D2.1 Data Management Plan

PRISM 2 – GA 101034377
Psychiatric Ratings using Intermediate Stratified Markers 2
WP 2 Data management, sustainability, and analysis

Lead contributor	Lee Lancashire & Maryan Zirkle (Org number – Cohen Veteran's Bioscience)
	Lee.Lancashire@cohenbio.org
Other contributors	Martien Kas (01 RUG) m.j.h.kas@rug.nl; Raj Jagesar (01 RUG) r.r.jagesar@rug.nl; Gerry Dawson (P1vital) gdawson@p1vital.com; Abigail Abrahams (02 P1vital) aabrahams@p1vital.com; Philippe l'Hostis ((07 BIOTRIAL) Philippe.lhostis@biotrial.com;

Due date	30 November 2022
Delivery date	06 February 2022
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1	18.05.2021

Document History

Version	Date	Description
V1.0	24 01 2022	First Draft
V1.1	31 01 2022	Comments
V2.0	06 02 2022	Final Version

Publishable Summary

The main goal of PRISM2, within this programme of work, is to show the reproducibility and generalizability of quantitative biological parameters which were identified in the original PRISM clinical study (Protocol number ABR59359) as having significant relationships with social dysfunction, in a transdiagnostic manner. Namely,

1. SFS total score and rostromedial PFC activity
2. SFS and Behapp composite score
3. Behapp composite score and FA
4. Behapp composite score and EEG
5. SFS and connectivity between the mPFC and amygdala

The DATA MANAGEMENT PLAN describes the general data flow, compute infrastructure, and data management considerations needed to consolidate results as well as the management of the data after the PRISM2 project has ended, to allow further data analysis.

Data Management Plan

WP2 Data Collection

In PRISM2 WP 2 and 3 is responsible for data collection using their resources, where data collection and first level data pre-processing will be detailed by each provider in their specific Data Transfer Specifications (DTS) document (see Appendix), that will outline data types, file formats, naming conventions, steps taken to de-identify or anonymize the data according to EU GDPR prior to uploading to the Cohen Veteran's Bioscience (CVB) BRAINCommons infrastructure, who has access, and how data is safeguarded.

WP2 Data Management
WP2 Data Stream

Domain and modality specific data will be uploaded to a centrally accessible cloud-based platform (BRAINCommons) provided by WP2 co-lead, CVB. Data here will be made accessible to WPs for downstream data analysis for development of statistical approaches to deliver on the replication and generalization objectives of this project.

The BRAINCommons (portal.braincommons.org) architecture leverages the expertise of Amazon Web Services (AWS), a FedRAMP approved cloud services provider, and leading AWS certified and accredited managed service providers. These services include, but are not limited to S3, RDS, EC2, Cognito, CloudFront, AP Gateway, Route 53, Dynamo, Lambda, EMR, EKS and AppStream. The BRAINCommons policies, procedures, and controls aggressively protect user data and user privacy and do so utilizing the best available cybersecurity practices. These practices are, where necessary, in full compliance with evolving laws, regulatory bodies, and privacy mandates. All data is encrypted when in transit (SSL/TLS) and at rest (AES 256 or greater). We work with security, privacy, and compliance experts at AWS to define protections for constantly evolving regulations, compliance and security needs of data contributors. We rigorously log system and security level information on a scale that can immediately alert us to threats. For auditing the environment, we utilize penetration testing and intrusion detection and prevention technology. We have instituted policies and procedures that constantly monitor, assess, and remediate throughout the environment to respond immediately to threats and non-compliance.

To protect the confidentiality and integrity of the data and ensure services are made available in a secure manner, the BRAINCommons has established the NIST CSF cyber security framework and controls to achieve a security baseline for the platform. The BRAINCommons is currently implementing controls and processes to meet a Tier 3 NIST CSF compliance framework implementation tier.

Data Upload
Upload Infrastructure

CVB provides WPs with BRAINCommons a secure, cloud accessible platform. CVB will provide PRISM2 data contributors with access credentials that will allow uploading of data to the BRAINCommons data lake where it will be curated and made available to authorised users. The data contributor would be the individual site Principal Investigator(s).

Upload Account Management

Data contributors should provide contact information for one specific person (data submitter) who will act as the main point of contact for data upload purposes. Upload accounts will be specific to the data contributor and their designated submitter.

Data upload is restricted to an IP address. Sessions only remain open while the API call is active and will disconnect or expire after confirmation of successful data transfer. Default value for socket to disconnect is 60 seconds, however, we can manually set this value after discussions with the submitter.

Upload Access

Data is uploaded directly into an S3 bucket. Ideally, the data contributor will maintain the data ready for submission in S3. We will then use cross account access to a RAW S3 bucket. Otherwise, an IAM user with access to the RAW S3 bucket would be created. Then the data contributor can use the AWS command line to upload the data ready for submission.

