PROMETEUS

preterm brain-oxygenation and metabolic eu-sensing

D8.4 – Data Management Plan

Partner:	Università degli Studi di Padova
Lead Author:	Università degli Studi di Padova
Version: F: final; D: draft; RD: revised draft	F
Date:	31/07/2023

European Innovation Council

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101099093

Grant Agreement	101099093
Acronym	Prometeus
Project full title	Preterm Brain-Oxygenation and Metabolic EU-Sensing: Feed the Brain

Deliverable	**D8.4**
Deliverable Name	**Data Management Plan**
Nature of deliverable	DMP: Data management plan
Dissemination level	PU (Public – fully open automatically posted online on the Project Results platforms)
Scheduled delivery date	31/07/2023
Actual delivery date	31/07/2023

Prepared by	Sabrina Brigadoi
Reviewed by	Marta Pozza and all WP leaders
Verified by	Sabrina Brigadoi

History of Changes

Revision	Date (dd/mm/yyyy)	Author	Changes	Status (Draft/Inreview/ Submitted)
v1_Dave	24/07/2023	Alberto Scarpa	Added information regarding WP5-6	Draft
v1_UGA	24/07/2023	Emmanuel Barbier	Added information regarding WP3	Draft
v1_POLIMI	24/07/2023	Davide Contini	Added information regarding WP1	Draft
v1_QLAB	25/07/2023	Idan Tamir	Added information regarding WP2	Draft
v1_UNIPD	25/07/2023	Chiara Dalla Man	Added information regarding WP3-4	Draft
v1_UNIPD	25/07/2023	Paola Rigo	Added information regarding WP7	Draft
v2	26/07/2023	Sabrina Brigadoi	Incorporated changes from all partners	Final

Table of contents

1. Project overview

Prometeus introduces a new-paradigm for personalized nutrition of prematurely born neonates in neonatal intensive care unit (NICU), by developing a ground-breaking technology for real-time adjustment of glucose and nutrients intakes to target neonatal brain needs. To do so, *Prometeus* will develop a metabolic model of the interaction between the three key brain fuels [glucose, lactate and beta-hydroxybutyrate (BHB)] and their effect on regional cerebral blood flow (CBF), oxygen saturation (StO_2) and metabolism ($CMRO_2$). The model will serve to individualize brain nutrition targeting "brain health" according to the inputs derived from two novel minimally invasive metabolic sensing systems: a wearable cap (neo-opticap) measuring regional CBF, StO_2 and $CMRO_2$ and a subcutaneous miniaturized metabolic sensor (CMM) for glucose, lactate and BHB. The system will create a metabolic "womb" to feed the brain of preterm neonates. *Prometeus* will be paralleled by a parent-dedicated interface, exploiting a purposely developed family adjusted visual language, that will inform parents of preterm babies during their NICU admission and stay. *Prometeus* will generate several types of data, including clinical data, simulated data, *in vivo* animal data and videos, which need to be managed following high security standards and privacy regulations.

2. Data summary

2.1 Type of studies and relation to the objectives of the project

In WP1, *phantom* and *in vivo* adult studies will be performed to test the precision and stability of the neo-opticap in relevant settings. *Phantom* experiments will be performed using standardized optical phantoms whereas *in vivo* measurements will be performed on healthy adults' head.

In WP2, an *in vivo* animal study in pigs will be performed to test the CMM sensor biocompatibility and the accuracy of glucose, lactate and BHB sensor measurements. Animal data will be collected at the animal facility of QLAB by seriated bio-compatibility experiments in 5 pigs to assess the local effect of the sensor materials. The same animals will be tested to measure accuracy of glucose, lactate and BHB sensor measurements during resting state after 6 hours starvation period (phase 1) and after glucose intravenous load (phase 2) by seriated blood sampling and in the presence of pharmacological disruptor commonly used in NICU (ampicillin, netilmicin, dopamine, fentanyl, ibuprofen, paracetamol, caffeine).

In WP3, an *in vivo* animal study in Sprague Dawley rat litters will be performed to acquire data to create the metabolic model of the interaction between the three key brain fuels (glucose, lactate and BHB) and their effect on regional cerebral blood flow (CBF), oxygen saturation (StO_2) and metabolism ($CMRO_2$). Groups of litters will follow different nutritional ratios of fat and protein and will be subject to either hypoglycemic hyperinsulinemic clamp or hyperglycemic clamp or euglycemic clamp. Tracers will be used to estimate glucose, lactate and BHB production and consumption. Measurements of cerebral hemodynamics will be performed with Magnetic Resonance Imaging (MRI) technique. Blood samples with be analysed with Liquid chromatography-mass spectrometry techniques and tissues histologically analysed.

