

IMI2 101034344 – EPND

European Platform for Neurodegenerative Diseases

WP4 – Cohort Interaction

D4.5 Summary metadata of 60+ cohorts from the catalogue, 30 cohorts linked to record-level discovery tool

WP4	
Lead contributor (partner organisation)	1- UM
Other contributors	6 – UNILU 26 – GV 2 – ULEIC

WP4	
Due date	31/10/2025
Delivery date	
Delivery type ¹	R
Dissemination level ²	PU

¹ Use one of the following codes:

R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

² Use one of the following codes:

PU = Public, fully open, e.g. web;

CO = Confidential, restricted under conditions set out in Model Grant Agreement;

CI = Classified, information as referred to in Commission Decision 2001/844/EC.

Description of Action

V1.0	
-------------	--

Document History

Version	Date	Status
V4.5.0	30/09/2025	First draft by UM
V4.5.1	13/10/2025	Draft including partner input
V4.5.2	30/10/2025	Final

Glossary

Cohort	Study
Center	Institution conducting the study
Dataset	Record-level data collected by a cohort
Biosample collection	Biosamples collected by a cohort available for sample analysis
Metadata	Information on available record-level data or biosamples
EPND Hub	Digital infrastructure to centralize and harmonize discovery and access to biosample collections and datasets for neurodegenerative disease research
EPND Catalog	Section of the EPND Hub including the cohort metadata

Abstract

Throughout years 1-4, one of the tasks of WP4 was to gather metadata from over 60 cohorts with data and/or biosamples on neurodegenerative diseases and 30 cohorts with record-level data to be made discoverable via the European Platform for Neurodegenerative Diseases (EPND)'s Hub. The EPND Hub is a digital infrastructure to centralize and harmonize discovery and access to biosample collections and datasets for neurodegenerative disease research. A total of 108 cohorts has been onboarded by WP4 into the EPND Hub, covering 277328 participants across 14 diseases. High level metadata from all cohorts was collected and in collaboration with WP1 was made available in the EPND Catalog (within the EPND Hub). To allow more in-depth cohort discovery, additional metadata on available dataset(s) was collected for

41 cohorts, and with available biosample collections for 35 cohorts. In both cases, information on use-conditions and related publications was also collected. For 17 of these cohorts, *ad-hoc* processes were

defined with cohort owners/managers and implemented to allow cohort data or samples to be requestable directly through the EPND Hub. In addition, 77 datasets requestable via the Alzheimer's Disease (AD) Data Initiative's AD Workbench have been made discoverable with associated metadata in the EPND Catalog through interoperability with the EPND Hub, for a total of 118 datasets discoverable via the EPND Hub. Finally, discovery of 78 record-level datasets has been enabled by making available pre-aggregated record-level information with core metadata via a federated discovery network licensed to EPND.

Introduction

One of the main tasks of WP4 throughout years 1-4 was to onboard over 60 cohorts with data or biosamples from patients with neurodegenerative diseases from different European centers and make their metadata available via the EPND Hub. Additionally, 30 cohorts had to be selected for record-level discovery. In this deliverable, we report the main procedures and outcomes of this task.

Methods

Metadata discovery

Between November 2021 and June 2022, WP4 defined a set of high-level cohort information to be requested to cohorts for initial onboarding into EPND. Starting from July 2022, WP4 disease area specialists (DAS; AD: UM and GV; PD: UNILU, DLB: KCL) reached out to cohorts who manifested interest in EPND during the project proposal stage (summarized in Table 3.3 in the Description of Action) to collect this high-level cohort information via an online form in the EPND Hub. When possible, WP4 pre-filled the known cohort information to assist cohort owners/managers. Between September 2022 and December 2023, WP4 DAS and leads worked closely with WP1 to expand the initial set of requested metadata with additional information on available data, biosamples, use conditions, and related publications and to implement it within a form in the EPND Hub. This exercise leveraged the DAS expertise across several neurodegenerative diseases and past initiatives (e.g., data fields used in the EMIF-AD catalog). We aimed to define a metadata set that was not burdensome for cohort owners/managers to complete, and that was user-friendly to fill out within the EPND Hub. We reached out to our larger networks and offered cohort owners/managers the possibility to enter metadata

information directly on the online Hub or via a structured Excel form which we could use to add information to the Hub afterwards. Additionally, we defined processes to allow EPND Hub users to contact cohorts via the EPND DAS for collaborative research. In collaboration with WP1, metadata of datasets listed in the AD Data Initiative's AD workbench were also made discoverable via interoperability with the EPND Hub. Since January 2025, WP4 has collected more detailed use conditions information of the cohorts, in line with established representations of consent and use conditions (DUC-CCE profiles) and incorporated more streamlined data and/or sample request procedures to facilitate access to the data and samples of cohorts by EPND Hub users.

