

IMI2 101034344 – EPND
European Platform for Neurodegenerative Diseases
WP8 –Management and Overall Coordination

D8.3 – Data Management Plan

Lead contributor (partner organization)	6 - UNILU
Other contributors	

WP8	
Due date	30/04/2022
Delivery date	04/05/2022
Delivery type¹	R
Dissemination level²	PU

Description of Action	
V1.0	15/10/2021

Document History

Version	Date	Status
V1.1.	01/04/2022	Draft
V1.2	05/04/2022	Comments
V1.3	28/04/2022	Comments
V1.4	29/04/2022	Final

Introduction:	3
Glossary:	4
1. Project introduction and objectives	6
2. Data Summary	6
3. FAIR data	10
3.1. Making data findable, including provisions for metadata	10
3.2. Making data openly accessible	12
3.3. Making data interoperable	12
3.4. Increase data re-use (through clarifying licences)	13
4. Allocation of resources	13
5. Data security	14
5.1. Formal information/data security standards	14
5.2. Data Protection Impact Assessment (DPIA)	14
5.3. Data Sharing and Access	15
5.3.1. Suitability for sharing	15
5.3.2. Discovery by potential users of the research data	15
5.3.3. Governance of access	15
5.3.4. Restrictions or delays to sharing, with planned actions to limit such restrictions	16
5.4. Main risks to data security	16
6. Ethical aspects	16
7. Other issues	17
8. Updates of this Document and feedback	17
9. Conclusion	17
10. Deviation	17

Introduction:

This document is a data management plan (DMP) whose purpose is to provide an analysis of the main elements of the data management policy that will be used by the consortium regarding the project research data. It will be consistent with the exploitation and IPR requirements detailed in the EPND Consortium Agreement (CA). The DMP covers the complete research data life cycle. It describes the types of research data that will be provided and/or generated during the project, the standards that will be used, how the research data will be exploited/shared/made accessible, and how they will be curated, FAIRified and preserved. In collaboration with work package (WP) 7, the DMP will also describe the sustainability of the data generated and the platform, tools/services/algorithms/code-base developed during the course of the EPND project.

The DMP will be a living document, (the current document is the first DMP version, planned as Deliverable (D) 8.3), and will be updated throughout the project whenever significant changes arise, such as new types of data or changes in the consortium policies or consortium composition. At a minimum, two updates will occur at mid-term and at project-end (D8.9, D8.14). The expected types of research data that will be collected and/or generated as part of EPND include the following data categories of high dimensional data: clinical and molecular data such as genomics, transcriptomics, proteomics, metabolomics (see Table A.1, Annex A). The data type set will be updated in the next DMP versions, as the definitions of common datasets and data-processing SOPs continue to evolve. As EPND is a highly data-centric project, there is a specific WP dedicated to data FAIRification (WP1), which will execute most of the data management strategy described in the DMP to increase discovery of research data and samples in line with the overall objectives of EPND to meet the needs of the larger global community of neurodegenerative disorder (NDD) researchers. The detailed aspects of making data FAIR, including the selection of data standards and relevant data extractions and transformations, will be covered by Task (T)1.2 and T3.3 and reported in D1.2 and D3.1, respectively.

This DMP presents an overview of the guidelines for data security and personal data protection, in line with the Regulation of the European Parliament and the Council on the protection of individuals regarding processing personal data, and on free movement of such data (GDPR)¹ and relevant Member State laws on sample and data sharing for research and associated research ethics approvals. The strategy for data protection (including the case law and the constantly evolving interpretative guidance issued by the European Data Protection Board - EDPB) will be specified in T2.1, in Whitepaper 1 (D2.2) and the Data Protection Impact Assessment (DPIA) (D2.1), continuously monitored by WP2 experts, and drawing on existing consortia models and guidelines.

Glossary:

AD	Alzheimer's Disease
ADWB	Alzheimer's Disease Workbench
ALS	Amyotrophic Lateral Sclerosis
API	Application Programming Interface
ATN	Amyloid-Tau-Neurogeneration
CA	Consortium Agreement
CAA	Cerebral Amyloid Angiopathy
CBD	Cortical-basal degeneration
CDISC	The Clinical Data Interchange Standards Consortium
CRF	Case Report Form
CSF	Cerebrospinal Fluid

¹ https://www.en.uni.lu/university/data_protection/data_protection_policy_and_faq

DEM	Dementia
DLB	Lewy body dementia
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
DPUK	Dementias Platform UK
DSAC	Data and Sample Access Committee
ELIXIR-LU	The Luxembourgish node of ELIXIR, the European infrastructure for life science information
ELSI	Ethical, Legal and Social Implications
EMIF-AD	European Medical Information Framework for AD
EPAD	European Prevention of AD
ETL	Extract, Transform, Load
ESAB	External Stakeholder Advisory Board
FAIR	Findable, Accessible, Interoperable, Reusable
FIP	FAIR Implementation Profile
FTD	Frontotemporal Dementia
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
IBBL	Integrated Biobank of Luxembourg
IPR	Intellectual Property Right
LIH	Luxembourg Institute of Health
MCI	Mild Cognitive Impairment
MDTA	Material and Data Transfer Agreement
MIP	Medical Informatics Platform
MSA	Multiple System Atrophy
NCBI	The National Center for Biotechnology Information
NDD(s)	Neurodegenerative Disorder(s)
NTK	NeuroToolkit
OMOP	Observational Medical Outcomes Partnership
PD	Parkinson's Disease
PET	Positron emission tomography
PHT	Personal Health Train
PREPAD	Participant Register for EPAD
PSP	Progressive Supranuclear Palsy

RBD	Isolated REM sleep behaviour disorder
REDCap	Research Electronic Data Capture
REMS	Resource Entitlement Management System
SCD	Subjective Cognitive Decline
SOP	Standard Operation Procedure
W3C DCAT	World Wide Web Consortium (W3C) Data Catalog Vocabulary
WP	Work Package

1. Project introduction and objectives

The EPND consortium aims to integrate existing initiatives and build, grow, and deliver the scalable and self-sustainable EPND for storage and analysis of high quality clinical and biological samples and data collections.

The specific objectives of the EPND project are to:

- Create a set of agreed ethical, legal, regulatory, business and sustainability principles to establish governance and operational processes to enable sharing and access to data and samples (Deliverables D2.2, D2.3).
- Establish a network that provides central and federated access to high quality samples and data from over 60 cohorts by combining existing data discovery and sharing initiatives within a global network of interoperable data sites and services provided by the Alzheimer's Disease Workbench (ADWB). In addition to enabling the sharing and access of data and samples, the platform will provide central and federated capabilities to ensure unified interpretation of biomarker data. (D1.1-D1.7).
- Provide visibility into the quality and standardization of the data and samples available in the platform from existing cohorts. Similarly, for prospective cohorts the development of SOPs will guarantee the quality of samples and data available, which will be important when focusing on the regulatory approval path for biomarkers (D3.3, D3.4).
- Perform a robust evaluation of the developing EPND platform through the execution of case studies on the Amyloid-Tau-Neurodegeneration (ATN) staging, complement activation, microbiome markers and molecular subtypes (D5.1-D5.3, D5.5).
- Embrace sustainability from the start of the project by designing and piloting exploitation strategies such that the platform is self-sustainable by the end of the project (D7.1-D7.4).