In WP4, an *in silico* test of the neo-controller (the nutritional clinical advisor) will be performed under different simulated scenarios to estimate its safety and efficacy. *in silico* testing is required to test the safety and expected outcomes of the neo-controller under common scenarios (e.g., hypo/hyper-glycemia). Data will be generated applying the neo-controller in different simulated scenarios and evaluating its outcome compared to standard clinical practices.

In WP6, 8 preterm neonates will be tested using the *Prometeus* system as a feasibility study. They will wear the *Prometeus* system and healthcare personnel will receive feedback and instructions

from the *Prometeus* system.

In WP7, interview, standardized questionnaire, and narrative sessions to evaluate the premature birth experience will be administered to parents of preterm babies and health care personnel working in NICU with the final aim of creating the first oral & visual archive of prematurity.

2.2 Types of data

WP1: quantitative data; generated from *phantom* and *in-vivo* test measurements (e.g., time series, optical measurements); projects and drawings of the instrument and modules design. Expected size: Tb

WP2: quantitative and qualitative data; images; expected size: 20-100 Gb

WP3: quantitative data; images, tissue samples, blood samples; expected size: Gb

WP4: quantitative data; generated with simulation; expected size: Gb

WP6: quantitative data; generated from clinical measurements and *Prometeus* system (e.g., time series, nutritional information, blood samples, plasma samples), personal data (demographic data of neonates and their health status [medical records relevant to the project]); expected size: Gb

WP7: qualitative and quantitative data generated from interviews and questionnaires, video/audio recordings, personal data (demographic data of both parents and HCPs [Age, sex, nationality, marital status, level of education, occupation, family income], e-mail and phone number, pictures, audio and video recording); expected size: Gb

2.3 Format and software

WP1: drawings and instrument projects will be performed using as much as possible industrial standards and will, mainly, be confidential. Publicly accessible versions for dissemination purposes will be generated if possible. Research application data on the testing of neo-opticap will be obtained by means of non-standard SW developed by each laboratory. These data will be stored in binary and text files. They will include links to analysis scripts (Matlab, R, Excel, custom-software). The processed data will be saved in a report format and will be publicly available once cleared in terms of IP, exploitation, and publication issues.

WP2: generated data will be in Word, Excel, Jpg and similar formats for the *in-vitro* and *in-vivo* studies.

WP3: images will be general in nifti formats.

WP4: generated *in silico* data will be in .csv format.

WP6: the data coming from the sensors will be received by the edge application via HTTPS REST or Bluetooth. These data will be stored inside the edge device on a local database that will be synchronised with a cloud database.

WP7: the qualitative and quantitative data generated from interviews and questionnaires will be in .csv and .xlsx format. The software used to administer interviews and questionnaires will be Qualtrics.com. The video and audio recordings generated from narrative sessions will be in .mp4

and .mp3 formats respectively. The program used to administer narrative sessions will be Zoom.com. The transcriptions of the audio recordings will be in .doc format. The program used to transcript the audio of the narrative sessions in text will be Cloud Speech-to-text from Google.

3. FAIR data

3.1 Findable

WP1: non confidential data which will be part of academic peer reviewed publications will be made available via a recognised suitable data sharing repository (e.g., Zenodo or other available repository). Metadata and standards of the repository used for sharing the data will be implemented.

WP2: the dataset produced in the project will be identifiable and locatable through DOI. Record names will follow a logical and predictable way using version numbering and searchable keywords will be employed to easily retrieve them.

WP3: Preclinical data will be stored using the following rules (directories are bolded):

/repository-root
- /sub_info.xlsx [contains information on the subjects]
- **/MRI-XXX**
  - /MRI-XXX_slice-XXXX.nii [brain slice scan of subject XXX; XXXX = physical slice position in µm, anterior to posterior]
- **/derived-data**
  - /analysis-X_info.json [structured info on applied analysis]
  - **/analysis-X**
    - /analysis-X_set-X.tsv [resulting data of analysis X that is defined in 'analysis_info.json]
- **/code**
  - /analysis-X.py [script that produces result data of analysis X]
- **/Blood-XXX**
  - /BLO-XXX-ZZZ_numberYYY [subject XXX timestamp ZZZ number YYY]
- **/Tissue-XXX**
  - /TIS-XXX-ZZZ_numberYYY [subject XXX timestamp ZZZ number YYY]

WP4: the simulated data will be stored in Research Data Unipd repository (the UNIPD repository for research data, a content provider of OpenAIRE and indexed in re3data.org and OpenDOAR). A DOI will be assigned to the dataset, which will be described with a linked metadata file.