Record-level federated discovery

Possibilities for federated discovery of datasets represented by metadata and record-level derived aggregate data were explored and piloted thoroughly by WP1, as explained in detail in D1.5. WP4 contributed to this process by providing input on metadata and aggregate data types useful for discovery, as well as by providing cohorts for piloting discovery solutions.

Results

Metadata discovery

The first set of high-level metadata to be requested to the cohorts for initial onboarding into the EPND Hub was finalized by WP4 in June 2022. The information requested in this set is presented in Table 1. In total, 108 cohorts have been onboarded by WP4 DAS and leads, and cohort metadata have been made available in the EPND Catalog within the EPND Hub (<https://discover.epnd.org/catalogue>, Figure 1). These cohorts represent a total of 277328 participants across 14 disease areas.

A second set of metadata was finalized by WP4 and WP1 in January 2024 to achieve more in-depth cohort characterization and discovery, also as related to available cohort datasets, biosamples, and related use conditions and publications. The requested information is summarized in Table 2. These more detailed metadata were collected on 41 datasets and 35 biosample collections and are available in the EPND Catalog after login. An example of a study that provides detailed metadata on datasets and biosample collections is shown in Figure 2. Further, detailed information on 77 additional datasets

was made available on the EPND Catalog via the AD workbench of the AD Data Initiative, bringing the total number of datasets discoverable via the EPND Catalog to 118.

To date, WP4 has collected more-detailed use-condition information for 17 cohorts, in line with DUC-CCE profiles, and incorporated in EPND more streamlined data and/or sample request procedures to facilitate access to the data and samples of these cohorts by EPND Catalog users. Most of these datasets were made requestable via the AD Workbench.

Table 1. Overview of high-level metadata collected for 108 cohorts.

Item	Subitem
Cohort name	
Cohort acronym	
Cohort description	
Primary site or institution	
Cohort primary investigator	Name, email
Cohort contact point	Name, email
Is the cohort multi-centre or multi-country?	Multi-site, single-site, multi-country
Year cohort started	
Are you still collecting data and/or samples?	Yes/No
Number of participants per disease area	• Control group • AD (preclinical, prodromal, dementia based on biomarkers, or clinical dementia) • PD (normal cognition, MCI, dementia) • DLB (MCI, dementia) • Isolated REM sleep behaviour disorder • CAA • FTD • ALS • PSP • CBS • MSA • Huntington • Ataxia
Have you collected biosamples?	Yes/No
Sample type	• CSF • Serum

	• Plasma • DNA • Saliva • Faeces • Urine
Do you have imaging data available?	Yes/No
Imaging types	• MRI • Amyloid PET • Tau PET • DaT scan
Do you have cognitive data?	• None • Cross-sectional • Longitudinal