Uploading Data

Prior to upload, contributors should complete a Data Contributor Agreement (DCA) that will be the legal basis for transfer of the data to BRAINCommons and help secondary users understand and reuse the data. This should include (where available):

- Project synopsis
- A description of the data
- The date of creation
- File formats and type
- Approximate data size
- Units of measurement
- Documentation on the methodologies used to process and transform data (using community standards where possible)
- Conditions in which data can be accessed (i.e., who can see/use it, and for how long). If data is from human participants, information around consent should be provided to ensure

data is permitted to be shared and reused, and whether any restrictions exist to permit this sharing and reuse.
- IRB approval number, if applicable
- Any restrictions on use of the data
- Statement of attribution for acknowledging use of the data

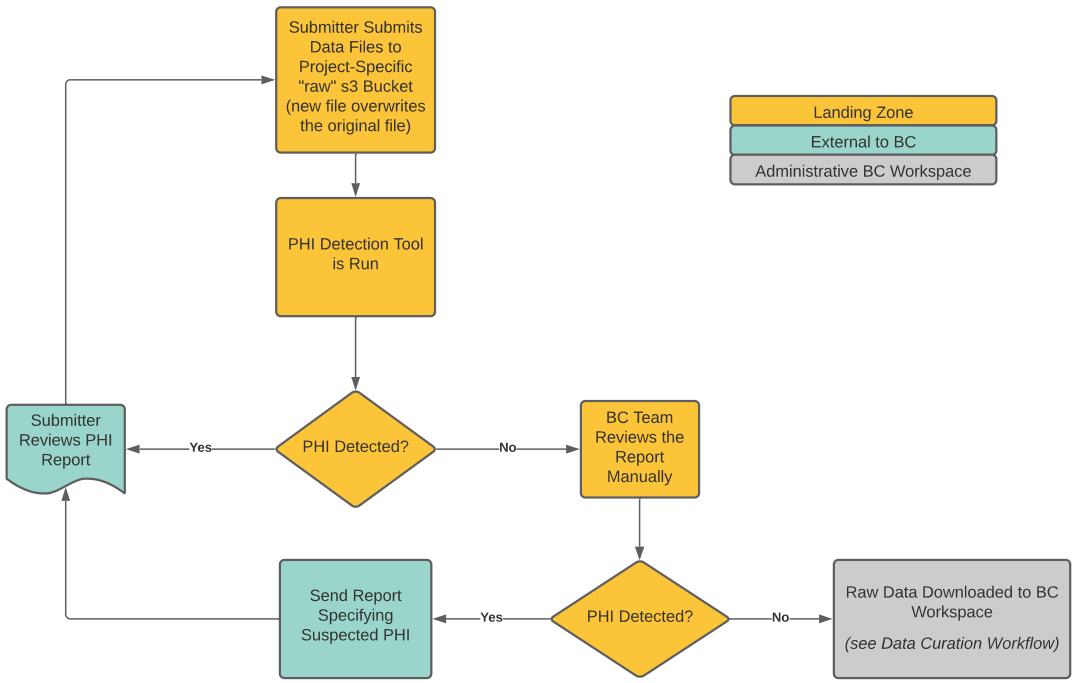
Domain and modality specifics for uploading data will be specified in respective DTS. Once the data submitter confirms that data is uploaded via the designated process, the BRAINCommons Data Steward will be given access to the corresponding RAW S3 bucket and perform several quality control checks to ensure all files designated for submission are available for curation.

Data uploaded to BRAINCommons cannot contain any participant identifying information, or personal health information (PHI) other than an anonymization number assigned by the data contributor.