2. Data Summary

The following subsections provide a brief description of the data that will be generated/provided and exploited/shared/made accessible as part of the EPND project, specifically addressing the relationship between the data and the project objectives and indicating data-related responsible working packages / tasks/ milestones/ deliverables. Updates on the progress of data management will be reported in next versions of DMP (in D8.9, D8.14).

Existing and prospective (yet-to-be-collected/generated) data will be made FAIR for the researcher community via the EPND platform and can be findable worldwide through a global network of interoperable data sites and services as a part of the ADWB. Below, we explain the relation between data collection/ generation to the project objectives:

Objective 1: Create a set of agreed principles to enable data access

EPND is being developed as a trusted place for data hosting and sharing for data providers and

researchers, as well as for making derived data available research-ready in a fair, responsible, transparent, and legally compliant way. This will be done for data collected as part of past and future cohort studies, and interventional clinical trials. Activities to build this trusted data space will have deliverables in D2.2, D2.3 and D6.2

Objective 2: Establish a network that hosts high quality data, with federated and centralised elements

EPND aims to provide access to retrospective data on over 124,000 individuals with NDD, including Alzheimer's disease (AD), Parkinson's disease (PD), Lewy body dementia (DLB), with annual recruitment of 13,000 new patients, and over 45,000 controls (Table A.1, Annex A), facilitating thus unique opportunities for large-scale clinical studies on the pathophysiology of NDDs. The initial data and sample set targets more than 60 cohorts available for various diagnostic groups in AD, PD and DLB including:

- mild cognitive impairment (MCI);
- dementia (DEM);
- Isolated REM sleep behaviour disorder (RBD);
- subjective cognitive decline (SCD);
- cerebral amyloid angiopathy (CAA);
- frontotemporal dementia (FTD);
- amyotrophic lateral sclerosis (ALS);
- Progressive supranuclear palsy (PSP);
- Cortical-basal degeneration (CBD);
- Multiple system atrophy (MSA).

In addition, at least one prospective longitudinal cohort collecting and/or planning to collect digital biomarker data will be incorporated into the platform (T5.6).

Objective 3: Establish fair and transparent governance and processes to enable data sharing and access

The governance and data protection framework for EPND will be defined in Whitepapers (D2.2, D2.3), which will specify the role of the data and sample access committee (DSAC) (T8.5) and provide guidance on how to balance the interests of science and society, data and sample providers and users.

Objective 4: Evaluate the platform through the execution of case studies

The functionality of the EPND platform will be tested and fine-tuned with case studies on emerging AD and PD biomarkers in blood, cerebrospinal fluid (CSF) and stool samples using existing data discovery tools for samples and data selection, to monitor and validate the process of sharing different data types (T5.1-5.4). During the first 2 years, we will run case studies in the same core cohort including three diagnostic groups, namely AD, PD and DLB - which share pathophysiological processes - to reveal common pathologies across diagnostic entities and novel drug targets, if applicable. Data from both the case studies and from prospective longitudinal cohorts will be uploaded to EPND and the case studies will be performed through the EPND platform (T5.5-T5.6). Learnings from these case studies will be dynamically incorporated into the project's governance and processes for sustainable data sharing and access (T5.5, T5.6, T7.3).

Objective 5: Self-sustainable entity by the end of the project

The EPND platform, allowing data deposition, sample selection and management into the (federated) environment, will follow a "sustainability-by-design" approach; this strategy will i) detail the business and legal framework in which the EPND platform will operate (T4.4) and ii) formulate an access costing model based on fairness, feasibility and ethical acceptability (D7.2), and will be refined with integration of existing best practices and novel strategic approaches (D7.4).

EPND data flow chart

The EPND platform will be established as a central FAIR data and analysis platform utilising key aspects of ADWB's architecture (details follow in the next section). Cohorts will have an option to harmonize using the common dataset, to be defined as consisting of ~50 clinical and biobanking variables for AD, DLB and PD, (reported in D3.1 in M6 and D3.2 in M60 - details follow in the next section). The minimal dataset will be used for mapping the cohort data to the EPND selected ontologies

included in the EPND glossary.

The cohorts can be loaded into the respective database systems on-site or centrally at ADWB or UNILU. Datasets can be deployed at different sites in EPND including the ADWB directly and the Virtual appliance (i.e the ADWB component). In both settings, individual datasets can be managed by one or more data stewards and operated independently including handling metadata, data management and approving data access requests. Data can be updated over time as new versions become available.

The ADWB Virtual appliance component is responsible for performing computations on this data, implementing the Common API standard, as well as providing the means to control the outgoing results. The Common API is the standard that a component must provide to become part of the federated network. The researcher can either directly access this Common API to carry out analysis, or they do it by proxy via tools and services that are aware of the standard. In this scenario, all results must be approved by the on-site data stewards before they are released by the ADWB component.

Analysis across data held in different sites requires orchestration. A user performing analysis as part of a research project may do so from an environment in the EPND network that is setup and authorised to access the Common API. The setup and authorisation may include a mix of network security and identity and access management. For example this can be done from within a secure Workspace in the ADWB.

Figure 1 illustrates the flow of the EPND data to date. Integration of the case study results is currently under discussion, and we will include details on this in the next DMP versions.

