. Units and Delay. Units) a checklist of 25 more specific ideas is used. Subjects get 1 point for each of these ideas that they are able to remember from the texts. The final score is a total of 25 points per text (25 x 3 in the Immediate since the second text is read twice, 25 x 2 in Delayed). The Delayed scores are performed 20 mins after reading the texts.
Logic. Mem. Delay. Units	See above.
Logic. Mem. Imm. Thematic	See above.
Logi. Mem. Delay. Thematic	See above.
Trail Making Test A & B	Reitan, 1958.
Change of rules	This test is performed with a deck of poker cards shown by the examiner. Initially, subjects are asked to say YES whenever a red card appears, and NO in the case of a black card. The examiner records the mistakes and the time spent for 20 cards. Afterwards, the instruction is changed and subjects are asked to pay attention to the card that is on the table (the previous card) as much as to the current card. If the color of those two cards matches, they are asked to say YES. If the color doesn't match, they have to say NO. The score is calculated as follows: 0 = > 10 errors; 1 = 7 – 9 errors; 2 = 4 – 6 errors; 3 = 1 – 3 errors; 4 = 0 errors.
FAQ	Pfeffer *et al.*, 1982.
Semantic Fluency Test	Controlled Oral Word Association Test; Benton *et al.*, 1983.

Phonemic Fluency Test	Controlled Oral Word Association Test; Benton *et al.*, 1983.
Mini-Mental State Examination	Folstein *et al.*, 1975.

Volumetric data

A complete set of brain volumes including medial temporal lobe volumes and the overall intracranial volume are also available under request. These volumes were obtained using the FreeSurfer software[7].

5 Description of work performed

5.1 Database implementation

The FDMC is available at the URL https://vbc.ucm.es, mounted on a virtual server of the UCM characterized by 2 cores, 8 GB RAM, and 1 TB Hard Disk (SAS Technology).

Following the BIDS standard, the data is structured in a main folder (`VBC_Madrid_BIDS`), and a derivatives subfolder comprising the derived data generated by the different processing pipelines (`derivatives/pipeline_preprocessing` and `derivatives/pipeline_sources`). In addition, an extra folder (`derivatives/TVB`) includes a set of TVB-compliant structural data and the time-series associated with the ROIs.

The BIDS tree structure of the FDMC is schematized in **Figure 3**.

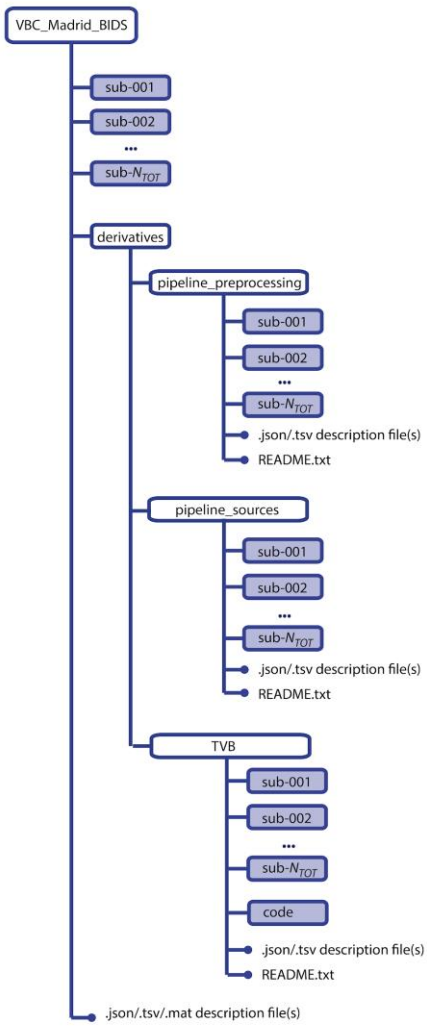

Figure 3. Tree structure of the FDMC arranged following the BIDS standard.

The derivatives subfolder hosts the data and metadata generated by the different processing steps applied to the MEG recordings. Specifically, the subfolder `/pipeline_preprocessing` stores the data obtained during MEG pre-processing (including information from the recording setup, artifact removal, ICA, and data segmentation), while the subfolder `/pipeline_sources` stores the information generated by the source reconstruction pipeline. Regarding the source reconstruction phase, the database includes the outputs generated with the UCM pipeline (see

sect. 5.2.1), while those regarding the UH pipeline can be requested to the related contacts (see sect. 5.2). Both pipelines run on Matlab with the Fieldtrip toolbox[8].

A detailed description of the content of each derivatives subfolder is provided in a README file found at their root. Additionally, a complete overview of the data and metadata included in the FDMC is given in **Table 5**.

Table 5. Description of the data and metadata provided for a *full participant* of the FDMC.

Filename	Short description	Reference n.imaging technique
FOLDER: VBC_Madrid_BIDS/sub-<participant_label>/ses-<session_label>/**meg**		
sub-<participant_label>_ses-<session_label>_coordsystem.json	This file contains the information relative to the **MEG coordinate system**.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_channels.tsv	This file contains the information relative to the **MEG sensors** (type, units, sampling frequency).	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_meg.json	This file contains the information relative to the **relevant acquisition parameters** (sampling frequency, anti-aliasing filtering, institution, instructions given to the participants, etc.).	**MEG**
FOLDER: VBC_Madrid_BIDS/sub-<participant_label>/ses-<session_label>/**anat**		
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_T1w.nii	This file contains the **defaced T1-weighted MRI** in .nii format.	**MRI**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_T1w.json	This file contains the information relative to the relevant parameters of the MRI acquisition protocol.	**MRI**
FOLDER: VBC_Madrid_BIDS/**derivatives**/**pipeline_preprocessing**/sub-<participant_label>/		
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_mixing.tsv	This table contains the **mixing matrix** for the Independent Component Analysis performed using Second Order Blind Identification (SOBI) algorithm. Rows represent the MEG channels; columns represent the Independent Components.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_mixing.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_mixing.tsv*.	**MEG**