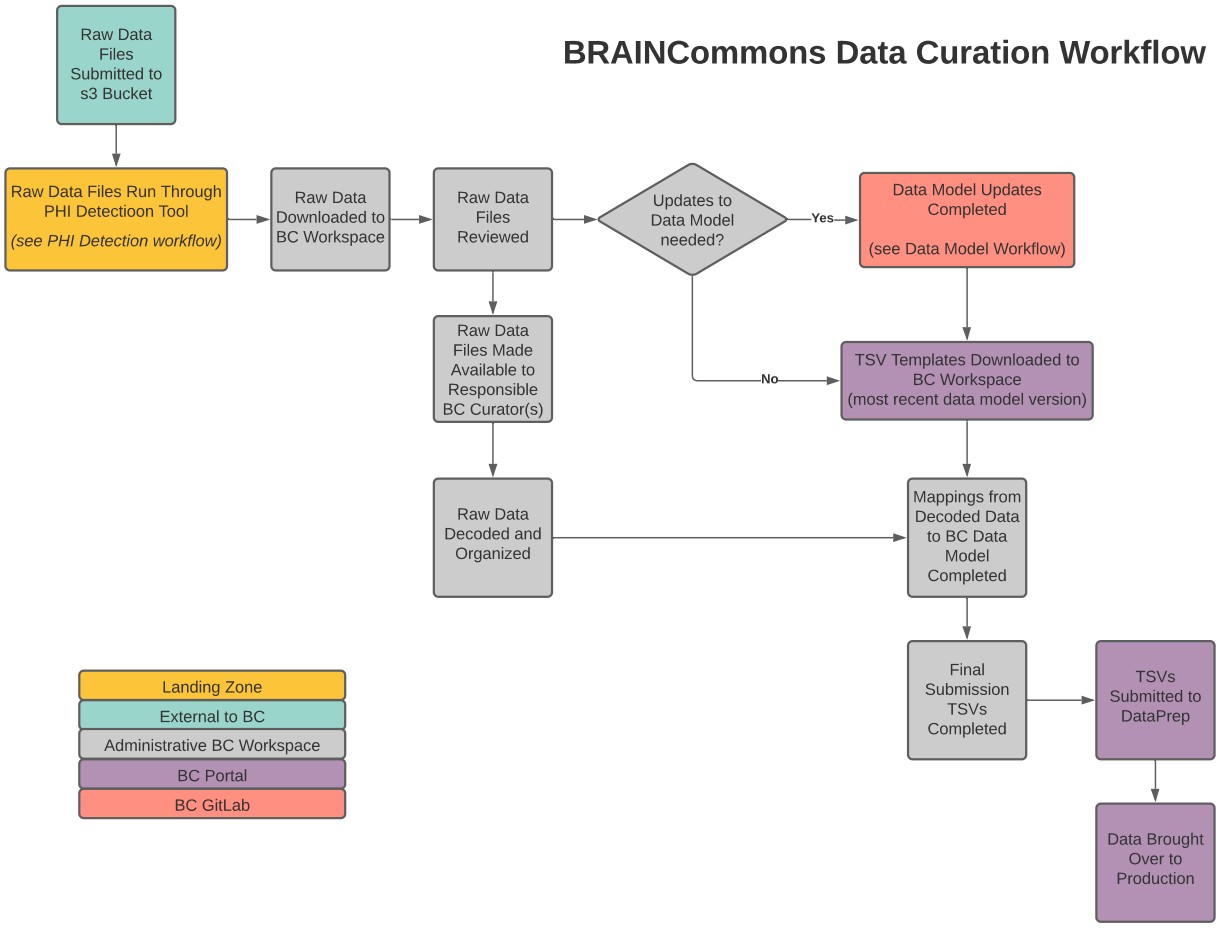
Data Consolidation

CVB will monitor the upload folders and perform validation of the de-identification of the submitted data. The PHI Detection Tool resides in the BRAINCommons (BC) secure landing zone and will alert the submitter of any personal health data detected. The submitter will be responsible for determining if the data needs additional processing to remove any PHI.

PHI Detection Workflow



Once the data is confirmed to be properly de-identified, the team will unpack the submitted data (raw data) within a BRAINCommons workspace to perform curation of the data and associated metadata so that it aligns with the BRAINCommons data model. This includes review of the submitted files by a designated curator who will evaluate the data to determine if it requires additional updates to the data model and decode and map the data to the data model.



Data Archiving

BRAINCommons leverages AWS managed services to provide 24x7x254 proactive monitoring, incident management, process automation, patching, logging, alerting and backup management of the environment. The BRAINCommons has established the NIST CSF cyber security framework and controls to achieve a security baseline for the platform. The BRAINCommons is currently implementing controls and processes to meet a Tier 3 NIST CSF compliance framework implementation tier for storage and backup.

Individual data providers should implement their own system for data backup and archiving of raw data, if only the derived data is being submitted to the BRAINCommons. They should provide a document that describes how often the data will be backed up and to which locations, including how many copies are being made. If third party services are used here, they should be described, along with contact information, to ensure that there is no conflict in terms of the legal jurisdiction in which data are held or the protection of sensitive data. If data is confidential (e.g. personal data not already in the public domain), data providers should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.

Data Availability
For WP2 Members

CVB will provide access credentials to members of WP2 to allow for data download, if applicable, and exploration and analysis in a secure BRAINCommons workspace, as needed, under a specific advanced compute contract that outlines the terms and conditions associated with use of storage, transfer and compute resources.

For IMI-PRISM2 Members

CVB will provide access credentials to members of PRISM2 to allow for data download, if applicable, and exploration and analysis in a secure BRAINCommons workspace, as needed, under a specific advanced compute contract that outlines the terms and conditions associated with use of storage, transfer and compute resources.

Acknowledgement

The PRISM 2 project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 101034377. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Cohen Veterans Bioscience. The information reflects only the authors' views neither IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein.

Appendices

WP Specific Data Management Plans (Behapp)

Data storage

1. What is going to be the primary data storage (IT infrastructure/database)?
   - A relational database combined with a data warehouse which can be queried using standard SQL -> Google Cloud SQL & BigQuery.
2. How is the data storage going to be accessed?
   - For exploratory and preliminary analysis, the data will be directly fetched from the databases to a local computer into an analytics workbench (e.g. R or SciPy) over an SSL connection. Automatic analysis will be performed on a compute instance in the private cloud environment. All data will remain encrypted in transit, at rest and while in-use.
3. Who will have access to the data storage and how do you handle user rights to protect deletion of overwriting of original files?
   - Only direct Behapp team members with a research role will have access to raw data. Researchers will only be given read-only access to the data.
4. What is the storage capacity?
   - NA
5. What is your back-up and how frequently will the back-up occur?
   - We will create full daily backups of the data.
6. How do you ensure that the original data remain untouched?
   - By consistently providing read-only access to authorized Behapp team members.

7. Who is responsible for management of the IT infrastructure and is the point of contact for any data issues (e.g. recovery of lost data)?
   - Raj Jagesar – r.r.jagesar@rug.nl
8. By whom and how often will the upload of key outcome measures be performed to the database provided by WP?
   - Key outcome measures will be generated on request and securely transferred to the specified central database.
9. What resources did you allocate to the above tasks (both infrastructure and FTEs)?
   - Est. max. 1000 – 2000 EUR per month (infrastructure) and approx. 0.2 FTE.
10. How long will data be kept? (Deleted as soon as processing is complete, retained as per legal requirement, retained forever, etc.)
   - Raw Behapp data will be kept for 25 years

Data structure and integrity

1. Where do the people whose data is being processed reside?
   - Multiple sites in the Netherlands and Spain (EU only)
2. Who is going to enter all the collected data into the database?
   - NA
3. What data structure will you use?
   - The outcome measures will be reported in a structured data form. The exact type of the outcome measure depends on its unit and on how the measure was recorded, this can vary per outcome measure. Example:

Behapp ID	Total time spent at home (minutes)	Number of unique places visited	Average time spent on incoming calls (minutes)	...
PRISM2_XXXX	200 (int)	5 (int)	360.5 (float)	...