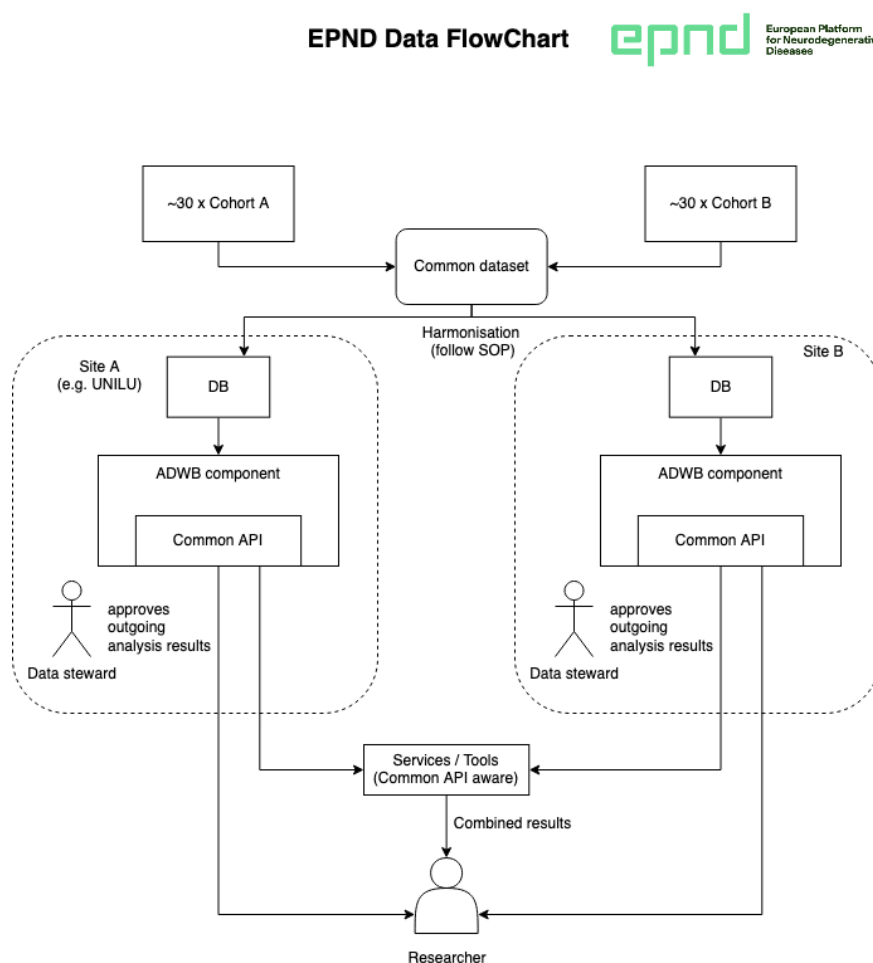


Figure 1 Data flow for the EPND project

Types of data that will be generated/collected/made accessible in the EPND project are dependent on the EPND case studies (coordinated by WP5), which focus on investigation of biomarkers in blood, CSF and stool samples in AD, PD and DLB diagnostic groups. A non-exhaustive list of the

EPND data and samples types includes:

- Plasma and CSF markers and positron emission tomography (PET) when available measures in n=350 individuals with AD, PD, DLB and controls (to be investigated in the first case study - CS1 Amyloid-Tau-Neurodegeneration (ATN) staging). CS1 will also include Neurochip genotyping of this Core Cohort. Complement pathway markers will be measured in mostly in CSF in the Core Cohort to identify potential correlation of complement activation to ATN stage (in CS2 Complement activation). Of note the O'link 3000 profiling of the Core cohort will provide much broader protein markers than just complement activation.
- Microbiome data from meta-analysis of existing data (n=2000) and stool analysis of a total of n=150 individuals with PD, AD and DLB (in CS3 Microbiome).
- CSF untargeted mass spectroscopy proteomics will be performed in n=1000 individuals with PD, DLB, and AD and controls to discover molecular subtypes and shared pathologies (CS4 Molecular subtypes).
- At least one additional case study on enrichment of existing prospective cohort studies is envisaged from year 3 (CS6 Enrichment of existing prospective cohort studies with an additional follow-up including collection of biosamples and digital biomarkers). Specifically, it aims at longitudinal data collection in at least one prospective study including AD, DLB, or PD individuals.

Datasets used or generated in this project as mentioned in the Table A.1 (Annex A) can be classified in three different data groups:

- structured data in the form of electronic medical records, XML or JSON;
- non-structured (raw) data including high-dimensional genome, proteome, metabolome, and microbiome;
- computational analyses developed using (1) and (2) – to be stored as whole models or using their parameters.

Details of further case studies and of the data types involved will be included in future versions of the DMP.

As stated in Objective 2 (above), EPND aims to provide access to retrospective data on over 124,000 individuals with NDDs (including AD, PD, DLB), with annual recruitment of 13,000 new patients, and over 45,000 controls (Table A1, Annex A). Data available in the 60 cohorts that were initially reviewed (based on the project proposal) is shown in Table 1 below; details on data size will be provided in the next versions of this DMP. Also, at least one prospective longitudinal cohort collecting and/or planning to collect digital biomarker data is planned to be incorporated into the platform (T5.6).

Table 1 Estimated size of the data available in the 60 cohorts initially reviewed

Type of data	Number of individuals with diagnostic markers available
Amyloid PET	6000
PET-Tau	1000
Neuropathology	4000
DaTscan	4000
Digital biomarker	1000
APOE	70000
MRI	70000

3. FAIR data

The FAIRness of all the research data will be improved by applying the principles of FAIR and using the tools (process, model, and cookbooks) developed by the GO-FAIR foundation (www.go-fair.org) as broadly as possible. The strategies to make EPND research data FAIR such as i) the selection of data standards and ii) relevant data extractions and transformations will be detailed in T1.2 (D1.2). Specifically, we will create a FAIR Implementation Profile (FIP) during the project (T1.2, M1.4) by collecting and evaluating CRFs and data dictionaries from cohorts and biobanks to survey the range of sample annotations, diagnostic criteria, clinical phenotypes and outcomes relevant to EPND. We will start from the ADWB & NTK platform metadata and ontologies/terminologies.

Via the EPND FIP, we will establish compatible 'common data elements' and core data models with respect to subject and sample annotations (including clinical parameters) and research/trial datasets. The cohorts (WP4) will regularly provision their relevant data fields using Extract-Transform-Load (ETL) and FAIRification services (WP1) along with curation/harmonization services that support custom and standard formats (e.g., CDISC, OMOP-CDM). This effort will be refined through the initial case studies (WP5) and further informed by regulatory and ELSI perspectives (WP2), which generally focus on establishing provenance metadata (when, how, by whom, on what (legal) grounds and why a data element came to exist), data/sample use conditions and restrictions.

We will test the first version of the FIP (M1.4) focusing on the needs of the case studies from WP5. After reaching a sufficient maturity level, we will extend it to all EPND partner cohorts and case studies. The full FIP (D1.3) will encompass:

- Core descriptive metadata
- Common EPND dataset supporting discovery and search
- Full datasets with metadata at field and controlled vocabulary level
- Semi-automated data request workflows (integrated with existing systems)
- Data delivery (direct transfer, remote query, federated API access)
- Full datasets harmonized to EPND common data elements.
- Sample metadata and availability (based on standard to be developed by the project)
- Semi-automated sample access request workflows

This FAIRification work will underpin data preparation for case studies, the production of common datasets, and integration of academic cohort datasets. Platform capabilities will later be extended to support high-dimensional data (e.g., genomics, metabolomics, proteomics, microbiome, imaging) and digital biomarkers, including connections to the IMI RADAR-BASE platform at KCL (www.radar-base.org). Details on the integration of the case study results in the database (and on the way these results will enrich it) will be included in the next versions of the DMP.

Here, we briefly present initial steps towards data FAIRification. Details will be included in updated DMP versions. Various EPND WPs collaborate to make current and prospective data FAIR for the research community via the EPND platform, which will rely on services offered by ADWB, as follows.

3.1. Making data findable, including provisions for metadata

The EPND platform will enable direct and indirect (remotely) data FAIRification, as well as integration of the data into the EPND Platform, alongside data indexation, discoverability, and research-readiness for secondary use. The EPND platform will be established as a central FAIR data and analysis platform using the ADWB's architecture, which will be developed by installing ADWB interoperability components and adapting existing data interfaces at UNILU in connection to an associated ADWB instance established on Azure via Aridhia (WP1).