sub-<participant_label>_ses-<session_label>_task sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_unmixing.tsv	This file contains the **pseudo-inverse of the mixing matrix** for the Independent Component Analysis performed using Second Order Blind Identification (SOBI) algorithm. Rows represent the Independent Components; columns represent the MEG channels.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_unmixing.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_unmixing.tsv.*	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_epochs.tsv	This file contains the **temporal definition of the epochs**. The following CustomMetadata fields are provided: - 'duration'. Duration of the segmented epochs. In seconds. - 'beg_time'. Beginning time. Time with respect to the start of the recording when the current epoch starts regardless of the padding. In seconds. - 'end_time'. End time. Time with respect to the start of the recording when the current epoch ends regardless of the padding (in seconds). - 'beg_padd'. Beginning time with padding. Time with respect to the start of the recording when the current epoch starts with added padding (in seconds). - 'end_padd'. End time with padding. Time with respect to the start of the recording when the current epoch ends with added padding (in seconds). - 'padd_length'. Padding length. Length of the padding that was added (in seconds). - 'trial_type'. Description of the type of trial.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_epochs.json	This file provides the information required for *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_epochs.tsv*	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-recording_annotations.tsv	This table contains the **annotations of the time interval corresponding to the eyes-closed resting state condition**. Only this time interval was used for further pre-analyses. - *'onset'*. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'. - *'duration'*. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'. - *'label'*. Label for the annotation. It can be a list of annotations separated by a comma: e.g., 'rest_eyesclosed' for eyes-closed resting-state. - *'channels'*. Label specific to a channel. If 'n/a', applies to all the channels.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-recording_annotations.json	This file contains the information relative to the description of the file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-recording_annotations.tsv.*	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-artifacts_annotations.tsv	This table contains the information relative to the **annotations of the resulting definition of the remaining artifacts** after completion of the pre-processing pipeline (including ICA component removal). It includes the information (onset, duration and type) of the artifacted segments (eye_blink, muscle, flux_jumps).	**MEG**

	- *'onset'*. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'. - *'duration'*. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'. - *'label'*. Label for the annotation. It can be a list of annotations separated by a comma: ● 'eye_blink'. Ocular artifact caused by blinking or eye movement. ● 'flux_jump'. SQUID jump artifact. ● 'muscle'. Muscular artifact caused by muscular contraction or swallowing. - 'channels'. Label specific to a channel. If 'n/a', applies to all the channels.	
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-artifacts_annotations.json	This file contains the information relative to the description of the file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-artifacts_annotations.tsv.*	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_annotations.tsv	This table contains the information relative to the **annotations of the resulting definition of undesired components after Independent Component Analysis (ICA)** in both time and frequency domains. It includes the information (type and name of ICA components) of the undesired components - *'onset'*. Onset in seconds from the beginning of the recording. 'n/a' if it affects the entire recording. If 'n/a', duration must also be 'n/a'. - *'duration'*. Duration in seconds. 0 for instantaneous events. 'n/a' if it affects the entire recording. If 'n/a', onset must also be 'n/a'. - *'label'*. Label for the annotation. Can be a list of annotations separated by a comma: ● 'eye_blink'. ICA component caused by blinking or eye movements. ● 'heart_beat'. ICA component of cardiac activity. ● 'noise'. ICA component of undefined noise. - '*channels*'. Label specific to a channel. If 'n/a', applies to all the channels. Else, it must be a list of channel names or labels for the time series (e.g., name of ICA component).	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_annotations.json	This file contains the information relative to the description of the file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-SOBI_annotations.tsv.*	**MEG**
FOLDER: VBC_Madrid_BIDS/**derivatives**/**pipeline_sources**/sub-<participant_label>/		
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_headmodel.zip	This compressed folder contains the data relative to the **triangulated surface mesh** used to provide the geometrical description of the volume conduction model of the head. Coordinates are given in Elekta-Neuromag space. The following files are provided in TVB format: - 'triangles.txt'. Matrix of (triangles x 3) indicating the combinations of indexes from the vertices.txt array forming each triangle. - 'vertices.txt'. Matrix of (NumberOfVertices x 3) indicating the coordinates (x, y, z) of each vertex in a space-separated values file. - 'normals.txt'. Matrix of (NumberOfVertices x 3) indicating the normal vector (x, y, z) to each vertex in a space-separated values file.	**MEG/MRI**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_headmodel.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_headmodel.zip.* The following CustomMetadata fields are provided: - 'HeadmodelType'. Volume conduction model, e.g., 'single-shell',	**MEG/MRI**

	'local spheres'. - 'Tissue'. Tissue that is being modeled, e.g., 'brain'. - 'Unit'. Measure unit (International System units). - 'NumberOfVertices'. Number of vertices used to build the geometrical headmodel. - 'FTCode'. Fieldtrip script used to compute the headmodel - 'FTMethod'. Fieldtrip method used in the call to FTCode.	
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_leadfield.mat	This file contains the forward solution in the volume conduction model, expressed as the **leadfield matrix** (sources x channels x 3) and relevant metadata, in Matlab's .mat format. The leadfield data are found in *grid.leadfield*. The grid used as the sourcemodel is found in *grid.pos*. This file also includes information on the **sourcemodel** used for source reconstruction of the MEG data. It provides a description of the Elekta-Neuromag coordinates of the equivalent current dipoles considered for source reconstruction	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_leadfield.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_leadfield.mat*.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_lcmvfilter.mat	This table contains the **beamformer filters** to estimate the activity at given locations in the brain, computed using Linearly Constrained Minimum Variance (LCMV) method that minimizes the source variance and is subjected to a unit-gain constraint. Data is expressed as a matrix of dimensions (sources x channels). 'n/a' is associated with sources that are not inside the headmodel.	**MEG**
sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band-label>_lcmvfilter.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_task-<task_label>_run-<run_label>_desc-<band_label>_lcmvfilter.mat*. The following CustomMetadata fields are provided: - 'Method'. Method used for the computation of the inverse solution. - 'Regularization'. Regularization factor (%) used in the computation of the inverse solution. - 'Whitener'. Whitener applied to the data to account for differences in sensor type (magnetometers versus gradiometers) (e.g., 'None', 'Scaling', 'PCA').	**MEG**
FOLDER: VBC_Madrid_BIDS/**derivatives**/**TVB**/sub-<participant_label>		
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_SC.zip	This compressed file contains the **structural connectivity** (SC) data. The following CustomFiles files are provided in TVB format: - 'centres.txt'. Matrix (rois x 4) of SC centres. Each row corresponds to the center coordinates of a region. The 1st column is the region label. 2nd to 4th columns are the (x, y, z) coordinates of the centres for that particular region. - 'weight.txt'. Matrix (rois x rois) of SC weights. - 'tract.txt'. Matrix (rois x rois) of SC tract lengths. - 'cortical.txt'. 1 = cortical region. 0 = subcortical region. Logical. - 'hemisphere.txt'. 1 = right hemisphere region. 0 = left hemisphere region. Logical.	**dw-MRI**
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_SC.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_SC.json* The following CustomMetadata fields are provided:	**dw-MRI**