4. What exact data is going to be stored (original, raw, processed) and in which formats (e.g. how do you handle questionnaires)?
   - Behapp will store raw smartphone based data, for PRISM 2 we will collect:
     - Location data
     - Realtime foreground application usage
     - Interval based Wi-Fi AP scans
     - Calls logs
     - SMS logs
     - Ambient light
     - Screen states
     - Physical activity labels
5. What is your QC procedure to ensure correct data entries?
   - Data quality is directly related to the uptime and correct general functionality of our app. We intend to actively monitor the incoming data performing automated checks for conformity.
6. How do you plan to integrate the data into a single dataset (query scripts, manual recoding)?
   - The outcome measures will be reported as a pseudonymised overview containing a Behapp ID. This ID can be used to merge the Behapp outcome measures into the larger consortium data sets.
7. In what format do you intend to share data with the BRAIN Commons?
   - CSV
8. How will you ensure that shared data is relevant to the purposes of the research project (data minimization)
   - NA

9. What resources did you allocate to the above tasks (FTEs)?
    - Approx. 0.2 FTE.

Personal data

1. How will you ensure that data is appropriately deidentified (pseudonymized) or anonymized prior to transfer to BRAINCommons?
    - We have a standardized data processing pipeline which is continuously evaluated by the team.
2. Please describe the processing activity (collection, use, storage and deletion) involving personal data and special category personal data
    - Raw smartphone data related to social behavior, such as location data, will be collected using an app and sent over to the central Behapp service for further analysis aligned with the PRISM II research purposes. All data is encrypted at rest and in transit
3. Where will you record and store data protection informed consent forms and information sheets?
    - NA
4. Do you have a legal basis for processing data? (Select all that apply: Necessary for the performance of a contract with the individual, necessary for the legitimate interests of the organization processing, necessary for compliance with a legal obligation, necessary for the performance of a task carried out in the public interest, necessary to protect the vital interests of individuals, consent of the individual)
    - Yes, informed consent.
5. Where will you record and store confirmation that there is lawful basis for the data processing and that appropriate safeguarding of subjects has been provided?
    - NA
6. What will be the frequency of processing? (Constantly, daily, weekly, monthly, one off/less frequently, etc.)
    - One off, upon request of the PRISM II consortium.
7. What is the relationship between your organization and the individuals whose data is being processed? Describe the context of the processing
    - The University of Groningen will be a joint controller of the data but offers Behapp as a facilitating service without direct interaction with study participants. Instead study manager at the respective sites will be trained on the inclusion flow for Behapp.
8. Are vulnerable people's (Children, disabled people, those suffering from medical conditions, Political exposed persons, other, or none of the above) data being collected/processed?
    - Minors are not involved in PRISM II, however the research is done with people from patient groups suffering from brain / mental disease.
9. Have you conducted a data protection impact assessment (DPIA)?
    - Yes
10. Who is responsible for 1-4 above?
    - Raj Jagesar – r.r.jagesar@rug.nl

WP Specific Data Management Plans (Biotrial)

Data storage

1. What is going to be the primary data storage (IT infrastructure/database)?
    - EEG raw data will be collected and stored on our internal server (Netapp).
2. How is the data storage going to be accessed?
    - NA
3. Who will have access to the data storage and how do you handle user rights to protect deletion of overwriting of original files?
    - Access rights are defined per study basis and adjusted to the role of each staff.

4. What is the storage capacity?
   - Global storage capacity is adapted to the needs.
5. What is your back-up and how frequently will the back-up occur?
   - Daily back up (incremental) and full back up every week stored in an external location.
6. How do you ensure that the original data remain untouched?
   - Original data is kept in read only access. Only duplicated copies are processed.
7. Who is responsible for management of the IT infrastructure and is the point of contact for any data issues (e.g. recovery of lost data)?
   - Dedicated IT department at Biotrial (Director = Gavin Lewis)
8. By whom and how often will the upload of key outcome measures be performed to the database provided by WP?
   - To be defined in the data transfer specifications. For such project, it can be one test transfer, one interim and one final. It is open for discussion..
9. What resources did you allocate to the above tasks (both infrastructure and FTEs)?
   - IT department for infrastructure management) and study team:
   - Project leader (1+1)
   - EEG infrastructure (1+1)
   - EEG technician (1+1)
   - EEG scientist (2)
   - Data manager (1+1)
10. How long will data be kept? (Deleted as soon as processing is complete, retained as per legal requirement, retained forever, etc.)
   - Reatained for at least 15 years

<u>Data structure and integrity</u>

1. Where do the people whose data is being processed reside?
   - France
2. Who is going to enter all the collected data into the database?
   - Database is populated following the processing of the EEG raw data. Ad hoc Matlab scripts are designed to match with study requirements.
3. What data structure will you use?
   - Matlab and SAS datasets
4. What exact data is going to be stored (original, raw, processed) and in which formats (e.g. how do you handle questionnaires)?
   - Data is kept at each step:
     - original
     - Raw
     - processed
   - EEG signal recorded using EDF format.
5. What is your QC procedure to ensure correct data entries?
   - A 100% quality control is carried out for manual entry.
6. How do you plan to integrate the data into a single dataset (query scripts, manual recoding)?
   - A mapping interface is developed to ensure the extraction of the endpoint according to a predefined file format. The file format is approved by the DM organization prior to the study start. A test transfer is performed to check the quality of the extraction
7. In what format do you intend to share data with the BRAIN Commons?
   - CSV file is preferred but SAS datasets is also possible
8. How will you ensure that shared data is relevant to the purposes of the research project (data minimization)
   - The information provided will be defined according to the study protocol.
9. What resources did you allocate to the above tasks (FTEs)?
   - One Data manager will be assigned to this task