We will follow the recipes in FAIR Cookbook² to generate and manage our dataset identifiers and we

² FAIR Cookbook: <https://faircookbook.elixir-europe.org>

will detail on the approach in the next DMP versions.

The EPND will provide data discovery services at several levels, as follows:

I. Cohort discovery:

We will develop the EPND catalogue that will enable cohorts to be discovered based on cohort level data. Specifically, for each cohort resource we will populate and regularly update the cohort level data model and employ this within the EPND platform to enable users to discover what cohort biobanks and datasets exist in EPND (T1.3).

As a basis for an EPND cohort resource "cohort level model", we will adopt the W3C DCAT standard used in the ADWB and part of the FIP to support best practice models used in existing cohort networks (T1.3). The scope of the DCAT model will be extended to include cohort level data elements (e.g., resource contacts, dataset provenance, scope of clinical data that could be provided), technical data, dataset dictionaries and properties, aggregate level statistics on samples/subjects in the cohorts, focusing on the needs from partner cohorts and biobanks (WP4) and case studies (WP5). Initially, a part of all EPND consortium partner cohorts, (e.g. EMIF-AD and EPAD) will be discoverable via the EPND platform, albeit with limited information; the number and depth of metadata records for driving cohort discovery will continually increase throughout the project.

The centralized catalogue of cohort resources will integrate the EMIF-AD catalogue (60 cohorts; software developed by BMD); BBMRI-ERIC biobank sample catalogue (668 biobanks, using MOLGENIS software developed by UMCG and also ELIXIR recommended interoperability resource); and the EATRIS catalogue (45 biobanks) into the ADWB catalogue on the EPND platform. There will be a single query portal that interrogates metadata information in the platform, with syndicated connections to also cross-query other relevant domain catalogues already established by project partners (e.g., EMIF-AD, EPAD, DPUK, BBMRI, EATRIS).

To enhance the visibility of cohort data offerings to the wider community we will develop the EPND Cohort Classification Scheme (D8.4, month 6), a clear classification scheme using three groups (bronze, silver, and gold) that will indicate the availability of data and samples to described scenarios. For example, distinguishing a Bronze level contribution (e.g. discoverability of simple cohort metadata) from a Gold level contribution (e.g. data harmonisation and discovery of patient level data). This clarity and visibility should both enhance cohort participation in the EPND platform, through aligning incentives to participating cohorts, and also demonstrate a robust metric upon which the value of the platform and hence the sustainability programme will be developed; e.g. a Gold level cohort will gain a range of support elements from the platform, and a platform of many Gold level contributions will be more attractive to customers/users of the ultimate delivery.

II. Data discovery at the record-level/ participant-level discovery:

EPND aims to enable safe, federated data discovery at the participant, or record-level by integrating and extending existing services of this type from the MIP (CHUV), PREPAD (ULEIC) and PHT (UM) environments, which currently encompass over 15 cohorts. It is planned to add 30 new cohorts to these discovery tools (T1.4).

Specifically, we will implement a central EPND query portal for samples and data subjects at the record level (T1.4). This query engine will operate in conjunction with the cohort discovery portal (T1.3). Industry and academic teams within and beyond EPND who are involved in neurodegeneration-related research will be able to use it to search for specific samples/subjects/data, or employ it to further explore the contents of cohorts of interest discovered at the cohort level (T1.3). Operation of the record-level query service will entail using a data model describing a common set of fields and associated export pipelines (T1.2) to regularly generate pseudonymised, minimal representations of samples and associated subjects. The query will be performed directly into the EPND platform or remotely via a tiered layer of existing federated query services. For biosamples, guided by WP3 and WP4, we will concentrate on basic information that is collected as part of standard clinical care and clinical trials. In addition to sample type, volume/size, these extra variables will describe the sample donor (demographics, clinical phenotype, interventions), and the basis of the sampling (visit number, timing, method used).

3.2. Making data openly accessible

The EPND platform will host retrospective and newly generated data via the Luxembourgish node of ELIXIR, the European infrastructure for life science information (ELIXIR-LU) platform, which also has the capacity to store high-dimensional genome, proteome, metabolome, and microbiome data if that becomes relevant to the project. Digital biomarker data will also be hosted, potentially enhanced via capacity on the IMI RADAR-BASE platform or similar platforms, with extracted features stored on the EPND platform utilizing some of ADWB's services in this case. Open-source software dockers will be provided (UM and ROCHE) to facilitate data analysis across the ADWB-enabled global network of interoperable data sites and services. The ELIXIR-LU node can provide access to central biobank facilities at Luxembourg Institute of Health (LIH) - Integrated Biobank of Luxembourg (IBBL) for WP5 case studies, as needed.

Safe and responsible procedures for requesting, accessing and handling data and materials will be supported by mechanisms in the ADWB, working in synergy with software in ELIXIR-LU (e.g. REMS, REDCap and ADA pipelines) (T1.5) and supervised by the Data and Sample Access Committee (DSAC, T8.5) and informed by guidance from Whitepapers 1 and 2 (D2.2, D2.3).

User queries and responses will be logged and monitored for any suspicious behaviour. Data privacy considerations (WP2) will guide query limitations, content obfuscations, differential privacy evaluations, and response restrictions on different classes of user, to further protect cohort subjects (T2.2). With respect to GDPR, data interrogated by these services will be "anonymous in context" and so safe to employ as envisaged. An elemental version will be in place by the end of the first year, with a complete system being in place by month 30 (M1.6) as guidance comes through from WP2 regarding data protection and ethics. Syndication with other record-level querying services of project partners will be based on the OpenAPI specification, standard query APIs (e.g., Beacon-2³ from Global Alliance for Genomics and Health (GA4GH)), and approved user permissions plus inter-site trust tokens by which record-level queries initiated in the EPND Platform can also interrogate a federated network of 30 additional cohorts and/or biobanks already connected to existing query tools of project partners (PREPAD, NCDC via PHT, MIP, BBMRI). This queried information will be compatible with the same core data models and be "anonymous in context".

3.3. Making data interoperable

In T3.1, we will define a common dataset of 50 variables that capture key clinical variables and sample annotation variables for AD, DLB and PD, based on the analysis of datasets used in ongoing multicentre studies such as the Parkinson PPMI study, data platforms such as GAAIN, C-Path consortium, DPUK, and consensus recommendations from JPND working groups (PMID: 28412978; PMID: 26523894). The **common dataset** (reported in D3.1 in month 6 and D3.8 in month 60) will serve for the ontology mapping (in WP1) and will be used in WP1 participant level data discovery and WP5 case studies. It will include patient-level data describing disease progression and patient phenotypes, integrated with key established biomarkers and genetically-driven targets. The common dataset will also benefit from key Pharma involvement (NOV, UCB, TAKEDA) and collaboration with the Critical Path for Parkinson's (CPP) consortium - a precompetitive initiative aimed at advancing novel drug development tools for regulatory endorsement. The SOP will be revised based on experience with case studies in T5.1-T5.6. (D3.8) and may include variables from other NDDs if applicable.

