European Platform for Neurodegenerative Diseases

IMI2 101034344 – EPND

European Platform for Neurodegenerative Diseases

WP3 – SOPs

D3.1 - SOP minimal dataset for data glossary

Lead contributor (partner organisation)	15 – UOXF		
Other contributors	**1-UM**	**6-UNILU**	**17-KCL**
	2-ULEIC	**9-UGOT**	**23-SARD**
	4-DZNE	**14-UNIGE**	**26-GV**
	5-VUmc	**16-Aridhia**	

Document History

Version	Date	Status
V1	**19/01/2022**	**Draft**
V2	**28/04/2022**	**Comments**
V3	**03/05/2022**	**Comments**
V4	**19/05/2022**	**Final**

Glossary:

AD	Alzheimer's Disease
ADDI	AD Data Initiative
AMP-PD	Accelerating Medicines Partnership - Parkinson's disease
AMYPAD	Amyloid imaging to Prevent Alzheimer's Disease research initiative
CPP	Critical Path for Parkinson's consortium
DLB	Lewy body dementia
GAAIN	Global Alzheimer's Association Interactive Network
NEURONET	IMI project to create an overall platform for efficient collaboration, communication and operational synergies among present and future IMI neurodegenerative diseases projects
PD	Parkinson's Disease
SOP	Standard Operation Procedure

Abstract

The purpose of this SOP is to define a minimally-sized common use concept model, serialized as a data model, that captures key clinical variables and sample annotation variables for Alzheimer's disease (AD), Lewy Body Dementia (DLB), and Parkinson's disease (PD). It includes relevant ontology mappings and will be used to support WP5 case studies. It will also provide the basis for a simpler and more constrained data model needed by WP1 for patient-level data discovery.

The model encompasses patient-level data describing disease progression and patient phenotypes, integrated with key established biomarkers and genetically-driven targets. The model design had input from AD, DLB, and PD disease experts within the consortium, and the leads of the first case studies, in collaboration with the Critical Path for Parkinson's (CPP) consortium - a precompetitive initiative aimed at advancing novel drug development tools for regulatory endorsement. It is expected this common data model will be revised by the end of the project, based on experience with case studies in WP5, and may eventually incorporate variables from other neurodegenerative disorders, if applicable.

Methods

Regular meetings were held with key stakeholders across relevant work packages (WP1,3,4,5) for the model delivery. After several discussions, two approaches to the core dataset variables were considered, and we decided to divide the data model into an analysis and discovery core (see below).

The ***Analysis Core*** comprises *patient-level data* describing disease progression and patient phenotypes integrated with established biomarkers. It focuses on major cohort datasets that will be most useful for secondary use, the inputs and result datasets from the WP5 case studies, and as many other enthusiastic external cohorts that can be found and logistically supported for preparation and connection to their analysis content. The analysis core model was *limited to ~50 data fields per disease for ease of use*.

The Analysis Core:

- consists of approximately 50 data fields per disease (AD, DLB or PD) rather than 100s-1000s
- is split into common building block domains applied across all 3 disorders (see below)

- contains harmonised values (e.g. units, z-scores, precision, ontologies, etc.)
- has a clear rationale for the functionality/purpose of each included element
- shows alignment to existing models via an agreed framework/representation
- was informed by models from global leading initiatives

The ***Discovery Core*** will be derived from the Analysis Core in conjunction with WP1, concentrating on a basal set of 10-20 concepts and associated data fields that are modified in design to facilitate meaningful sample and patient discovery. This will be guided by cohort datasets and data platforms used by all existing cohort discovery networks in EPND, ~15 new cohorts, and all cohorts which will be involved in the WP5 case studies.

There are key existing cohorts, initiatives, and data platforms across the three neurodegenerative disease areas (AD, DLB, and PD), several being IMI-funded, that helped to shape and inform EPND's minimal dataset, and are listed below:

AD focused:

- AMYPAD: Amyloid imaging to prevent Alzheimer's disease
- EMIF-Switchbox: predominant AD data sources IT platform
- ADDI: Alzheimer's Disease Data Initiative, AD and related dementias
- EPAD: European Prevention of Alzheimer's Dementia Consortium
- NEURONET: an initiative connecting IMI-funded projects in AD, PD, MS, etc
- GAAIN: Global Alzheimer's Association Interactive Network

PD focused:

- AMP PD (Accelerating Medicines Partnership) dataset- includes BioFINDER, HBS (Harvard biomarkers study), PDBP (PD biomarkers program) and PPMI cohorts.
- GP2 AMP PD core/minimal clinical data elements
- GEoPD minimal dataset

DLB focused:

- E-ELB and ENLIST DLB Consortia core/minimal clinical data elements

For biosample datasets, these were compared with Critical Path for Parkinson's (CPP) Consortium PD Biomarker Inventory.

Dataset Domains

For ease of use, the analysis core is divided into **higher level concept blocks, or domains, that are commonly applied across the three diseases.** Within the domain, data variables are grouped to reflect:

a) an overarching variable (e.g. amyloid biomarkers which include Amyloid PET, amyloid- b CSF and amyloid-b blood measures); or
b) that they can be interpolated from each other (e.g. MOCA and MMSE, Sniffin 16 item and UPSIT, which are parallel measures of cognitive/olfactory function respectively).

Data variables will also be stated as continuous or categorical elements, various data types, and level of optionality. Each analysis core building block is split into two parts. The first holding variable information, and the second holding linked variables within the same building block.