Personal data

1. How will you ensure that data is appropriately deidentified (pseudonymized) or anonymized prior to transfer to BRAINCommons?
   - Data received at Biotrial will be deidentified and this is kept throughout the process.
2. Please describe the processing activity (collection, use, storage and deletion) involving personal data and special category personal data
   - EEG data will be collected on site using Biotrial equipment. Raw data transfer will be carried out using secured platform and will be stored at Biotrial on internal server. Data processing will be setup according to the study protocol to extract endpoints. Study data will be kept during at least 15 years and deletion will be carried out after sponsor approval.
3. Where will you record and store data protection informed consent forms and information sheets?
   - Informed consent will be stored at clinical site level, not at Biotrial.
4. Do you have a legal basis for processing data? (Select all that apply: Necessary for the performance of a contract with the individual, necessary for the legitimate interests of the organization processing, necessary for compliance with a legal obligation, necessary for the performance of a task carried out in the public interest, necessary to protect the vital interests of individuals, consent of the individual)
   - Consortium agreement
5. Where will you record and store confirmation that there is lawful basis for the data processing and that appropriate safeguarding of subjects has been provided?
   - Maintained at Biotrial QA department.
6. What will be the frequency of processing? (Constantly, daily, weekly, monthly, one off/less frequently, etc.)
   - Processing will be conducted on a regular basis. The schedule will be driven by the general planning f the project in terms of patient recruitment as well as need for results availability.
7. What is the relationship between your organization and the individuals whose data is being processed? Describe the context of the processing
   - Data processing will be carried out by Biotrial employees (Core Lab unit).
8. Are vulnerable people's (Children, disabled people, those suffering from medical conditions, Political exposed persons, other, or none of the above) data being collected/processed?
   - Data collected on patients suffering from various neuropsychiatric issues (please refer to the study protocol).
9. Have you conducted a data protection impact assessment (DPIA)?
   - Yes
10. Who is responsible for 1-4 above?
   - Biotrial QA department – Viviane Guermont, Director, Data protection Officer

WP Specific Data Management Plans (P1vital)

Data storage

1. What is going to be the primary data storage (IT infrastructure/database)?
   - MRI: MRI data will be shared by sites with P1vital via ShareFile. Data will also be kept on relevant company machines and on the P1vital server as needed.
   - ePRO: data collected via ePRO will be stored on a server hosted by Amazon Web Services (AWS).
   - OpenClinica: data collected via OpenClinica will be stored on a server hosted by AWS. Subcontractor quality and risk assessments have been completed for OpenClinica in accordance with company's 'Subcontractor Selection and Management' Standard Operating Procedure (SOP).
   - ShareFile: Data will be stored or shared via ShareFile (cloud-based file sharing and storage service which is part of Citrix Systems) which is hosted on AWS servers. Subcontractor

quality and risk assessments have been completed for ShareFile in accordance with company 'Subcontractor Selection and Management' SOP.
- P1vital Server: data / documents stored on P1vital server will be stored on a server hosted by Rackspace. Subcontractor quality and risk assessments have been completed for Rackspace in accordance with company 'Subcontractor Selection and Management' SOP.

2. How is the data storage going to be accessed?
- MRI: Data will be accessed via ShareFile (please see ShareFile paragraph in this section regarding access to data storage).
- ePRO: ePRO can be accessed by the users whose account has been set up (study participants, site user and P1vital Admin users). Access will be granted after completion and approval of System Access Request Form or equivalent as agreed with the client and in accordance with company SOP on 'ePRO Access Control'. AWS administration is accessed by P1vital account holder (Jonathan Kingslake) and P1vital's ePRO Systems Administration team (Elysium) who manage AWS on P1vital's behalf.
- OpenClinica: OpenClinica can be access by users whose account has been set up on study environment (site users and P1vital Admin users). Access will be gained after completion and approval of System Access Request Form or equivalent as agreed with the client. AWS administration is done by OpenClinica. Subcontractor assessments have been completed for OpenClinica and AWS.
- ShareFile: ShareFile is a secure, encrypted web-based file sharing service. ShareFile is access via employee account (by P1vital employees) and client user account (by study site users and other project stakeholders). Password and 2 Factor Authentications are enabled for account access. AWS administration is done by ShareFile. Subcontractor assessments have been completed for ShareFile and AWS.
- P1vital Server: access to P1vital server is restricted to P1vital employees. Access will be granted after completion and approval of System Access Request Form. Access to restricted folders must be approved. Server administration tasks are completed by P1vital IT Consultant and Rackspace administration is done by Rackspace. Subcontractor assessment has been completed for Rackspace.