We will extend the ADWB data dictionary (T1.2) to establish data and related metadata core standards, including an 'EPND Glossary', by building on global standards of this type (e.g., NCBI sample metadata) via collaborative work between WP1 and domain experts and clinical partners from WP3 (T3.3/D3.3), WP4 and WP5.

Data harmonisation will be performed at different levels:

- **Cohort level** (WP4 and WP1): detailed mapping of the cohorts who have joined EPND, providing details on the cohort characteristics as defined by clinical and biosample data (quality, completeness, source, typology). The main objective is not to harmonize data capture and SOPs for all cohorts, but to agree on requirements for effective cohort exchange and pooling. WP4 is responsible for mapping appropriate clinical variables, and WP1 will establish the ontologies and perform ETL mapping of cohort data to the EPND common data model and associated glossary.
- **High dimensional data:** Defined as clinical and molecular data such as genomics, transcriptomics,

³ <https://beacon-project.io/#beacon-v2-scope>

proteomics, metabolomics, and associated metadata will be curated and harmonised using standard terminologies and controlled vocabularies for the case studies (CS) (WP5): CS3 (T5.3: microbiome) and CS4 (T5.4: CSF proteomics).

3.4. Increase data re-use (through clarifying licences)

All foreground and background IP necessary to exploit the platform will be licensed by the owners to the exploiters of the platform as defined in the sustainability plan (see Long-term data sustainability plan in the next section). The terms will be captured in a license agreement (T7.4 in month 48). After month 48, the sustainability plan will remain a living document, updated according to emerging knowledge, including the outcomes of the pilot review in T7.3 in month M50, or from the WP4 webinar organised with EPND contributing cohorts in T4.3.1 in month M54. This section will be updated in the next DMP versions. WP2 will develop guidance around standard data use conditions and agreements in the White Paper to facilitate sharing and re-use.

4. Allocation of resources

List of Responsibilities

WP8 coordinates FAIR data management planning across the project in collaboration with all WPs: WP1 (platform), WP2 (legal/ethical aspects), WP3 (clinical and sampling SOP), WP4(cohort interaction), WP5 (case studies), WP6 (stakeholder involvement) and WP7 (sustainability). In summary,

- WP1 coordinates development of a FAIR data platform, which aims to increase discovery of research data and samples in line with the overall objectives of EPND to meet the needs of the larger global community of ND researchers, including the selection of data standards and relevant data ETLs as described in T1.2 (D1.2) and reported as a “FAIR Implementation Profile” (M1.4) and online demonstration of various FAIR-related platform components (D1.3 - D1.6).
- WP2 has an ELSI officer to assist partners with ethical and legal compliance, and to evaluate compliance costs for Cohorts to inform value-added ELSI services for the final platform, and to assist WP7 on final sustainability plan.
- WP3 coordinates development of an EPND glossary as the basis for creating common datasets of ~50 clinical and biobanking variables for each disease area (AD, DLB, and PD), which will serve record-level discovery, case studies, and prospective data collection.
- WP4 coordinates FAIRification at cohort level and mapping of minimal dataset variables identified in WP3 from each cohort to the common data dictionary developed within WP1.
- WP5 coordinates harmonisation of high dimensional data (such as genomics, transcriptomics, proteomics, metabolomics data) as part of case studies: CS3 - T5.3: microbiome and CS4 - T5.4: CSF proteomics.
- WP6 coordinates stakeholder feedback including regulatory considerations that facilitates project sustainability.
- WP7 coordinates the final sustainability plan of the project (T7.4, D7.4), in collaboration with WP2.

Long-term data sustainability plan

Self-sustainability of the platform by the end of the project is the main priority of EPND, and therefore it will be tackled from the start of the project, guided by the sustainability and exploitation plan (D7.1 - D7.4), a living document to be updated as the needs of the community, service gaps, and technical potential, are better understood. EPND has a dedicated work package (WP7), whose overall goal is to define and execute an exploitation strategy ensuring long-term sustainability of the EPND platform, through provision of high value services and resources to the international research community. Specifically,

- WP7 will develop the sustainability and exploitation initial plan (D7.1) with exploitation strategies during the course of the project;

- WP7 will collaborate closely with WP2 and the External Stakeholder Advisory Board (ESAB) to define the legal, regulatory, and operational principles relevant to sustainability (D2.2);
- WP7 will deliver the Final Sustainability Plan (D7.4), including the Minimum Viable Product (MVP), and potential additional services and updates from the needs of the community, service gaps, and technical potential. The Final Sustainability Plan will comprise detailed information on operations, business, funding plan, governance structure, principles governing the relationship between sample/data depositors, the platform and users, policies and templates for handling of access requests, IP, confidentiality, re-use of data, and publication for each user type (industry, academic, citizen). It will also include checklists for data/sample deposition (with WP4), future platform operations processes, and quality-related requirements developed in WP3.

5. Data security

5.1. Formal information/data security standards

The EPND data platform (WP1) will naturally emphasize technical security and data privacy, in close collaboration with WP2. A 'privacy-by-design' workshop to review the necessary security and privacy measures will be held between WP1 and WP2 in month 8 (T2.2). These efforts will be bolstered by the ADWB environment which by then will already provide many key aspects, not least logging of data related activities (viewings, queries, downloads, analyses, etc.) and authentication and authorization of users.

5.2. Data Protection Impact Assessment (DPIA)

The workshop referenced in section 5.1 will support WP2 in conducting a DPIA of the EPND platform and identify possibilities for improvement, which we will implement thereafter (D2.1). We will include updates in the further DMP versions. The DPIA⁴ of the EPND platform will ensure the processing of personal data through the platform takes place in a manner that protects the fundamental rights and freedoms of data subjects. Another important goal of the DPIA is to give cohorts confidence that the use of the platform is GDPR-compliant, promoting its wider adoption by the EPND consortium.

In accordance with Art. 35(7) of the GDPR, the DPIA will encompass the following components:

- a systematic description of the envisaged processing operations within the EPND and the purposes of the processing
- an assessment of the necessity and proportionality of the processing operations in relation to the purposes;
- an assessment of the risks to the rights and freedoms of data subjects
- the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned

Importantly, as the DPIA concerns the EPND platform and ADWB, it cannot address specific future research projects that will be utilizing personal data through the platform in an exhaustive manner. Hence, the scope of the DPIA will be platform-centered, primarily focusing on the services and functionalities of the EPND platform that relies on ADWB functionality and services. However, there is an opportunity to enhance the value of the DPIA by laying down guiding principles and recommendations concerning yet-to-be-defined future use cases (research projects) involving Cohorts providing data to external research organisations. This will benefit the community of future data controllers by providing them with a helpful framework for identifying legal responsibilities and mitigating global risks relevant to their research project, as well as demonstrating GDPR compliance.

It should be emphasized that the deliverable D2.1 in M10 will be the first iteration of the DPIA, whose content is expected to evolve substantively in the subsequent years. As the exact design,

⁴<https://gdpr.eu/data-protection-impact-assessment-template/>

functionality, and governance of the platform will still be under development, the DPIA will be a living document that will be refined over the course of the project.