	- 'Atlas'. Parcellation atlas used to parcel the SC.	
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_parceldata.dat	This file contains the matrix representation of the source-space, ROI-aggregated, reconstructed time-series. Data is stored in binary format and can be recovered to .mat formatting using the code provided in the folder *code* (TVB folder).	**MEG**
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_parceldata.json	This file contains the details of the parameter combinations used to generate the file: *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_parceldata.dat* , in addition to the parameters used for the signal acquisition and the following metadata fields: - MatrixDimensions, - Atlas, - FrequencyBand, - FrequencyLimits, - Padding, - PaddingLength, - CodeURL.	**MEG**
sub-<participant_label_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_region_mapping.txt	This file defines a **map** between *sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.zip* and the connectivity. This file holds a vector of length equal to the number of vertices in the cortical surface. The numeric values are in the interval (0, 1, ..., n-1), where n is the number of rois in the connectivity.	**MRI**
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_region_mapping.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_region_mapping.txt.* The following CustomMetadata fields are provided: - 'Atlas'. Parcellation atlas used to parcel the cortical surface.	**MRI**
*sub-<participant_label>_ses-<session_label>_run-<run_label>_**cortical_surface**.zip*	This compressed file contains the **triangulated cortical (pial) surface mesh**. The following files are provided in TVB format: - 'triangles.txt'. Matrix of (triangles x 3) indicating the combinations of indexes from the vertices.txt array forming each triangle. - 'vertices.txt'.Matrix of (NumberOfVertex x 3) indicating the coordinates (x, y, z) of each vertex in a space-separated values file. - 'normals.txt'. Matrix of (NumberOfVertex x 3) indicating the normal vector (x, y, z) to each vertex in a space-separated values file.	**MRI**
sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_cortical_surface.zip.*	**MRI**
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_leadfield_rois.mat	This file contains the forward solution in the volume conduction model, expressed as the **leadfield matrix** (rois x channels x 3) (and relevant metadata) in Matlab's .mat format. The leadfield data are found in grid.leadfield. The grid used as the sourcemodel is found in *grid.pos*.	**MEG/MRI**
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_leadfield_rois.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_leadfield_rois.mat.* The following CustomMetadata fields are provided: - 'Atlas'. Parcellation atlas used to parcel the cortical surface.	**MEG/MRI**
sub-<participant_label>_ses-	This file contains the **beamformer filters** (and relevant metadata)	**MEG**

<session_label>_run-<run_label>_atlas-<atlas_label>_lcmvfilter_rois.mat	used to estimate the activity at the centroids of each roi, computed using Linearly Constrained Minimum Variance (LCMV) method that minimizes the source variance and is subjected to unit-gain constraint. Data is provided in Matlab's .mat format. The beamformer filters for each classical frequency band (e.g., 'Alpha', 'Beta', 'Broadband') are found in band(i).sources.filter. The estimated power of the neuronal sources for each classical frequency band are found in band(i).sources.pow. The frequency limits for each classical frequency band are found in band(i).edges.	
sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_lcmvfilter_rois.json	This file contains the details of the parameter combinations used to generate the derived file *sub-<participant_label>_ses-<session_label>_run-<run_label>_atlas-<atlas_label>_lcmvfilter_rois.mat* The following CustomMetadata fields are provided: - 'Method'. Method used for the computation of the inverse solution. - 'Regularization'. Regularization factor (%) used in the computation of the inverse solution. - 'Whitener'. Whitener applied to the data to account for differences in sensor type (magnetometers versus gradiometers) (e.g., 'None', 'Scaling', 'PCA'). - 'Atlas'. Parcellation atlas used to parcel the cortical surface.	**MEG**

All the participants of the FDMC have at least one session subfolder (i.e., `ses-1`), both in the main folder and in the derivatives subfolders. The subset of participants with longitudinal follow-up presents an additional session subfolder (i.e., `ses-2`) for the second MEG recording.

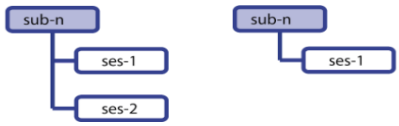

Figure 4. Subfolder structure for a longitudinal (left) and non-longitudinal (right) subject.

For the sake of simplicity, the list of the participants pertaining to the longitudinal study is given at the root of the main folder, with the respective outcome (stable MCI /progressive MCI). Finally, a the demographics of the FDMC participants, the neuropsychological scores of the longitudinal patients, and the additional data for longitudinal participants (see section 3.2.3) is provided in the related BIDS-compliant files.

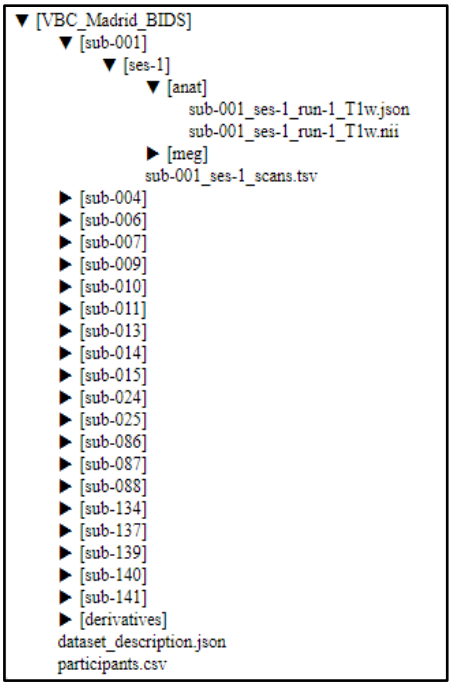

Figure 5. Example structure of the BIDS tree viewer in the UCM portal.

The website embeds an interactive *tree viewer* (see **Figure 5**) to provide the user with a visual reference of the BIDS structure, useful to facilitate the downloading steps. Another separate section is dedicated to the *file manager,* where the user will be able to download the FDMC dataset.

5.1.1 Data storage

In **Table 6,** we indicate the average amount of memory needed for the storage of a complete set of data for a full participant of the FDMC.

Table 6. Data sizes per subject and subfolder, considering a *full participant* of the FDMC.

Subfolder	**MB** (approx)
VBC_Madrid_BIDS/sub-<participant_label>/	50
VBC_Madrid_BIDS/derivatives/pipeline_preprocessing/sub-<participant_label>/	Up to 50
VBC_Madrid_BIDS/derivatives/pipeline_sources/sub-<participant_label>/	Up to 400
VBC_Madrid_BIDS/derivatives/TVB/sub-<participant_label>/	Up to 500
TOTAL	**1 000**

5.1.2 Data protection and measures adopted to prevent misuse

The core aspects of data protection applied to the FDMC reside in the pseudonymization of the data and the implementation of security measures to control data access and download. For an extensive summary of the data-sharing protocols within TVB-Cloud we refer the reader to the Annex procured by UNIVIE in **D3.4**.