WP6: data collected during the clinical trial will be recorded with the following metadata: patient ID, health care provider ID, hospital ID, sensor type and starting and ending date-time. These metadata can be used to search and extract data.

WP7: the transcribed and anonymized interviews and questionnaires data will be stored in Research Data Unipd repository (the UNIPD repository for research data, a content provider of OpenAIRE and indexed in re3data.org and OpenDOAR). A DOI will be assigned to the dataset, which will be described with a linked metadata file.
For the audio-video recorded narrative sessions: (i) the raw data will be stored in an external encrypted hard-disk system and (ii) the encoded, anonymized, and aggregated data will be stored in Research Data Unipd repository (the UNIPD repository for research data, a content provider of OpenAIRE and indexed in re3data.org and OpenDOAR). A DOI will be assigned to the dataset, which will be described with a linked metadata file.

The visual & oral archive of prematurity will be openly licensed and permanently archived in Phaidra (https://phaidra.cab.unipd.it), the UNIPD repository for long-term preservation and dissemination of digital collections. Phaidra is CoreTrustSeal certified as a Trustworthy Digital Repository which guarantees the reliability of data in compliance with the standards for digital preservation, data interoperability and FAIR principles. The archive will be assigned a permanent link and persistent identifier.

3.2 Accessible

WP1: all the data considered confidential in terms of IP protection, or the basis of a trade secret, will be secured and not published.

WP2: recorded data that will require IP protection will be kept closed until such protection can be secured. All other data will be deposited using DOI format and accessible through common search tools.

WP3: preclinical data and metadata will be made available with a DOI on zenodo.org. An embargo period will be maintained until the first publication of the data. Data will be readable using standard software in the field.

WP4: the simulated data will be made openly available in Research Data Unipd repository.

WP6: *Prometeus* Exploitation Committee (PEC) will identify proper licenses for data sharing and re-use (e.g., Creative Commons, Open Data Commons). Data will be stored for the entire duration of the project and up to maximum 3 years afterwards. Partners of the project will have access to the data.

WP7: the transcribed and anonymized interviews and questionnaires data and the encoded, anonymized and aggregated data of narrative sessions will be made openly available in Research Data Unipd repository. The visual & oral archive will be openly licensed in Phaidra.

3.3 Interoperable

WP1: Forward compatible formats such as text files (tabulated values, open-source formats such as R data-tables), and/or open-source binary formats (such as open document spreadsheets, open document text) and/or custom-made binary formats (with definition files stored in standard text formats) will be applied with associated descriptive documentation.

WP2: the produced data will be interoperable through common SW tools.

WP3: preclinical data produced by *Prometeus* will be interoperable, according to the standards of each field.

WP4: .csv file is a standard format that can be easily exported in Matlab, Python, R etc., used by most the bioengineers.

WP6: data produced by *Prometeus* system will be interoperable, according to the standards of each field (e.g., neo-opticap will be produced according to the standards of cerebral oximeters)

WP7: the data produced are interoperable. Phaidra, with comprehensive mapping of object

metadata to the Dublin Core profile, offers excellent interoperability with other platforms, such as Europeana.

3.4 Reusable

UNIPD in collaboration with library services of academic partners and the *Prometeus* Exploitation Committee (PEC) will identify proper licenses for data sharing and re-use (e.g., Creative Commons, Open Data Commons), evaluating each time the new dataset. Data will be stored for the entire duration of the project and up to maximum 5 years afterwards (except for the oral & visual archive of prematurity, which will be stored permanently in Phaidra).

WP1: all data will be stored in the cloud and in addition on a secure hard-drive that is backed up following the best practices of the consortium partners. All computers used by the personnel involved in the project are managed following the best practices internally defined by the IT departments of each partners involved.

WP2: data gathered throughout the project will be reusable under license after project finalization.

WP3: the data generate in WP3 will reusable according to the consortium guidelines.

WP4: simulated data will be stored in the repository for the entire duration of the project and up to maximum 5 years afterwards.

WP6: the data generate in WP6 will be reusable according to the consortium guidelines.

WP7: the anonymized data from interviews and questionnaires and the anonymized and aggregated data from narrative sessions generated in WP7 will be reusable according to the consortium guidelines.

4. Allocation of resources

Each partner will identify a Data Manager who is responsible of data storage and preservation, under the supervision of the coordinator.

WP1: no additional resources and/or specialist expertise is needed because each partner involved has well established internal procedures for data management. The costs are budgeted within the project and internally by each partners involved.