5.3. Data Sharing and Access

5.3.1. Suitability for sharing

The data sets made available within EPND will be limited to those data items relevant to NDD research (data minimisation principle). The number of data transfers will be minimized and in EPND, we envision to perform federated analysis to reduce the risk of accidental further data dissemination (data protection by design principle). Any data transfer and data access will be carried out in accordance with a signed material and data transfer agreement (MDTA). This agreement, to be elaborated by WP2 based on the template MDTA within the CA, will ensure data are only used for approved purposes, that recipients establish appropriate technical and organizational safeguards, and that data are only kept for the duration necessary to complete the proposed use. Since cohort owners will remain the source of data/sample access and use conditions (i.e., determining what constitutes 'approved purposes'), it is anticipated that they may have flexibility to introduce certain data specific access and use conditions. WP2 will provide support and guidance for this process.

WP1 will provide support to WP8 in handling data requests (both central and federated) and data access (T1.5). We will use the REMS system (already operational at UNILU) and potentially additional components available through other partners (e.g., GA4GH passports, BBMRI 'negotiator' request tooling) to develop IT support solutions that will be provided in the EPND platform. Such developments will thus add to existing capabilities provided by the ADWB for access request management, namely: customizable forms, business process support and data delivery procedures when supporting data partners. These functions will be (a) implemented for each partner providing biosample or data resources, accomplished via a limited set of templates, and (b) enhanced to support the additional requirements of managing access to both biorepositories and cohort assets. Additional capabilities will be established to log all actual sharing events in a manner that ensures follow-up and compliance with access/sharing agreements. During Phase I (i.e., the first year), to enable data use as required by WP5, the project will employ traditional communication and documentation mechanisms, operated through the EMIF-AD and BBMRI catalogue interfaces and in collaboration with WP2 and WP6. Support will transition from the basic methods to ADWB-based solutions as they become operational, with this being independent of cohort availability timelines set by the work of WP4.

5.3.2. Discovery by potential users of the research data

EPND data can be findable worldwide through a global network of inter-operable data sites and services as a part of the ADWB while respecting the ethical and legal guidelines set out by the data governance. The current plan has following options:

1. Primary discovery tooling at epnd.org (EPND node);
2. Secondary discovery as part of ADWB;
3. Optional Syndication at EPND partner catalogues.

Details will be provided in the next DMP versions.

5.3.3. Governance of access

The Whitepapers (D2.2, D2.3) will define the governance and data protection framework for EPND: will specify the role of the data and sample access committee (DSAC) (T8.5) and provide guidance on how to balance the interests of science and society, data and sample providers and users.

The Patient Expert Group (PEG, WP6) will also contribute to the development of EPND principles for data and sample sharing (T2.1).

International data access by scientists from outside Europe is overseen by the DSAC and the access policies governed by these DSAC will be set up the first Whitepaper (D2.2 by WP2), after identifying appropriate legal mechanisms and appropriate safeguards for international transfers. The EPND's federated design will also help to minimize the risks associated with international transfers, where foreign researchers are required to conduct their analysis in a secure European computing environment.

5.3.4. Restrictions or delays to sharing, with planned actions to limit such restrictions

User queries and responses will be logged and monitored for any suspicious behaviour. Data privacy considerations (WP2) will guide query limitations, content obfuscations, differential privacy evaluations, and response restrictions on different classes of user, to further protect cohort subjects (T2.2).

5.4. Main risks to data security

This section will be updated in further DMP versions based on outcomes from the ‘privacy-by-design’ workshop (T2.2 in month 8) and DPIA (D2.1 in month 10).

6. Ethical aspects

The dedicated WP2 (Legal and ethical regulations) is responsible for coordinating the work related to ethical issues bearing on the EPND. The scope of this work is multifaceted, encompassing governance, and data protection aspects, alongside strictly ethics considerations.

An overarching objective of WP2 is to ensure respect for the rights, freedoms, and reasonable expectations of the data subjects whose personal data and biological samples are used in the EPND. To this end, WP2 will coordinate multiple related activities.

First, WP2 will seek to document, where feasible, the expressed preferences of the data subjects regarding the use of their personal data and samples in research projects. This activity will rely on the review of informed consent forms and their associated patient and/or research participant information materials used by EPND cohorts. These documents are valuable sources for identifying data and sample use conditions that are in line with data subjects’ reasonable expectations, values, and preferences. The review will be part of a broader activity that entails surveying EPND cohorts to identify data/sample use conditions or restrictions specific to each cohort. The ethical-legal questionnaire, including review of informed consent forms and information materials, has been initiated as part of the development of the MDTA for the case studies in WP5 and will continue to be used as additional cohorts are onboarded to EPND.

Second, WP2 will prepare specific recommendations for addressing cases where the reasonable expectations of data subjects cannot be readily established. These cases may arise where informed consent is missing or has been deemed inadequate in addressing secondary use of samples/data in future research projects. The challenge of non-existent or inadequate informed consent for future research is common in the context of legacy samples and their associated data, which are collections obtained opportunistically in a clinical setting without a clearly defined roadmap for subsequent reuse. Assessing the lawfulness of reusing such collections in the EPND is essential for determining whether, and under what conditions, these resources can be included in the platform.

WP2 will also document that all necessary approvals such as research ethics committee approvals are obtained where applicable to data/sample sharing as well as research projects using data and samples.

Third, in close collaboration with WP1 (particularly BBMRI), WP2 will help build data and access governance framework around EPND resources such as sample and data collections. The framework will include the Data and Sample Access Committee (DSAC) tasked with assessing and deciding on data/sample access requests from external researchers. The access governance framework will additionally include various policies and contractual instruments, such as a data/sample access policy and data/sample transfer or use agreements. These instruments will act as additional safeguards for ensuring the protection of the rights and reasonable expectations of the data subjects.

The blueprint for these policies and frameworks will be laid out in the EPND White Papers, led by WP2. The White Paper v.1.0 (D2.2; M12) is expected to be particularly instrumental in this regard, as it explores the most critical challenges facing the EPND from ethical, legal, governance, regulatory, and sustainability perspectives. In addition to the ethics-centred part of the White Paper, WP2 will be driving the work related to legal aspects, focusing primarily on data protection and GDPR compliance. Some of the specific legal issues related to data protection will include:

- Assessing the lawfulness of data sharing for research purposes under the different legal bases, within the meaning of GDPR Article 6(1).

- Determining whether and to what extent the rights of the data subjects apply, especially in the context of GDPR Articles 15 – 22
- Ensuring transparency requirements towards data subjects, in accordance with GDPR Articles 13 and 14
- Facilitating compliance with other GDPR principles including, among others, data minimisation, purpose limitation, and accountability, as well as additional relevant data protection obligations

The Whitepaper will rely on the substantive input from multiple EPND partners and will result in concrete recommendations for informing the governance of the EPND platform, alongside ethical and legal compliance matters.

In addition to tackling the overarching ethical and legal issues in a comprehensive manner, WP2 will provide ongoing support to the consortium, offering ad-hoc guidance regarding ethical and legal compliance. This work has been initiated by WP2 in the early months of the EPND and is currently focused on assisting consortium members with the implementation of case studies (led by WP5). WP2 will work closely with WP6 that coordinates the engagement of stakeholders including regulators, industry, patient associations, researchers and care providers.