- **Data pseudonymization and defacing.** Upon enrolling, participants were assigned a numeric code provided at the medical center where recruitment was carried out. No personal data (name, date of birth, personal ID, personal address, etc.) were collected during the experimental procedure at the UCM; only general demographics. MEG data were labeled using the assigned numeric code. This original numeric code was converted to a new ID (different from the one assigned at the medical center and unique for the

TVB-Cloud project) according to BIDS practice (i.e., sub-XXX). The only link between the original (not provided to partners) numeric code and the personal data (the name and signature of the participants) would be found in the signed informed consents that are safely stored in compliance with National and European Laws. Additionally, T1-weighted MRI data have been defaced to remove identifiable features from the images (see **Figure 6**). MEG raw data is not to be shared, only data that has already been processed to some extent.

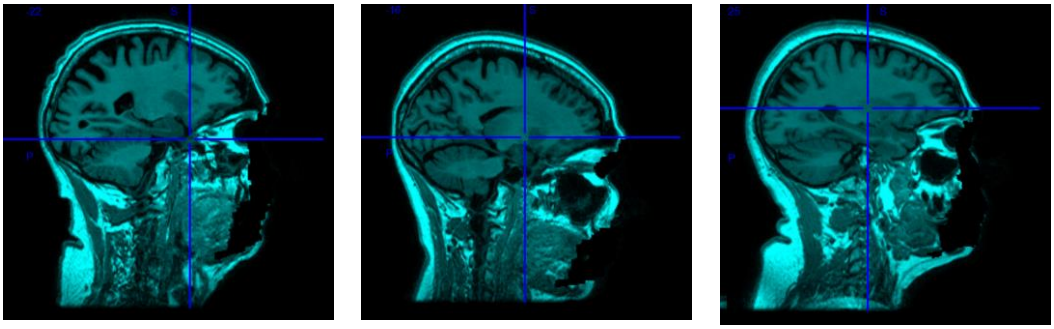

Figure 6. Examples of defaced MRIs.

- **Secure access.** FDMC data is available through a UCM server accessible at the URL https://vbc.ucm.es/login.php. This data will only be available for research purposes within the TVB-Cloud project. The iter to access the FDMC is the following:

  - Through the button *Access Request* (see **Figure 7**), the applicant is able to request access to the co-IP. In order to request access, the applicant must specify *Name, Surname, Institution, and TVB-Cloud WP.*

  - The applicant will receive the documentation required to establish a data-sharing bilateral agreement with the UCM. Note that this step is **mandatory** before any data exchange takes place.

  - Under IP approval, once the agreement has been settled, login credentials will be created and sent to the applicant who will then be able to access the FDMC data through the dedicated button *Login*. Access rights to third parties outside TVB-Cloud will be denied.

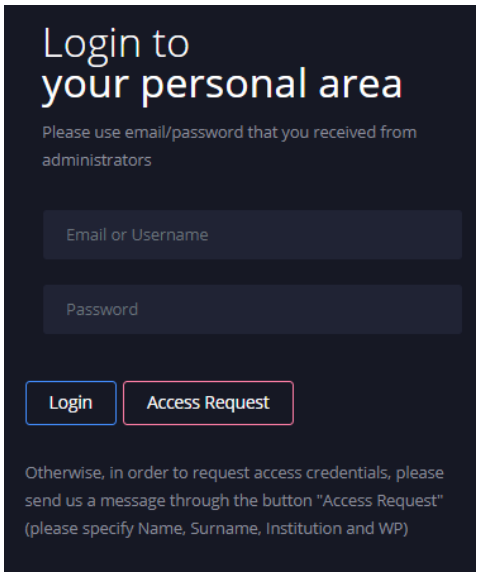

Figure 7. A screenshot of the access request page https://vbc.ucm.es/login.php

Additionally, a record of download requests (*data recipient/data description/exchange date)* will be automatically generated on an electronic document to maintain control over the data flow from the server. Security levels at the UCM server are guaranteed in compliance with Spanish and European Laws, as established by an initial external audit, and follow-up audits carried out by the DPO every 6 months. To keep track of the downloads, the UCM server integrates the above-described log system which allows for automatic tracking of the filename, solicitant name, and timestamp for any file that has been requested for download.

As example, the GDPR-compliant Joint-Controller Agreement among UCM, UH and Charitè is provided as Annex. In addition to this, an additional internal confidentiality agreement between UCM and the other two parties involved has been signed.

5.2 MEG pipelines

In this section, we introduce the two pipelines used for the processing of MEG data, from tSSS to source-space signal reconstruction at UCM and UH. The pipelines are divided into two main procedures: 1) **MEG pre-processing** and 2) **MEG source-reconstruction**. In MEG pre-processing the general steps were agreed by the UH and the UCM (at a group meeting held in Helsinki in May 2019), while in MEG source-reconstruction we have utilized two different methods i.e. Linearly Constrained Minimum Variance (LCMV) beamformer and Minimum Norm Estimates (MNE), for estimation of forward and inverse operators, at UCM and UH, respectively. The code associated with UCM pipeline can be requested by sending an email to co-IP Fernando Maestú: fernando.maestu@ctb.upm.es (CC: isabel.suarez@ctb.upm.es and gianluca.susi@ctb.upm.es) and for UH pipline to co-IP Matias J. Palva: matias.palva@helsinki.fi (CC: ehtasham.javed@helsinki.fi). All the scripts of the pipelines were run on Matlab and use the Fieldtrip toolbox[8] for MEG analysis (https://www.fieldtriptoolbox.org/).

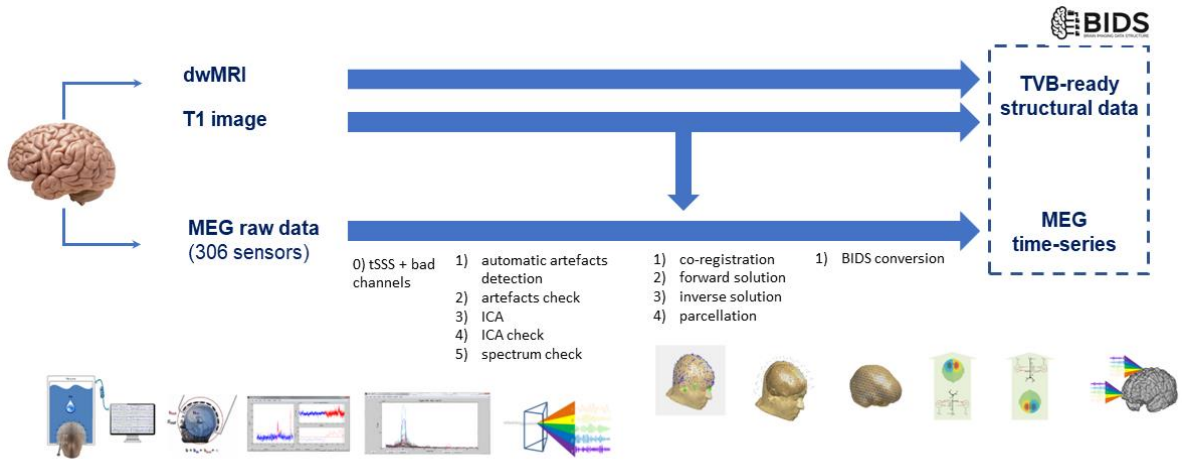

Figure 8. Summary of the steps in the MEG pipeline.