3. Who will have access to the data storage and how do you handle user rights to protect deletion of overwriting of original files?
- MRI: P1vital will keep a ShareFile configuration document that keeps track of all the users and their privileges. ShareFile access will be set up in accordance with ShareFile description in this document. The ability to upload, download and delete will be limited as required.
- ePRO: Elysium, P1vital Admin and Clinician users. Revise function switched on for P1vital Admin. Full Audit Trail is available. Clinician users may review data entered, however, they are not able to edit data collected. Deletion is not possible.
- OpenClinica: P1vital Admin and Clinician users will be able to update previously entered records. Full Audit Trail is available. Deletion is not possible.
- ShareFile: Access restrictions will be in place. Only users with employee or client user accounts will be able to log in to ShareFile. These users will need to be given access to relevant folder by P1vital employee user and their rights will be restricted (e.g. ability to upload, download, delete etc.). ShareFile has document version control; therefore, data cannot be overwritten. Where document / data is deleted, these can be retrieved from recycle bin or backup.
- P1vital Server: Access restrictions are in place for P1vital Server. Access to relevant folders must be approved via completion and approval of System Access Request Form. Servers are backed up daily, therefore, previous versions of document can be retrieved.

4. What is the storage capacity?
   - MRI: unlimited
   - ePRO: unlimited
   - OpenClinica: unlimited
   - ShareFile: unlimited
   - P1vital server: unlimited

5. What is your back-up and how frequently will the back-up occur?
   - MRI: The data will be transferred to P1vital via ShareFile. Documents may be stored on P1vital server which is backed up daily.
   - ePRO: EBS volume snapshots are taken daily using EC2 snapshots functionality. EBS snapshots are retained for 14 days in S3 (replicated across different availability zones as standard). RDS snapshots are taken daily using RDS snapshots functionality. RDS snapshots are retained for 14 days in S3 (replicated across different availability zones as standard).
   - OpenClinica: System Administration will schedule full file and database backups to be completed daily in the OpenClinica Optimised Hosting environment, with multi-site retention.
   - ShareFile: If document is deleted, it stays in Recycle Bin for 7 days. After 7 days, it can be recovered via Management Console. It is not possible to recover a file after 45 days.
   - P1vital Server: The incremental backups are done daily by Rackspace and a full weekly backup is also performed.

6. How do you ensure that the original data remain untouched?
   - MRI: The data integrity will be ensured via ShareFile access and version control.
   - ePRO: full audit trail is captured.
   - OpenClinica: full audit trail is captured.
   - ShareFile: ShareFile indicates which user uploaded the document and it is time-stamped. For every file, the version history can be accessed. Therefore, any changes to the original data can be tracked and queried as needed. Access to relevant folders will be access controlled via: ShareFile access provision to relevant parties only and via folder permissions e.g. download, upload, delete etc.
   - P1vital Server: Documents can be password-protected or saved in folders with restricted access as required

7. Who is responsible for management of the IT infrastructure and is the point of contact for any data issues (e.g. recovery of lost data)?
   - MRI: Asad Malik, amalik@p1vital.com
   - ePRO, OpenClinica and ShareFile: itsupport@p1vital.com

8. By whom and how often will the upload of key outcome measures be performed to the database provided by WP?
   - MRI: Data transfers will be conducted in accordance with P1vital PRISM II Data Management and Validation Plan.
   - ePRO: Data transfers will be conducted in accordance with P1vital PRISM II Data Management and Validation Plan.
   - OpenClinica: Data transfers will be conducted in accordance with P1vital PRISM II Data Management and Validation Plan.
   - ShareFile: Data transfers will be conducted in accordance with P1vital PRISM II Data Management and Validation Plan.

9. What resources did you allocate to the above tasks (both infrastructure and FTEs)?
   - Resources will be allocated accordingly.
   - MRI (P1vital Science Team) – 2 FTEs

- ePRO, ShareFile and OpenClinica (Data Management Team) – 2 FTEs

10. How long will data be kept? (Deleted as soon as processing is complete, retained as per legal requirement, retained forever, etc.)

MRI - All data (Investigator Site Team PII, where applicable and participants' pseudonymised data) will be stored on P1vital ShareFile and / or P1vital server for the duration of the study.
At the end of the study, the data will be archived in accordance with Archiving SOP for a period of 25 years.

ePRO - All data (Investigator Site Team PII and participants' pseudonymised data) will be stored on a secure ePRO's AWS server for the duration of the study.
At the end of the study, the system will be decommissioned and archived. Data will be archived in accordance with Archiving SOP for a period of 25 years.

OpenClinica - All data (Investigator Site Team PII and participants' pseudonymised data) will be stored on a secure OpenClinica's AWS server for the duration of the study.
At the end of the study, the system will be decommissioned and archived. Data will be archived in accordance with Archiving SOP for a period of 25 years.

Data structure and integrity

1. Where do the people whose data is being processed reside?
European Union - Madrid, Spain; Leiden, Netherlands and Amsterdam, Netherlands.

2. Who is going to enter all the collected data into the database?
- MRI: sites will upload raw data. P1vital will upload pre-processed data. Sites will upload analysed data/endpoints. This will be described in the MRI Processing Manual.
- ePRO: Study participants will do data entry on ePRO.
- OpenClinica: Site staff will do EDC data entry on OpenClinica.

3. What data structure will you use?
- MRI: raw data will be in DICOM/PAR-REC files. Pre-processed data and whole-brain endpoints will be in NIfTI files. ROI endpoints will be in CSV files.
- ePRO: CSV files in CDISC compatible format as agreed in Data Transfer Specifications.
- OpenClinica: CSV files in CDISC compatible format as agreed in Data Transfer Specifications.