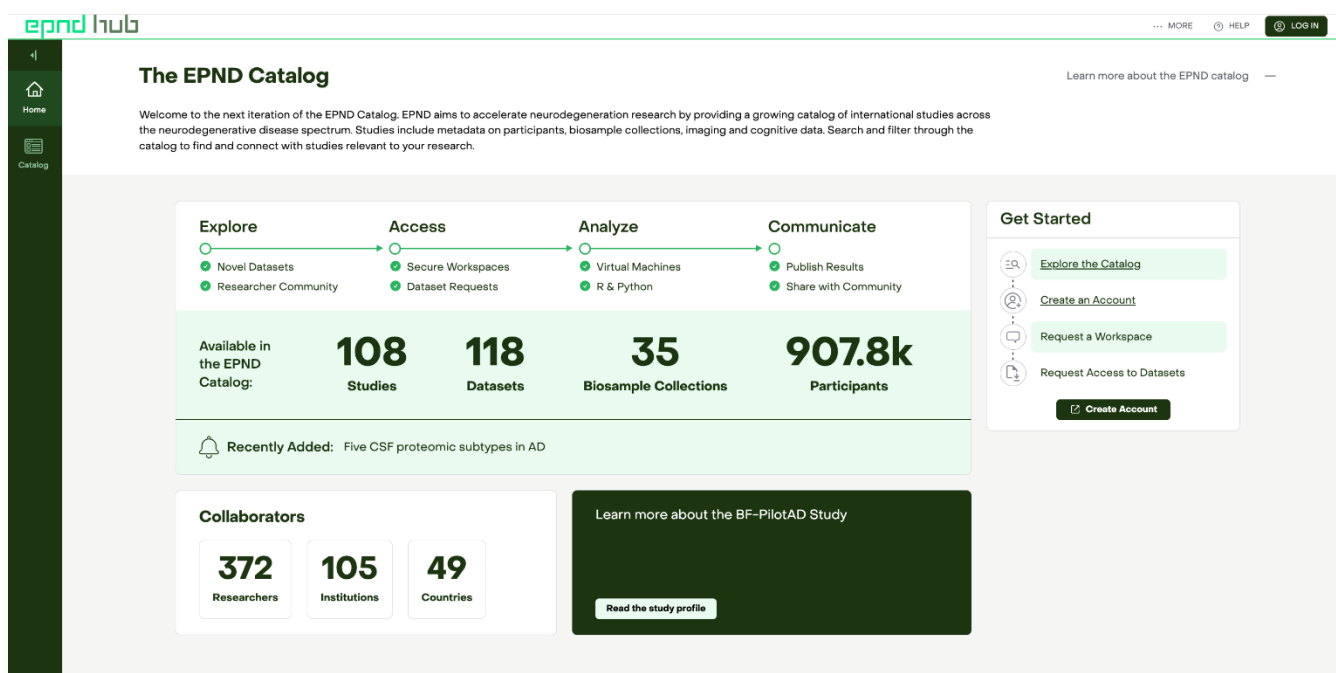
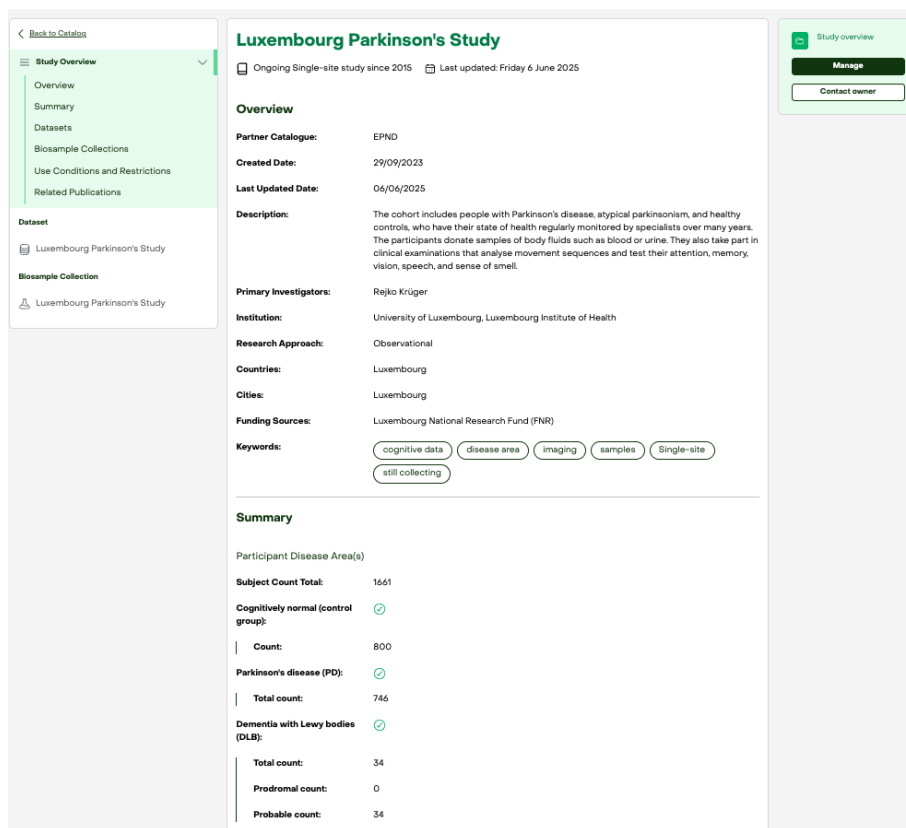


Figure 1. Overview of cohort metadata available via the EPND Catalog (<https://discover.epnd.org/>)

Table 2. Overview of in-depth metadata collected for a subset of EPND cohorts.