5.2.1 UCM pipeline

PRE-PROCESSING:

- *Automatic artifact detection*. This step uses Fieldtrip's automatic artifact detection algorithm to build a temporal definition of three types of artifacts (blinks, muscular, and SQUID jumps). The input is MEG data after tSSS.

- *Visual artifact inspection*. To correct potential misidentifications and check for undetected artifacts, data is inspected visually. If available, data from the EOG electrodes is presented alongside the MEG channels to help the detection of blink artifacts (see **Figure 9a**).
- *Independent component extraction*. A blind source separation algorithm based on second-order statistics (Second Order Blind Identification, SOBI) is used to extract the mixing matrix to obtain statistically independent components.
- *Independent component removal*. The SOBI components extracted from the previous step are visually inspected to detect EOG and EKG activity. The artifacted segments previously identified are highlighted to facilitate the selection of components of EOG activity. Topographic distributions of the SOBI components are depicted alongside each component's signal to facilitate the identification (see **Figure 9b**).
- *Quality checks*. Following SOBI, the artifact definition might be inspected again, e.g., to recuperate data segments previously labeled as EOG artifacts, now clean after the removal of EOG components. The spectra of data and SOBI components are also visually inspected to check for undetected noisy components (see **Figure 9c**).
- *Data segmentation*. The data is segmented in 4-seconds epochs of artifact-free activity.

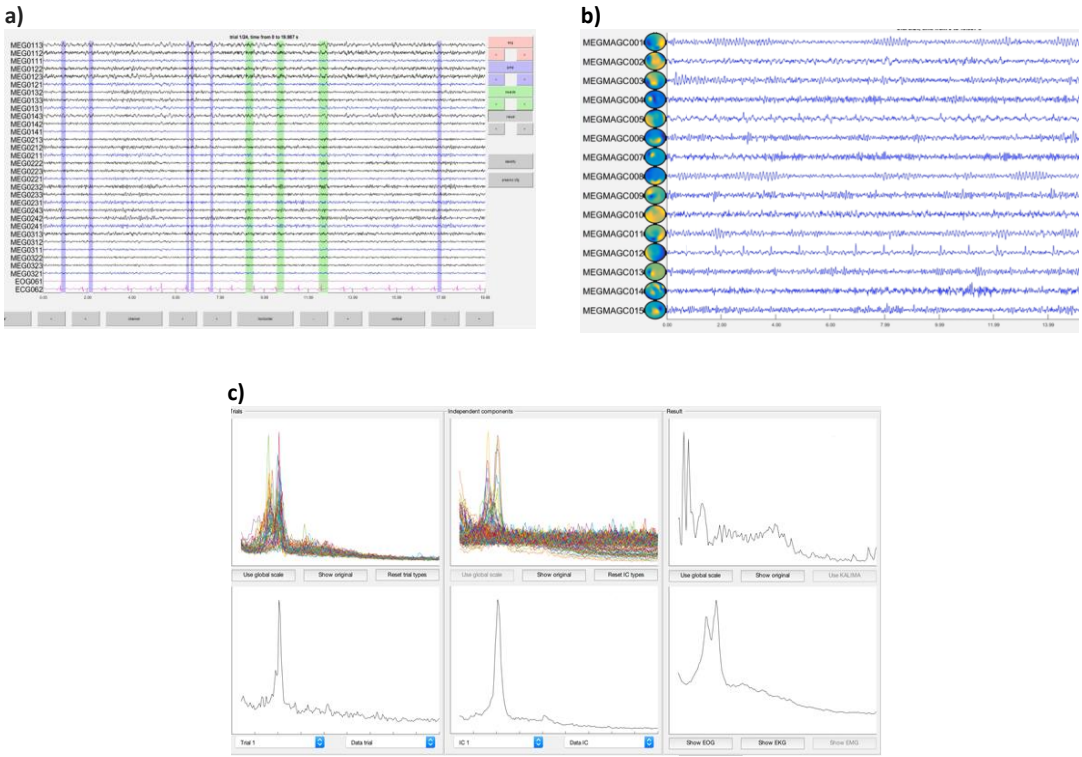

Figure 9. User interfaces of the tools used for: **a)** Visual artifact inspection. **b)** SOBI component identification. **c)** Quality check based on the spectra of epochs and SOBI components.

SOURCE-RECONSTRUCTION:

- *Fiducial identification, segmentation, and mask generation.* Three fiducial points (nasion, right preauricular, left preauricular) and three SPM landmarks are identified in the T1-weighted image. The software SPM12 is used to segment the MRI data into probability maps for the different brain tissues (white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF)). Following the probability segmentation, we obtain a mask for the brain, the skull, and the scalp.

- *Headmodel.* This step provides a single-shell head model with a unique boundary defined by the inner skull (the combination of WM, GM, and CSF) generated from the individual T1-weighted MRI.
- *Sourcemodel*. The source model (2459 sources placed in a homogeneous grid of 1 cm in a Montreal Neurological Institute (MNI) template) is converted to subject space by an affine transformation.
- *Co-registration* between the MEG and the T1-MRI using the fiducial and scalp coordinates digitized during the MEG acquisition, and the fiducial points identified on the T1-weighted image.
- *Forward solution*. The *leadfield* is calculated with the single-shell head model.
- *Inverse solution*. Source reconstruction is performed independently for each subject using a LCMV beamformer. The beamformer filters are obtained for each classical frequency band (*theta* [4 - 8] Hz, *alpha* [8 - 12] Hz, *beta* [12 - 30] Hz, *lower beta* [12 - 20] Hz, *upper beta* [20 - 30] Hz, *gamma* [30 - 45] Hz, and *broadband* [2 - 45] Hz) using the previously computed leadfield, the epoch-averaged covariance matrix, and a 20% regularization factor.
- *Parcellation.* Source time-series are then parcellated using a specific anatomical atlas (e.g., *Desikan Killiany*).