4. What exact data is going to be stored (original, raw, processed) and in which formats (e.g. how do you handle questionnaires)?
- MRI: original/raw data and processed data will be stored. (See 2 above for formats).
- ePRO: .csv .xlsx formats, raw / original
- OpenClinica: .csv format, raw / original

5. What is your QC procedure to ensure correct data entries?
- MRI: QC on the original data will be described in the MRI Procedures Manual and QC on the processed data will be described in the MRI Processing Manual.
- ePRO: ePRO QC will be described and completed in accordance with P1vital Data Management and Validation Plan.
- OpenClinica: OpenClinica QC will be described and completed in accordance with P1vital Data Management and Validation Plan.
- ShareFile: ShareFile configuration document will be produced.

6. How do you plan to integrate the data into a single dataset (query scripts, manual recoding)?
   - A MRI: This will not be done – further clarification with coordination team may be required.
   - ePRO: Currently we do not plan to have single dataset but Pentaho data integration tool will be used to create datasets as per requirements defined in the Data Transfer Specification.
   - OpenClinica: Currently we do not plan to have single dataset but Pentaho data integration tool will be used to create datasets as per requirements defined in the Data Transfer Specification.

7. In what format do you intend to share data with the BRAIN Commons?
   - MRI: We can provide the DICOM/PAR-REC and NIfTI files – depending on what is required.
   - ePRO: .csv or .xlsx
   - OpenClinica: .csv or .xlsx

8. How will you ensure that shared data is relevant to the purposes of the research project (data minimization)
   - MRI / ePRO / OpenClinica: Only users that require access to the data will be provided access. Data minimisation is discussed during study set up with the sponsor to ensure only data that is absolutely required is collected. Decision on what data is collected is made by the sponsor and P1vital builds its systems in accordance with client specifications.

9. What resources did you allocate to the above tasks (FTEs)?
   - MRI (Science Team) – 2 FTEs
   - ePRO and OpenClinica (Data Management Team) – 1FTE

<u>Personal data</u>

1. How will you ensure that data is appropriately deidentified (pseudonymized) or anonymized prior to transfer to BRAINCommons?
   - MRI: The clinical sites will only transfer personally deidentified data to ShareFile. P1vital will check that the data is correctly anonymised. If it is not, the data will be deleted, and the sites will be asked to anonymise the data and then upload it again. P1vital will deface the T1w scans and analysts will only have access to the defaced T1w scans. P1vital can provide BRAINCommons only the defaced scans if that is what is required.
   - ePRO: no personal data for study participants will be collected, only pseudonymised data will be collected on ePRO (study participants will be identified by their Participant ID).
   - OpenClinica: no personal data for study participants will be collected, only pseudonymised data will be collected (study participants will be identified by their Participant ID).

2. Please describe the processing activity (collection, use, storage and deletion) involving personal data and special category personal data

MRI imaging data:
Source of data:
Study team members (Users) – they will access MRI console which will collect and store:
Investigator Site Users' data when logging in
Participant's data after Investigator Site User sets up a Participant – data will be pseudonymised i.e., only participant code will be used to identify participant.

Data collection, usage, storage & deletion:
The following data will be collected: At the MRI console, study team members (Users) will be asked to enter the participant's name and date of birth. This information is contained in the DICOM files that will be shared by the Users with P1vital. The Users have been instructed to

enter pseudonymised data by setting the participant's name as the subject code and the date of birth as a generic date of birth (e.g., 01Jan1900).

MRI data will be shared by Investigator Sites with P1vital via P1vital ShareFile. Data will also be kept by P1vital on relevant Company machines and on the P1vital server as needed:

P1vital ShareFile - ShareFile is a secure, encrypted web-based file sharing service. ShareFile is access via employee account (by P1vital employees) and client user account (by study site users and other project stakeholders). Password and 2 Factor Authentications are enabled for account access. AWS administration is done by ShareFile. Subcontractor assessments have been completed for ShareFile and AWS.

P1vital server: access to P1vital server is restricted to P1vital employees. Access will be granted after completion and approval of System Access Request Form. Access to restricted folders must be approved. Server administration tasks are completed by P1vital IT Consultant and Rackspace administration is done by Rackspace. Subcontractor assessment has been completed for Rackspace.

System access:

P1vital ShareFile: P1vital will keep a ShareFile configuration document that keeps track of all the users and their privileges. ShareFile access will be set up in accordance with ShareFile description in this document. The ability to upload, download and delete will be limited as required.

P1vital Server: Access restrictions are in place for P1vital Server. Access to relevant folders must be approved via completion and approval of System Access Request Form. Servers are backed up daily, therefore, previous versions of document can be retrieved.

Sharing of data:

Investigator Study Sites will share pseudonymised MRI data with P1vital and P1vital will share pseudonymised MRI data with data controller and analysts. Data will be shared via P1vital ShareFile. P1vital will keep a ShareFile configuration document that keeps track of all the users and their privileges. ShareFile access will be set up in accordance with ShareFile description in this document. The ability to upload, download and delete will be limited as required.

P1vital® ePRO questionnaire data:

Source of data:

Investigator Site Team - they will access ePRO, a web-based application which will collect and store:

Investigator Site Users' data when setting up their account.

Participants' pseudonymised data after Investigator Site User logs into their account and sets up a Participant on the system – data will be pseudonymised i.e. only participant code will be used to identify participant.

Data collection, usage, storage & deletion:

The following data are being collected: (1) study team members: Contact (Name, Postcode, Email address, Telephone number) will be collected, stored and used for system access, sending email notifications and for user identification purposes only. (2) study participants: health data and participant code.