Domain	Item	Subitem
Cohort overview	Basic information (Table 1)	
	Cities	
	Funding sources	

	Number of participants per disease area	
	Number of biosamples per sample type	Cross-sectional or longitudinal collection
Datasets	Demographic coverage	Age range Sex coverage
	Dataset information	Published date Total number of participants Disease areas
	Variables collected	Clinical information Functional ratings Neuropsychiatric Assessment Neuropsychological Assessment Quality of Life Sleep scales Diagnostic markers Digital data Imaging Electrophysiology Neuropathology
Biosample collections	Demographic coverage	Age range Sex coverage Ethnicity coverage
	Basic information	Total number of donors Disease areas Variables collected
	Collection details	SOPs Quality
	Biosample Collection Types collected per each disease area	Sample count Cross-sectional or longitudinal collection
Use conditions	Use Conditions and Restrictions	
	Digital Use Conditions	
Related Publications	List of key cohort publications	



Luxembourg Parkinson's Study
Ongoing Single-site study since 2015 | Last updated: Friday 6 June 2025

Overview

Partner Catalogue: EPND

Created Date: 29/09/2023

Last Updated Date: 04/06/2025

Description: The cohort includes people with Parkinson's disease, atypical parkinsonism, and healthy controls, who have their state of health regularly monitored by specialists over many years. The participants donate samples of body fluids such as blood or urine. They also take part in clinical examinations that analyse movement sequences and test their attention, memory, vision, speech, and sense of smell.

Primary Investigators: Rejko Krüger

Institution: University of Luxembourg, Luxembourg Institute of Health

Research Approach: Observational

Countries: Luxembourg

Cities: Luxembourg

Funding Source: Luxembourg National Research Fund (FNIR)

Keywords: cognitive data, disease area, imaging, samples, Single-site, still collecting

Summary

Participant Disease Area(s)

Subject Count Total:	1661
Cognitively normal (control group):	800
Parkinson's disease (PD):	746
Dementia with Lewy bodies (DLB):	34
Prodromal count:	0
Probable count:	34

Figure 2. Example of a study with detailed metadata in the EPND Catalog.

Federated dataset discovery

Direct discovery of record-level data posed major challenges for both technical operations and data collection, with cohort owners/managers expressing concerns about potential data breaches and facing legal and administrative barriers such as patient consent requirements. Thus, as detailed in D1.5, an intermediate solution between pure metadata and pure record-level data was chosen, consisting of making available core metadata plus pre-aggregated record-level information via federated discovery. Via this approach, discovery queries can be launched in one catalogue, propagated through a network of other catalogues, and finally presented to the user as an aggregated set of results.

Specifically, as explained in D1.5, a dedicated application programming interface (API) was designed by WP1 to support federated discovery, and this itself was based on the GA4GH standard called Beacon-2. Both by coding a Beacon-2 service from scratch, and by adopting Café Variome v3 (CV3) software (which has inbuilt Beacon-2 functionality), five EPND Partners together created a federated discovery network to pilot this concept. As part of this work, CV3 was licensed to EPND by ULEIC, and

its component modules were adapted to support the relevant metadata/data models, query logic, result interfaces, and network messages. Using these solutions, a federated discovery demonstrator network has been enabled across 5 nodes (DZNE, CHUV, ULEIC, BMD, and UNILU) comprising 78 datasets/cohorts, along with an empty copy of the CV3 software which acts as the main query portal for this network. Registered EPND Hub users can explore this network, available at <https://cafevpilot.alzheimersdata.org>. The cohort and dataset metadata model discoverable via this network is available at <https://github.com/BrookesLab/standards/blob/main/cafe-variome-3/draft/schema/meta-source-dataset.json>.