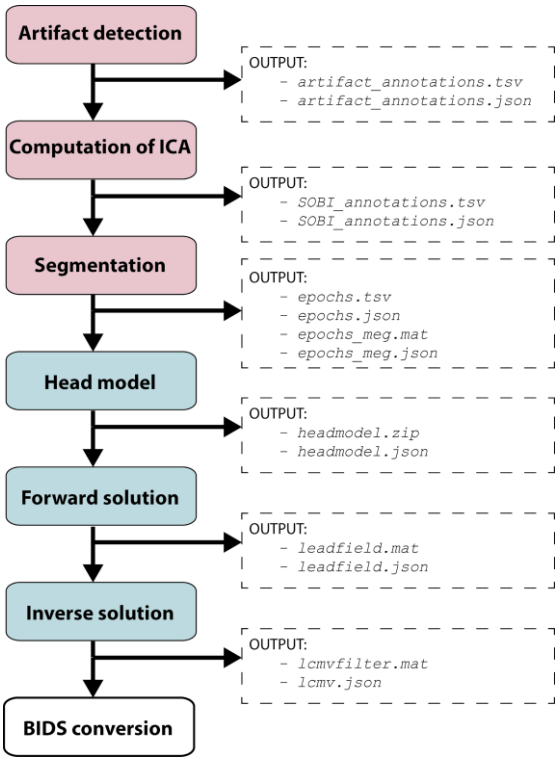

Figure 10. Processing chain and outputs/annotations stored in the BIDS folder. The pre-processing and source-reconstruction steps individuated in the text are here indicated in light red and light blue, respectively.

5.2.2 UH Pipeline

PRE-PROCESSING:

- *Automatic artifact detection*. In this step, Fieldtrip's automatic artifact detection algorithm is used to identify blinks, muscular, and SQUID jumps. The input was MEG data after tSSS. If available, data from the EOG electrodes is presented alongside the MEG channels to help the detection of blink artifacts
- *Visual artifact inspection*. Visual inspection is done to correct potential misidentifications and check for undetected artifacts (**Figure 9a**).
- *Independent component extraction*. Independent components are extracted using Second Order Blind Identification (SOBI) algorithm.
- *Independent component removal*. The components extracted in preceding step are visually inspected to detect EOG and EKG activity. Topographic distributions of the independent components are to differentiate potential non-neuronal signals (**Figure 9b**).
- *Quality checks*. In this step, the spectra of data and SOBI components are also visually inspected to check for undetected noisy segments. Also, segments with sudden jumps are identified which were not removed in any of above steps (**Figure 9c**).
- *Data Interpolation*. In this last step of artifact removal, to recuperate the segments, instead of removing them, identified in 'quality checks', they were replaced with interpolated segments formed using Golay filter and clean data segments.

SOURCE-RECONSTRUCTION:

- *Fiducial identification, segmentation, and mask generation.* Similar to Madrids' pipeline, we obtain a mask for the brain, the skull, and the scalp using three fiducial points (nasion, right preauricular, left preauricular) and three SPM landmarks identified in the T1-weighted image. The SPM12 is used to segment the MRI data into probability maps for the different brain tissues (white matter (WM, GM, and CSF).

- *Headmodel.* A single-shell model with a unique boundary defined by the inner skull (the combination of WM, GM, and CSF) is used to generate headmodel from the individual T1-weighted MRI.

- *Sourcemodel*. White-matter surface is reconstructed using FreeSurfer software. The dipoles are defined at the vertices of white matter surface with ~7mm inter-dipole distance.

- *Co-registration.* The fiducial and scalp coordinates digitized during the MEG acquisition, and the fiducial points identified on the T1-weighted image are used to align MEG and the T1-MRI.

- *Forward solution*. The single-shell head model is used for calculation of leadfield.

- *Inverse solution*. Source reconstruction is performed independently for each subject using Fieldtrip's MNE algorithm. The inverse operator is estimated using the previously computed leadfield, the 4 second epoch-averaged covariance matrix, and a 0.11 regularization factor.

- *Fidelity Weights.* Using forward and inverse operator along with simulated data, fidelity weights for each dipole are estimated to avoid erroneous estimation of source time-series.

- *Parcellation.* The fidelity weighted source time-series are then parcellated into Schaefer's 400 parcels time-series.

5.3 MRI processing

The T1-weighted images were processed using FreeSurfer v.6.0 *recon-all* function that includes motion correction, intensity non-uniformity correction, intensity normalization, segmentation of the different brain tissues, and constructs a cortical surface mesh for each image. This cortical surface mesh is inflated to a sphere and registered to a common surface-space. An anatomical atlas was used to assign neuroanatomical labels to each native brain voxel. Lastly, the T1-space cortical atlas was registered to each subject's dw-MRI space using FSL *flirt* with 7 degrees of freedom.

dw-MRI data was processed using the MRtrix3 software (version 3.0.2)[9] and included the following steps: dw-MRI denoising, Gibbs ringing artifacts removal, eddy current and movement correction, dw-MRI bias field correction, generation of a tissue-type segmented image for anatomically constrained tractography, and the estimation of the WM, GM, and CSF response functions. The single-Shell 3-Tissue CSD (SS3T-CSD) method was applied to obtain WM-like fiber orientation distributions as well as GM-like and CSF-like compartments in all voxels using the MRtrix3Tissue fork (https://3Tissue.github.io). Finally, we performed multi-tissue informed log-domain intensity normalization, and the generation of the tractogram (25 million streamlines, maximum tract length = 250, FA cutoff = 0.06, dynamical seeding).

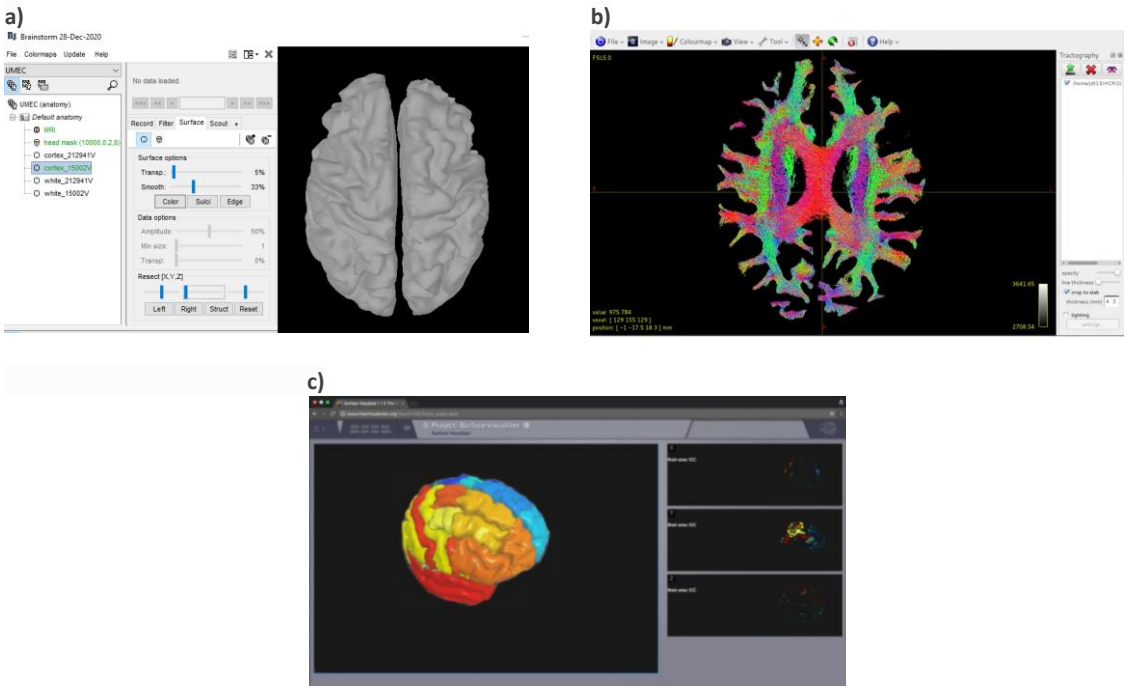

Figure 8. MRI data extraction: **a)** cortical surface, b) tractography data, and c) subsequent import to TVB.