WP2: during the project's duration data storage will be performed in the cloud (Google or AWS). Following project finalization, the relevant data will be forwarded to the Coordinator for further storage and sharing.

WP3: Data will be made available on zenodo.org (no cost).

WP4: Research Data Unipd does not require any expense for data storage and preservation. WP3 Leader (Chiara Dalla Man) will be responsible for data management of these data.

WP6: clinical trial data will be stored on cloud databases services. Their cost is part of the external costs of Dave. The WP5 Leader (Alberto Scarpa) will be responsible for data management of this data.

WP7: Research Data Unipd and Phaidra do not require any expense for data storage and preservation. The Coordinator (Sabrina Brigadoi) will be responsible for data management of these data.

5. Data security

WP1: all sensible data will be stored in a secured way following the best practices of the consortium partners. Transmission of data will be performed using the private area of the project website or using the certified platforms of each partner involved.

WP2: no personal data is expected to be collected. Access to other non-disclosable information will be secured through the cloud using personal usernames and passwords.

WP3: preclinical data will be stored in centralized servers of University Grenoble Alpes. This system is continuously duplicated in a different location. Data access is controlled by login and password. Access to the data is granted by the PI of UGA Partner. Preclinical data will be acquired and validated by UGA.

WP4: no problem of security is envisioned for the simulated data. They are not collected from patients nor relatives.

WP6: access to the system is protected with login and password. Data generated by doctors are validated for type and range of validity. Data generated by sensors are validated for type and range of validity. Data at transit from sensors to edge device are protected with encryption like HTTPS/TLS. Data on edge device are stored in local databases protected with password and encrypted. Data at transit between edge device and cloud are protected with encryption like HTTPS/TLS. Data at rest on the cloud are stored in databases protected with password and encrypted. All components of the cloud platform are protected with login and password and there is an auditing service to verify accesses. Data on cloud have a daily backup during the clinical tests and a weekly backup later.

WP7: Ensure integrity – The PI of each site (UNIPD and UCC) and the authorized research staff working for the *Prometeus* project will be responsible for collection/processing of personal data. Only PI and authorized research staff will have access to the personal data. Furthermore, UNIPD and UCC will impose input validation to preclude the entering of invalid data and error detection/data validation to identify errors in data transmission. Unauthorized access: encrypted NAS or server with institutional restricted access will be used for storage of electronic data. No access will be given except to the investigating staff and authorized study monitors from the sponsoring institution. Ensure recoverability: all data will be backed up keeping the same security standards for back-ups than for original files. Successful backed-up of data will be checked every two weeks.

6. Ethical aspects

Personal data collection will be compliant with General Data Protection Regulation (GDPR) EU 2016/679.

UNIPD and UCC will implement organizational measure to safeguard the privacy of the subjects involved in the study: the DPOs have already been appointed (privacy@unipd.it, gdpr@ucc.ie). Both UNIPD and UCC have developed procedure for reporting personal data breaches. Participants will be allowed to accept or refuse the publication of the interview on the *Prometeus* archive or to publish only the anonymized script of the interview. Participants will be allowed to withdraw the interview any time during or after the study. Participants will receive a detailed privacy information sheet and will be given the possibility to decide to publish the audio and video recording of the interview, the recording of the audio only or anonymized script of the interview. A privacy information sheet will be provided to the study participants in order to explain them: a) lawful basis and scientific reason for sensitive data storage and processing; b) Data Controller detail; c) Data Protection Officer (DPO) contact detail; d) measure to ensure data security and countermeasure in case of data breach; e) rights of the study participants.
Ensure confidentiality – Coding all subject data with a unique identification number will minimize risk of loss of subject confidentiality. Each Center's Log, linking Study ID Number to patient identity, will remain with the site directors in a locked file. None of the CRFs will contain any personal identifying information, so that all information received by the Coordinator Center (UNIPD) and the Data management facility at Dave will have no identifiable patient data. At each site, all information that includes patient identifiers (e.g., copies of CRFs, etc.) will be placed in locked file cabinets. Any publication arising from this study will maintain the anonymity of study participants. Subject-specific hospital data and completed questionnaire data will be made available only to the research staff. The only dataset with subject identifier information will be the subject tracking system used to follow-up and contact families. All other datasets will label subject records with a unique study number and be stripped of other identifying information; specifically, clinical data will not reside with identifying data. All results will be described in aggregate without identification of individual subjects.
All results deriving from the pilot and pivotal study will be shared among partners of consortium and used only in aggregated form and fully anonymized. No exchange of personal data will take place among partners.