Discussion

We have reached and exceeded the planned discovery of metadata from 60 cohorts and the record-level federated discovery of 30 cohorts. To date, we have gathered for discovery via the EPND Catalog high-level metadata from over 100 cohorts, representing more than 270000 participants across 14 disease areas. We have additionally collected more detailed information on 41 datasets and 35 biosample collections, together with use-conditions and related publications. Additionally, we have enabled discovery of other 77 datasets on the EPND Catalog via the AD Data Initiative's AD workbench, bringing the total datasets discoverable via EPND to 118. Further, for 17 cohorts we have obtained information on use conditions in line with DUC-CCE profiles and facilitated more streamlined data and/or sample request procedures for data and sample access. Overall, the metadata collected in the EPND Catalog brings together the European neurodegenerative research community and can enable collaborative research efforts and advancements. As a tangible result, we have to date received cohort collaboration requests from 32 EPND Catalog users. Requests are first reviewed for completeness by the EPND Disease Area Specialists and are then forwarded to the cohort owners and managers. For the 17 cohorts that opted for closer involvement with the EPND Hub and specified more detailed use conditions, tailored processes were defined in the Hub to streamline data and/or sample request procedures and allow them to route directly to cohorts.

The initial cohort onboarding via high-level information was well-received and promptly accepted by cohort owners/managers, who welcomed the opportunity to increase the discoverability and visibility of their studies through EPND and expressed enthusiasm about joining a large community on neurodegenerative disease research. However, gathering more detailed cohort information at a second stage to support in-depth discovery and tailored collaboration requests were challenging for different

reasons. First, this was constrained by the limited time available to cohort owners/managers, who gave lower priority to providing more detailed metadata after having already provided high-level cohort information. The fact that clear incentives for providing more detailed cohort information were still being defined within EPND during that request phase further hindered the process. Another challenge we encountered was that some cohorts without a clearly organized infrastructure for data curation and sharing were reluctant to engage in further data sharing efforts. Additionally, due to project timelines, cohort onboarding took place while the catalog interface was still under development, which at times made the user experience of cohort owners/managers difficult.

To address these challenges, we took some corrective actions. First, we reached out to cohort owners/managers in a personalized manner, tailoring outreach email templates and identifying EPND members that could conduct specific outreach where possible. Additionally, we shared Excel templates to facilitate data entry for cohort owners who preferred this format over the Catalog and offered phone calls to provide individual support. We also developed manuals and “how-to” video demonstrations to support Catalog use, and organised onboarding webinars and Q&A sessions in collaboration with WP6. Furthermore, we identified new cohorts and onboarded them directly with more detailed information, thus expanding the initial target of 60 cohorts.

Related challenges were encountered when exploring options for record-level discovery. Sharing or providing access to record-level data, even in a decentralized manner, was generally not well received by cohort owners/managers due to concerns about data breach, time-consuming legal/administrative procedures or limited participant consent forms. To overcome this, WP4 supported WP1 to explore different options for record-level data discovery which could be suitable for cohorts, technically feasible and sustainable for EPND (reported in D1.5). A licensing contract had to be put in place for software (CV3) needed by EPND to test and enable alternative ways to use record-level information for data discovery. Because of this, timelines for record-level federated discovery took longer than initially planned and postponed D4.5.

Repository of primary data

All cohort subject level data is stored on the local servers at the cohort sites. Metadata for cohorts are available through the EPND Hub via the EPND Catalog (<https://discover.epnd.org/catalogue>), where associated datasets and biosample collections can be requested in accordance with defined access and use conditions. Registered EPND Hub users can explore record-level federated cohort metadata at <https://cafevpilot.alzheimersdata.org>. The associated metadata model is available at <https://github.com/BrookesLab/standards/blob/main/cafe-variome-3/draft/schema/meta-source-dataset.json>.

Conclusion

Overall, we collected metadata from 108 cohorts for discovery via the EPND Catalogue, covering 277328 participants across 14 diseases. Additionally, we supported WP1 in enabling federated discovery of 78 record-level datasets. While gathering high-level cohort metadata was relatively straightforward, obtaining information on more-detailed cohort characteristics or record-level assets proved more challenging, and was solved by pre-aggregating record-level content. The resulting EPND Catalog is a promising tool to promote collaborative research efforts and advancements. We will continue expanding the Catalog by onboarding new cohorts and collecting more detailed information from existing ones.