6 Conclusions

In addition to representing a valuable source for the study of neurodegenerative diseases (e.g., unveiling possible disease trajectories), the FDMC dataset can be used to optimize and validate computational models of large-scale brain dynamics in TVB. The dataset is arranged using the BIDS standard, to foster interoperability and to address the heterogeneity of data organization. Importantly, the subset of TVB-compliant data enables personalized simulations, and the FDMC on the whole can be used as a test bench for computational neuroscience methods and machine learning within the TVB-Cloud project.

7 Glossary

AD Alzheimer's disease

APOE Apolipoprotein E

BIDS Brain Imaging Data Structure

CTB Centre for Biomedical Technology

CSF Cerebrospinal fluid

dw-MRI Diffusion-weighted magnetic resonance imaging

EOG Electrooculogram

EKG Electrocardiogram

FAIR Findable, accessible, interoperable, and re-usable

FDMC Full dataset of the Madrid cohort

GM Grey Matter

HC Healthy control

HPI Head position indicator

ICA Independent component analysis

LCMV Linearly constrained minimum variance

LNCyC Laboratory of Cognitive and Computational Neuroscience

MCI Mild cognitive impairment

MNI Montreal Neurological Institute

MRI Magnetic resonance imaging

MEG Magnetoencephalography

NIA-AA National Institute on Aging-Alzheimer's Association

NINCDS-ADRDA National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association

ROI Region of interest

SCD Subjective cognitive decline

SNR Signal-to-noise ratio

SOBI Second-order blind identification

tSSS Temporal signal space separation

TVB The Virtual Brain

UCM Universidad Complutense de Madrid

UPM Universidad Politécnica de Madrid

URL Uniform resource locator

WM White matter

WP Work package

8 References

1. Jessen, F. Subjective and objective cognitive decline at the pre-dementia stage of Alzheimer's disease. *Eur. Arch. Psychiatry Clin. Neurosci.* **264 Suppl 1**, S3–7 (2014).
2. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* **7**, 270–279 (2011).
3. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* **7**, 263–269 (2011).
4. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* vol. 34 939–939 (1984).
5. Pusil, S. *et al.* Hypersynchronization in mild cognitive impairment: the 'X' model. *Brain* vol. 142 3936–3950 (2019).
6. Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
7. Fischl, B. FreeSurfer. *NeuroImage* vol. 62 774–781 (2012).
8. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
9. Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. doi:10.1101/551739.

ANNEX 1

Joint Controllership Agreement

Preamble

The VirtualBrainCloud (TVB-Cloud), a Horizon 2020 research project under grant agreement No 826421 integrates the data of large cohorts of patients and healthy controls through multi-scale brain simulation using The Virtual Brain (or TVB) simulator. There is a need for infrastructures for sharing and processing health data that comply with the EU general data protection regulations (or GDPR). The VirtualBrainCloud consortium closes this gap, making health data actionable. As part of this endeavour, involved partners will need to share personal and sensitive data, thus requiring a contractual arrangement that will enter into force upon signature of the involved parties and remain valid until the end of the project.

§ 1

(1) This agreement determines the rights and obligations of the controllers (hereinafter also referred to as "parties") for the joint processing of personal data. It applies to all activities of the parties, or processors appointed by a party, when processing personal data. The parties have jointly determined the purposes and means of processing personal data in accordance with Article 26 GDPR.

a. The purpose of the joint processing is as follows:
    i. extract personalized brain indices for The Virtual Brain (TVB) parameter validation, using different approaches and methodologies
    ii. multi-perspective analysis for the early diagnosis of AD
b. The means of the joint processing are as follows:
    i. The data transfer will be made using a password protected institutional server (UCM) with download tracking;
    ii. Data protection and proper measures are adopted to prevent misuse (e.g., pseudonymization and defacing)

(2) This Agreement is made between:

Party 1: CHARITE - UNIVERSITAETSMEDIZIN BERLIN (CHARITE), established in Chariteplatz 1, BERLIN 10117, Germany
Party 2: Z HELSINGIN YLIOPISTO (UH), established in FABIANINKATU 33, HELSINGIN YLIOPISTO 00014, Finland
Party 3: UNIVERSIDAD COMPLUTENSE DE MADRID (UCM), established in AVENIDA DE SENECA 2, MADRID 28040, Spain

The parties determine the sections in which personal data is processed under joint controllership (Article 26 GDPR).
For the other sections of processing, where the parties do not jointly determine the purposes and means of data processing, each contracting party is a controller pursuant to Article 4(7) GDPR. As far as the contracting parties are joint controllers pursuant to Article 26 GDPR, it is agreed as follows:

§ 2

(1) In context of joint controllership, Party 1 is competent for the processing of personal data shared by UCM and downloaded to a local server with restrictive access (only to members of CHARITE associated with TVB-c). Operating range includes joint decision making on methods used to pre-process and assessment of personalized brain indices for research purposes. The processing may concern the following categories of data: MEG, neuropsychology, MRI and genetics data. The legal basis for the processing of personal data is Article 6(4) + Article 5(1)(b) + Article 9(2)(j) GDPR.
(2) In the context of joint controllership, Party 2 is competent for the processing of pseudonymized data shared by UCM and downloaded to a local server with restrictive access (only to members of UH associated with TVB-c). Operating range includes joint decision making on methods used to pre-process and assessment of personalized brain indices for research purposes. The processing on this data is performed locally. The processing may concern the following categories of data: MEG, neuropsychology, MRI and genetics data. The legal basis for the processing of personal data is Article 6(4) + Article 5(1)(b) + Article 9(2)(j) GDPR.
(3) In the context of joint controllership, Party 3 is competent for the processing of personal data gathered by UCM and stored in the UCM server. The UCM server includes pre-processed and processed versions of the personal data. The processing is mostly performed locally. The processing may concern the following categories of data: Magnetoencephalography (MEG) data, genetic data, neuropsychology data, magnetic resonance imaging (MRI) data. This does not imply that all aforementioned categories of data will in fact be used for the research purposes mentioned in § 1 (1) a. The legal basis for the processing of personal data is Article 6(4) + Article 5(1)(b) + Article 9(2)(j) GDPR in connection with the national implementing law of Spain Article 6(4) + Article 5(1)(b) + Article 9(2)(j) GDPR in connection with the national implementing law of Spain (Artículo 97 Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales con Disposición adicional decimoséptima - *Tratamientos de datos de salud*).