Usage data for Investigator Site Team will be collected and stored on ePRO and used for keeping a record of user system activity, for understanding usage of the system and

providing user support, if required. P1vital Admin Users will have access to Investigator Site Team's PII and Investigator Site Team's PII will be retained for audit trail purposes.

Participant pseudonymised data from ePRO will be accessible to Investigator Site Users, P1vital Admin Users and super users at Elysium (P1vital's subcontractor) who hold an encryption key in case of emergencies.

When data will be extracted from ePRO, data management team will use validated Pentaho scripts to process the data.

Special categories of data:
The personal data concern the following special categories of data: User's data includes Article 9 Special category of personal data – health data collected via ePRO.

How much data will you be collecting and using? How many individuals are affected?
Individuals affected: ~180 participants (40 participants with Alzheimer's, 40 participants with Schizophrenia, 40 participants with Major depressive disorder and 60 healthy controls and approximately 5 clinical study site users per site.

What geographical are(s) does it cover?
Madrid, Spain; Leiden, Netherlands and Amsterdam, Netherlands.

Sharing of data:
Data will be shared with data controller. Encrypted data will be stored on ePRO which will be hosted on AWS server in Europe. Data will be shared using P1vital ShareFile which is also hosted on AWS server in Europe.

OpenClinica
Source of data:
Investigator Site Team - they will access OpenClinica, a web-based application which will collect and store:
Investigator Site Users' PII will be provided when they set up their account
Participants' pseudonymised data will be recorded by Investigator Site Team. Investigator Site Team will record participant's data on paper CRF (out of scope for this DPIA) and then input pseudonymised participant data in an OpenClinica system.

Investigator Site Team members will access OpenClinica which will collect and store study participant pseudonymised data. Participants will not have access to OpenClinica.

Data collection, usage, storage & deletion:
The following data will be collected: (1) Investigator Site Team users: contacts details for account set up and (2) participants: participant pseudonymised demographic and health data will be recorded by Investigator Site Team users.

Investigator Site usage data will be collected and stored on OpenClinica and used for keeping a record of user system activity, for understanding usage of the system and providing user support, if required. Usage data will be accessible to P1vital staff.

Investigator Site Team Users' PII (e.g.; Name, Email address, password) and pseudonymised participant eCRF data will be stored in an encrypted manner within OpenClinica. Investigator

Site PII and pseudonymised participant eCRF data from OpenClinica will be accessible only to Investigator Site Team users and super users at P1vital.

Special categories of data:

The personal data concern the following special categories of data: User's data includes Article 9 Special category of personal data – health data collected via OpenClinica.

How much data will you be collecting and using? How many individuals are affected?

Individuals affected: approximately 5 clinical study site user per site who will have account and ~180 participants for whom eCRF records will be created.

What geographical are(s) does it cover?

Madrid, Spain; Leiden, Netherlands and Amsterdam, Netherlands.

3. Where will you record and store data protection informed consent forms and information sheets?

- n/a, this will be collected by and stored at Investigator Sites.

4. Do you have a legal basis for processing data? (Select all that apply: Necessary for the performance of a contract with the individual, necessary for the legitimate interests of the organization processing, necessary for compliance with a legal obligation, necessary for the performance of a task carried out in the public interest, necessary to protect the vital interests of individuals, consent of the individual)

Lawful basis for processing participant pseudonymised data based on Article 6(1)(a) – consent and Article 9(2)(a) – consent.

Participant data including health data (via ePRO, OpenClinica and MRI) will be collected from participants that give consent to take part in the clinical trial using a PIS and ICF which is approved by relevant Ethics Committee.

Lawful basis for processing Investigator Site Teams' PII is based on Article 6(1)(b) – contract.

PII will be collected from Investigator Site Users that have consented to the use of ePRO / OpenClinica via Privacy Notice. Investigator Site Teams

5. Where will you record and store confirmation that there is lawful basis for the data processing and that appropriate safeguarding of subjects has been provided?

- Data Protection Impact Assessment, Privacy Notices for ePRO, Privacy Notice for OpenClinica
- Study Participant's consent will be collected via Patient Information Sheet / Informed Consent Form by Investigator Sites and stored at Investigator Sites.

6. What will be the frequency of processing? (Constantly, daily, weekly, monthly, one off/less frequently, etc.)

Frequency of data collection: data will be collected in accordance with Time and Events table as specified in Protocol.

7. What is the relationship between your organization and the individuals whose data is being processed? Describe the context of the processing

PII will be collected from Investigator Site Users that have consented to the use of ePRO / OpenClinica via Privacy Notice acknowledgement. Investigator Site Teams have contract with a sponsor under which they are required to use ePRO and OpenClinica during study participation.

P1vital has no relationship with users.
Participant pseudonymised data including health data (via ePRO, OpenClinica and MRI) will be collected from participants that give consent to take part in the clinical trial using a PIS and ICF which is approved by relevant Ethics Committee. P1vital has no relationship with participants.

8. Are vulnerable people's (Children, disabled people, those suffering from medical conditions, Political exposed persons, other, or none of the above) data being collected/processed?

Patients with mild dementia, schizophrenia and major depressive disorder will be recruited into this study. However, as patients with mild dementia and schizophrenia participate in the use of ePRO / OpenClinica in PRISM 1 with no issues, there are no concerns with the type of data processing for these patient groups.

9. Have you conducted a data protection impact assessment (DPIA)?

- Yes

10. Who is responsible for 1-4 above?

- Data Protection Officer, dpo@p1vital.com