Using the approach of grouping the 150+ variables by domains, the three clinical leads within EPND for AD, DLB, and PD produced a short list of key disease-specific variables from existing cohort variables datasets collected (see list above). Existing cohort variables from these key cohorts were amalgamated into one table, then further refined, discussed, and adapted to ensure they were fit for purpose for EPND.

List of Analysis Core: Domain and subdomains

Domain	Subdomain	AD relevant	DLB relevant	PD relevant
AUTONOMIC AND SENSORY		Y	Y	Y
BIOMARKERS	CSF	Y	Y	Y
	DNA	Y	Y	Y
	EXOSOME	Y	Y	Y
	FAECES	Y	Y	Y
	PBMC	Y	Y	Y
	PLASMA	Y	Y	Y
	SALIVA	Y	Y	Y
	SERUM	Y	Y	Y
	SKIN/FIBROBLASTS	Y	Y	Y
	TISSUE	Y	Y	Y
	URINE	Y	Y	Y
COGNITIVE		Y	Y	Y
CURRENT MEDICATION STATUS		Y	Y	Y
DEMOGRAPHICS		Y	Y	Y
DIAGNOSIS		Y	Y	Y
DISEASE SPECIFIC		Y	Y	Y
ENVIRONMENTAL/BEHAVIOURAL HISTORY		Y	Y	Y
FAMILY HISTORY		Y	Y	Y
FUNCTIONALITY AND QUALITY OF LIFE		Y	Y	Y
GENERAL		Y	Y	Y
IMAGING BIOMARKERS		Y		
MEDICAL HISTORY		Y	Y	Y
MENTAL HEALTH		Y	Y	Y
MOTOR			Y	Y
SLEEP		Y	Y	Y
VITAL SIGNS		Y	Y	Y

Conclusion

EPND has produced a minimal, common-use dataset of 159 variables covering clinical features (eg diagnostic, demographic, medication status, family history, functionality, vital signs etc), as well as imaging and bio-sample variable features. The Excel file with the full list of variables and their descriptors is available upon request and will continue to be further refined within the 5 years of the project. Below is a screenshot of a subset of the Biomarkers domain.

Domain	Subdomain	Description	Name	Available	Units	Type	Format
BIOMARKERS	DNA	Have DNA samples been banked?		Y		factor	(Y,N)
BIOMARKERS	DNA	If yes, on which date were DNA samples banked?		Y		date	YYYY-MM-DD
BIOMARKERS	DNA	If yes, are DNA samples available for future research?		(Y,N)		factor	(Y,N)
BIOMARKERS	DNA	If yes, has genotyping been performed on the DNA samples?		Y		factor	(Y,N)
BIOMARKERS	DNA	If yes, which DNA genotyping platform was used?		Y		string or factor?	
BIOMARKERS	DNA	APOE (Apolipoprotein E) genotype		Y		factor	
BIOMARKERS	DNA	Are there any GBA (Glucosylceramidase beta) mutations present?		Y		factor	(Y,N)
BIOMARKERS	DNA	If yes, which GBA mutations are present?		Y		string or factor?	
BIOMARKERS	EXOSOME	Have exosome samples been banked?		(Y,N)		factor	(Y,N)
BIOMARKERS	EXOSOME	If yes, on which date were exosome samples banked?		(Y,N)		date	YYYY-MM-DD
BIOMARKERS	EXOSOME	If yes, are exosome samples available for future research?		(Y,N)		factor	(Y,N)
BIOMARKERS	FAECES	Have faeces samples been banked?		N		factor	(Y,N)
BIOMARKERS	FAECES	If yes, on which date were faeces samples banked?		N		date	YYYY-MM-DD
BIOMARKERS	FAECES	If yes, are faeces samples available for future research?		N		factor	(Y,N)
BIOMARKERS	PBMC	Have PBMC (Peripheral Blood Mononuclear Cells) samples been banked?		(Y,N)		factor	(Y,N)
BIOMARKERS	PBMC	If yes, on which date were PBMC samples banked?		(Y,N)		date	YYYY-MM-DD
BIOMARKERS	PBMC	If yes, are PBMC samples available for future research?		(Y,N)		factor	(Y,N)
BIOMARKERS	PLASMA	Have plasma samples been banked?		Y		factor	(Y,N)
BIOMARKERS	PLASMA	If yes, on which date were plasma samples banked?		Y		date	YYYY-MM-DD
BIOMARKERS	PLASMA	If yes, are plasma samples available for future research?		(Y,N)		factor	(Y,N)
BIOMARKERS	PLASMA	Plasma: Amyloid beta 1-42		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: Amyloid beta 1-40		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: NFL (Neurofilament Light Chain)		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: Phosphorylated Tau isoform 181		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: Phosphorylated Tau isoform 217		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: Phosphorylated Tau isoform 231		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: Total Tau		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Plasma: GFAP (Glial Fibrillary Acid Protein)		(Y,N)	pg/ml	decimal	
BIOMARKERS	PLASMA	Was plasma proteomics performed?		(Y,N)		factor	(Y,N)
BIOMARKERS	PLASMA	If yes, what plasma proteomics platform was used?		(Y,N)		string or factor?	
BIOMARKERS	SALIVA	Have saliva samples been banked?		(Y,N)		factor	(Y,N)
BIOMARKERS	SALIVA	If yes, on which date were saliva samples banked?		(Y,N)		date	YYYY-MM-DD
BIOMARKERS	SALIVA	If yes, are saliva samples available for future research?		(Y,N)		factor	(Y,N)