§ 3

Each party shall ensure compliance with the legal provisions of the GDPR, particularly in regards to the lawfulness of data processing under joint controllership. The parties shall take all necessary technical and organisational measures to ensure that the rights of data subjects, in particular those pursuant to Articles 12 to 22 GDPR, are guaranteed at all times within the statutory time limits.

§ 4

(1) The Parties shall store personal data specified in § 2 in a structured, transparent, commonly used, and machine-readable format.
(2) All Parties shall ensure that only personal data, which are strictly necessary for the legitimate conduct of the process, are collected. Moreover, all contracting parties agree to observe the principle of data minimisation within the meaning of Article 5 (1)(c) GDPR.

§ 5

The Parties commit themselves to provide the data subject with any information referred to in Articles 13 and 14 of the GDPR in a concise, transparent, intelligible, and easily accessible form, using clear and plain language. The information shall be provided free of charge. The Parties agree that Party 1 provides the information on the processing of personal data in operating range A, Party 2 provides the information on the processing of

personal data as in methods/results of analyses from pre-processing to personalized brain indices and Party 3 provides the information on the processing of personal data in operating range C.
For personal data that has not been obtained from the data subject, Article 14(1)-(4) shall apply unless one of the exceptions in paragraph (5) applies. In particular, where the provision of such information proves impossible or would involve a disproportionate effort, subject to the conditions and safeguards referred to in Article 89(1) or in so far as the obligation referred to in paragraph 1 of Article 14 is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases, the Parties shall be responsible for taking the appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

§ 6

The data subject may exercise his or her rights under Articles 15 to 22 GDPR against each of the joint controllers.

§ 7

(1) Party 1, party 2 and party 3 shall provide the data subject access according to Article 15 of the GDPR.
(2) Where the data subject requests access according to Article 15 GDPR, the parties shall provide this information.
Party 3 will make available the information required in digital form, by setting up a dedicated web page and credentials for the applicant. The information can also be requested via email under ofi.dpd@ucm.es. If necessary, the parties shall provide each other with the necessary information from their respective operating range. Competent contact persons for the parties are Petra Ritter (petra.ritter@charite.de) for Party 1, J. Matias Palva (matias.palva@helsinki.fi) for Party 2, Isabel-Cecilia del Castillo (ofi.dpd@ucm.es) for Party 3. Each party must immediately inform the other of any change of the contact person.

§ 8

(1) If a data subject exercises his or her rights against one of the parties, in particular of the rights of access, correction, or deletion of his or her personal data, the parties are obliged to forward this request to the other party without undue delay. This applies irrespective of the general obligation to guarantee the right of data subjects. The party receiving the request must immediately provide the information within its operating range to the requesting party.
(2) If personal data are to be deleted, the parties shall inform each other in advance. A party may object to the deletion for a legitimate interest, for example, if there is a legal obligation to retain the data set for deletion.

§ 9

The parties shall inform each other immediately if they notice errors or infringements regarding data protection provisions during the examination of the processing activities or the order results.

§ 10

The parties undertake to communicate the essential content of the joint controllership agreement to the data subjects (Article 26(2) GDPR).

§ 11

The parties are obliged to inform the supervisory authority and the data subjects affected by a violation of the protection of personal data in accordance with Articles 33 and 34 GDPR concerning their operating ranges. The parties shall inform each other about any such notification to the supervisory authority without undue delay. The parties also agree to forward the information required for the notification to one another without undue delay.

§ 12

If a data protection impact assessment pursuant to Article 35 GDPR is required, the parties shall support each other.

§ 13

Any documents within the meaning of Article 5(2) GDPR on the purpose of processing, which serve as proof of proper data processing, shall be archived by each party beyond the end of the contract in accordance with legal provisions and obligations.

§ 14

(1) Within their operating range, the parties shall ensure that all employees authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality in accordance with Articles 28(3), 29, and 32 GDPR for the duration of their employment, as well as after termination of their employment. The parties shall also ensure that they observe the data secrecy provisions prior to taking up their duties and are familiarised with the data protection legislation and rules relevant to them.
(2) The parties shall independently ensure that they are able to comply with all existing storage obligations with regard to the data. For this purpose, they must implement appropriate technical and organisational measures in accordance with Article 32 GDPR. This applies particularly in the case of termination of the cooperation/agreement.
(3) The implementation, default-setting, and operation of the systems shall be carried out in compliance with the requirements of the GDPR and other regulations. In particular, compliance with the principles of data protection by design and data protection by default will be achieved through the implementation of appropriate technological and organisational measures corresponding to the state of the art.
(4) The parties agree to store personal data which are processed on GDPR compliant IT infrastructures of the three parties (and Charité VRE, UH local server, UCM local server) in the course of the services on specially protected servers, and agree to store the personal data for no longer than is necessary in accordance with Article 5(1)(e) GDPR.

§ 15

(1) The parties commit themselves to conclude a contract in accordance with Article 28 GDPR when engaging processors within the scope of this agreement (see § 1) and to obtain the written consent of the other party before concluding the contract. Each party shall have the right to prohibit the engagement of a particular processor if there are important reasons to be held against it
(2) The parties shall inform each other in a timely manner of any intended change with regard to the involvement or replacement of subcontracted processors. The parties shall only commission subcontractors who meet the requirements of data protection legislation and the provisions of this agreement. Services which the contracting parties use from third parties to support the execution of the contract, such as telecommunications services and

maintenance, shall not be seen as services provided by subcontractors within the meaning of this contract. However, the parties are obligated to make appropriate contractual agreements in accordance with the law and to take controlling measures to guarantee the protection and security of personal data, even in the case of additional third party services.
(3) Only processors who are subject to the legal obligation to appoint a data protection officer shall be commissioned to perform services in connection with this contract.

§ 16

The parties shall include the processing operations in the records of processing activities pursuant to Article 30 (1) GDPR, in particular, with a comment on the nature of the processing operation as one of joint or sole responsibility.

§ 17

Notwithstanding the provisions of this contract, the parties shall be liable for damages resulting from processing that fails to comply with the GDPR. In external relations they are jointly liable to the persons concerned.
In the internal relationship the parties are liable, notwithstanding the provisions of this contract, only for damages which have arisen within their operating range.

§ 18

On behalf of Party 1:

_____

Date, Signature and Institution

On behalf of Party 2:

_____

Date, Signature and Institution

On behalf of Party 3:

_____

Date, Signature and Institution