7. Other issues

Non-research data management

In addition to managing the EPND research data (as described in previous sections), we will rely on GDPR-compliant processes while also adhering to the Consortium Agreement (e.g. protection of results, other fundamental IPR/ matters, dissemination and publication policies etc.) to manage data and information generated within the project. This includes information or data generated by various WP (e.g. surveys or process documentation) and information disseminated internally and externally (including communication internal to the project consortium and external with the scientific community, healthcare professionals, industry, workshop reports, meeting minutes, interviews, surveys etc.). This non-research data and information generated as part of the EPND project will be stored on the Project's Teams and accompanying SharePoint site, which are only accessible to project participants.

8. Updates of this Document and feedback

This DMP is a living document and will be reviewed and will be updated throughout the project whenever significant changes arise. At a minimum, two major updates will occur at mid-term and at project-end respectively (D8.9, D8.14).

9. Conclusion

The first version of the data management plan has been set up and agreed among the parties.

10. Deviation

There are no deviations from the project plan.

Annex A: Table A.1 Characteristics of third-party cohorts who have expressed interesting in joining EPND

Partner	Institute	Cohort	(co-)PI	Country	Setting	Record Disc	Focus					Data	Samples	Access							DE					
							Cog. nor	AD spec	PD spec	DLB spec	Open			Pre	Post	Open	Pre	Post	Open	Pre	Post	Open	Pre	Post		
P	CHUV	CLEMENS	Demonet	CH	C	MIP	AD	244	620	20	40	y	n	y	y	n	y	n	n	n	s	y	y	s	s	s
P	UNIGE	SNAP/gMAD	Frisoni	CH	C	MIP	AD	61	163	1	20	s	s	y	y	y	y	n	y	n	s	y	y	y	s	s
P	UNIGE	Geneva outpatient clinic	Frisoni	CH	C	PRE	AD	241	307	5	16	s	s	s	n	s	n	n	n	n	s	n	y	n	s	s
P	UM	BBACL	Verheij	NL	C	n	AD	342	349	4	0	y	y	s	y	y	y	n	n	n	n	y	y	n	n	n
P	VUmc	ADC	Scheltens	NL	C	PHT	AD	1200	2449	0	260	y	s	y	y	y	y	n	s	n	n	y	y	s	s	n
P	VUmc	EMIF PreAD/90+	Visser	NL	R	PHT	AD	210	90	0	0	y	y	y	y	y	y	n	n	n	s	y	y	y	n	
P	VUmc	EPAD-VUmc	Visser	NL	R	PHT	AD	150	37	0	0	y	y	y	y	y	y	y	n	n	n	y	y	n	n	n
P	UNIGE	Bus Sante	Stringhini	CH	P	PRE	Age	700	0	0	0	y	y	n	y	y	y	n	n	n	n	y	n	n	n	n
P	UM	Maastricht study	Stehouwer	NL	R	PHT	Age	9000	0	0	0	y	y	n	y	y	y	y	y	y	n	n	y	n	n	n
P	VUmc	LASA	Huisman	NL	P	PHT	Age	4000	0	0	0	y	y	n	s	s	s	s	n	s	n	s	n	n	n	n
P	DZNE	4 Cohorts	Klockgether	DE	R	n	Mix	1375	320	544	5	y	y	s	y	y	y	n	n	y	y	y	s	n	n	n
P	DZNE-Tuebingen	Tuepac	Gasser	DE	C	n	PD	94	0	1466	108	y	y	y	y	y	y	y	y	y	n	n	y	n	n	y
P	DZNE-Tuebingen	TREND	Gasser	DE	C	n	PD	1200	25	220	0	n	n	n	n	n	n	n	n	n	n	y	y	n	n	n
P	UNILU	NCER	Krueger	LU	C	n	PD	1632	0	681	17	y	y	s	y	y	y	y	s	y	y	n	s	n	n	s
P	VUmc	PROGRESS-PD	v.d. Berg	NL	C	n	PD	52	0	149	0	y	y	y	y	y	y	n	n	n	n	n	n	n	n	y
P	UOXF	Discovery	Hu	UK	C	n	PD	322	0	1241	0	y	y	s	y	y	y	n	n	n	y	y	s	n	n	s
A	Multicentre Europe	E-DLB	Aarsland	UK	C	n	DLB	0	0	0	900	y	y	s	s	s	y	n	n	n	n	s	s	n	n	s
A	Multicentre NL	PROPARK	v Hilten/v.d. Berg	NL	C	n	PD	265	0	1250	0	y	y	n	y	y	y	n	y	n	y	n	n	n	n	y
A	Uni. Cambridge	CamPaLIGN	Williams-Gray	UK	C	n	PD	160	0	179	0	y	y	n	s	s	s	n	n	n	n	n	n	n	n	n
A	Uni. Glasgow	ProBaND	Grosset	UK	C	n	PD	0	0	2266	0	y	y	n	n	y	y	n	n	n	y	n	n	n	n	n
A	Uni. Newcastle	ICICLE-PD	Lawson	UK	C	n	PD	99	0	257	0	y	y	s	s	s	s	n	n	n	y	s	s	n	n	n
A	Uni. Cambridge	PICNICS	Williams-Gray	UK	C	n	PD	118	0	276	1	y	y	n	s	s	s	n	n	n	n	n	n	n	n	n
A	KU Leuven	LEU_DEM	v.d.Berghe	BE	C	n	AD	495	237	0	20	n	n	n	y	y	y	n	n	n	y	s	s	n	n	n
A	KU Leuven	ARCK	v.d.Berghe	BE	C	n	AD	111	190	0	0	y	y	s	y	y	y	n	n	n	n	s	s	s	s	n
A	KU Leuven	F-PACK	v.d.Berghe	BE	R	n	AD	233	79	0	0	y	s	s	y	y	y	n	n	n	n	s	y	s	s	n
A	UMC Saint-Luc Brussels	UCL-2016/NCHM	Hanseeuw	BE	C	n	AD	294	228	15	17	s	n	s	s	s	s	n	n	n	n	s	s	s	s	s
A	IMI Pharmacog	PharmaCog	Frisoni	CH	R	n	AD	0	81	0	0	y	y	y	y	n	n	n	n	n	n	y	y	n	n	n
A	UMC Motol	CBAS	Hort	CZ	C	n	AD	92	324	7	25	s	s	y	s	n	n	n	n	n	n	s	n	n	s	s
A	UMC Berlin Charité	Memory Clinic	Peters	DE	C	n	AD	176	1137	0	0	y	y	y	y	y	n	n	n	n	y	s	s	n	n	n
A	UMC Munich	ActiGIA	Ewers	DE	R	n	AD	28	92	0	0	y	y	y	y	y	y	s	y	y	y	y	y	y	s	s
A	University of Erlangen	DCN multi center study	Kornhuber	DE	C	n	AD	243	1870	0	0	y	y	y	y	y	y	n	n	n	n	y	s	n	n	n
A	Danish Dem Biobank	DDB	Waldemar	DK	C	n	AD	230	1095	30	205	s	n	y	y	y	y	n	n	n	y	s	y	n	s	s
A	FACE Barcelona	Memory clinic	Ruiz	ES	C	n	AD	3930	11380	108	503	y	n	s	s	s	y	s	n	n	y	s	s	n	s	s
A	Hosp. St Pau Barcelona	SPIN	Lleó	ES	C	n	AD	330	484	0	135	y	y	y	y	y	y	n	n	n	n	y	s	n	s	s
A	CITA-AD Foundation	GAP/DEBA	Martinez Lage	ES	R	PRE	AD	499	165	0	0	y	y	y	y	y	y	n	n	s	n	y	y	n	n	n
A	Hosp. Salpetrière Paris	SOCRATES	Dubois	FA	R	n	AD	0	158	0	20	y	y	y	y	n	n	n	n	n	y	s	y	n	n	s
A	Lille Memory Clinic	COMAJ/YOD/Baltazar	Pasquier	FR	C	MIP	AD	70	760	0	16	y	y	y	y	n	y	n	n	n	n	n	y	n	n	n
A	Hosp Lariboisière Paris	BioCokBank	Henon	FR	C	n	AD	300	670	30	160	y	n	y	y	n	y	n	n	n	n	y	y	y	n	n
A	Hosp. Salpetrière Paris	Insight pre-AD	Dubois	FR	R	n	AD	230	103	0	1	y	y	s	y	n	n	n	n	n	n	y	y	n	y	n
A	Uni. Athens (NKUA)	ALBION	Scarmeas	GR	C	n	AD	90	65	0	0	y	y	s	s	s	s	n	s	n	y	s	s	n	n	n
A	Uni. Thessaloniki	Memory Clinic	Tsolaki	GR	C	n	AD	340	1700	65	125	y	y	y	y	n	y	s	s	n	s	y	y	n	n	n
A	IRCCS Brescia	ARWIBO	Redolfi	IT	C	MIP	AD	261	217	0	0	y	n	y	y	y	y	n	n	n	n	y	y	n	n	n
A	Uni. Genoa	Genoa-EADC	Nobili	IT	C	n	AD	70	600	220	80	y	n	s	s	s	n	n	n	n	n	s	s	n	s	s
A	Uni. Milano	CAA cohort	Piazza	IT	C	n	AD	0	600	0	0	y	y	s	y	y	y	n	n	n	n	s	y	n	s	n
A	Uni. Perugia	ReGAL 2.0	Mecocci	IT	C	n	AD	255	490	0	0	y	y	n	y	y	y	n	n	n	n	n	n	n	n	n
A	UMC Oslo	PROSP/OS/DDI	Fladby/Pihlstrøm	NO	C	n	AD	418	173	76	0	y	y	s	s	s	n	n	n	n	n	y	s	n	s	s
A	UMC Coimbra	Memory clinic	Santana	PT	C	n	AD	300	1270	140	140	y	s	s	s	s	y	n	n	n	n	y	s	n	s	n
A	Uni. Lisbon	BBA	Mendonça	PT	C	n	AD	0	228	0	0	y	n	y	y	y	y	n	n	n	s	n	y	y	n	n
A	UMC Carol Davila	ANALS	Spiru	RO	C	n	AD	20	45	25	17	n	n	s	y	y	s	y	s	y	y	s	y	n	n	n
A	Uni. Gothenburg	Gothenburg MCI	Wallin	SE	R	n	AD	362	500	0	0	y	y	y	y	y	y	n	n	n	n	y	y	n	n	n
A	Karolinska Institute	GEDOC	Winblad/Kivipelto	SE	C	PRE	AD	4700	6000	50	50	y	n	y	y	y	y	n	n	n	n	y	y	s	s	n
A	Uni. Athens (NKUA)	HELIAD	Scarmeas	GR	P	n	Age	3100	67	154	9	y	s	n	s	s	s	n	n	n	n	s	n	n	n	n
A	Erasmus MC	Rotterdam Study	Ikram	NL	P	PHT	Age	13000	800	0	0	y	y	n	y	y	s	n	n	n	y	y	s	n	s	n
A	RIVM	Doetichem cohort	Verschuren	NL	P	PHT	Age	4700	0	0	0	y	n	n	y	y	y	n	y	y	y	s	n	n	n	n
A	Lifelines	Lifelines	Dotsing	NL	R	PHT	Age	1600	0	0	0	n	n	n	y	y	y	n	n	n	n	y	s	n	n	n
A	Uni. Aveiro	iBiMED	Herdeiro	PT	C	n	Age	500	9	0	0	y	n	n	y	y	y	n	n	n	n	y	n	n	n	n
A	Imperial Col London	Chariot	Middleton	UK	R	n	Age	450	0	0	0	n	n	y	y	y	y	y	n	y	y	y	y	n	y	n
A	UMC Brussel	Memory clinic	Engelborghs	BE	C	n	Mix	100	200	150	50	y	y	s	s	s	s	n	n	n	n	s	s	n	s	s
A	Uni. Antwerp	Memory clinic	Engelborghs	BE	C	n	Mix	350	2750	260	100	s	s	s	s	s	s	n	n	n	n	s	s	n	s	n
A	KNAW	DutchBrainBank	Huitinga	NL	R	n	Mix	175	776	247	82	n	n	y	y	y	y	n	n	n	n	n	n	n	n	n
A	UMC Utrecht	SMART-MR	Geerlings	NL	R	PHT	Mix	1269	0	0	0	n	n	n	y	y	s	s	n	s	n	n	n	n	n	n
A	Uni. Dokuz Eylul	Turkish MCnetwork	Yener	TR	C	n	Mix	97	41	60	0	n	n	n	n	n	n	n	n	n	n	n	y	n	n	n
A	UMC Innsbruck	Memory clinic	Seppi	AT	C	n	PD	350	300	330	15	s	s	s	s	s	n	n	y	n	s	n	s	n	n	n
A	UMC Goettingen	Kassel	Mollenhauer	DE	C	n	PD	0	0	205	20	n	n	y	y	y	n	n	n	n	n	n	n	n	n	s
A	UMC Goettingen	KAMSA	Mollenhauer	DE	R	n	PD	0	0	250	0	n	n	n	y	y	n	y	y	y	n	n	y	n	n	s
A	UMC Goettingen	DeNoPa	Mollenhauer	DE	R	n	PD	80	0	150	0	y	y	s	y	y	y	y	y	y	n	y	s	n	s	n
A	UMC Virgen Rocio Sevilla	Movement Clinic	Mir	ES																						

Partner type: P=partner; A=affiliated partner; Record discovery: cohorts currently linked to MIP, PHT or Prepad; Setting: C=clinical, R=research, P=population-based; Main focus cohort: AD=Alzheimer's disease, PD=Parkinson's disease, DLB=Lewy Body dementia; Mix: mixed; Age: ageing study; Availability: n=no, y=yes, s=subset.