

IMI2 101034344 – EPND

European Platform for Neurodegenerative Diseases

WP1 – Establishment of the platform

D1.4 – Online demonstration of resource-level (catalogue) discovery service, with at least 30 participating cohorts/biobanks

Lead contributor	11 - BMD
Other contributors	18 - GV 2 - ULEIC

WP1	
Due date	30/04/2024
Delivery date	M30
Delivery type	DEM
Dissemination level	PU

Description of Action	
V3.0	15/08/2024

Document History

Version	Date	Status
V1.0	18/04/2024	Draft
V1.1	03/05/2024	Edits

V1.2	04/05/2024	Edits
V1.3	21/08/2024	Final

Abstract

This Deliverable describes progress made within EPND towards enabling online users to find dementia-related research assets of interest, including Cohorts, Datasets and Biosamples. We report success in creating a powerful catalog in a Central Hub, powered by the development of an extendable and interoperable metadata model that echoes the DCAT standard. A remote demonstrator node that syndicates (i.e., allows download, via APIs) their metadata records is described, along with a proof-of-principle version of a Beacon-v2 based federated discovery capability. In total over 70 cohorts have been onboarded to become discoverable within this EPND system, greatly exceeding the number initially targeted for this Deliverable.

Methods

During the progress of the **European Platform for Neurodegenerative Diseases** (EPND) initiative, it was established – within “Task 1.4” – that the use of federation methodologies and cohort-supplied “metadata” would be employed to promote interoperability between the EPND Catalogue and participating cohorts and biorepositories.

Unlike more standard, centralized approaches (i.e., those being singular entities aggregating and storing all project data) federated methodologies allow our multiple cohorts to keep their data and choose to share it in a network of other nodes. The burden of keeping, managing, and versioning data is now distributed by each cohort, instead of the EPND Catalogue being a singular point of consultation and data storage.

Metadata Selection

For the federation and the EPND Catalogue, it was decided that the Catalogue is to be a central discovery “hub”: it serves as a primary consultation point for biorepositories and cohorts (i.e., the federated nodes) that are participating in the federated network. Users (researchers, scientists) may visit the Catalogue to discover where relevant datasets are stored and what cohorts manage these. Requests to the Catalogue involve consulting all participant nodes – by programmatic means, mainly REST API – for the presence of relevant data for given user-submitted queries.

To make datasets discoverable in cohort sites, some form of data must be shared to identify what sorts of datasets are available. The information shared regarding each dataset must not be explicitly the private/sensitive information, but only enough “basic” fields to help describe and identify what sorts of datasets each are, such as:

- Dataset name, description, country, city, created/updated date
- Site identity, primary investigators, institution name
- Number of research participants

- Type of dataset (biosample collection type, ...), keywords (serves as additional query context)

Data of this sort, which may be aggregated and grouped in user-requested queries, are classified as “metadata”. To ensure that this federated network of cohorts and biorepositories shares reliable, standardized metadata, it was agreed to formulate a “Metadata Model” which must be followed to provide information in a predictable format for easier and automated processing of query results.

Bronze/Silver Models

The model cannot be deemed to be, or used as, a “complete” solution without validation by real users, as we would be blindly creating a complete model that would very likely have to be adjusted due to issues with adoption by the cohorts – which would take more time. Additionally, requesting cohorts to adopt a new method of sharing data that requires following a large model would be an onerous undertaking for cohorts.

Given the above, the model was split into 3 tiers that represent increasingly granular levels of metadata (i.e., number of fields) that the cohorts would be asked to provide. Each tier of the model builds off the previous, representing metadata completeness:

1. Bronze model
 - Minimal data for onboarding of only Cohort-related information
 - General fields such as research participants, investigators, ...
 - The smallest, and easiest to integrate for cohorts
2. Silver model
 - Higher level of metadata model, for Cohorts, Datasets and Biosample Collections
 - Includes Minimal data (Bronze)
 - Additional fields regarding the type of dataset (general participant/biosample data)
 - Broad demographics and collection details
 - Cohorts can build off their work integrating the bronze model for the additional requirements.
3. Gold model
 - Highest-tier, for Cohorts, Datasets and Biosample Collections - currently in development
 - Includes Minimal data (Bronze)
 - Includes Additional context (Silver)
 - Even further context on a per-dataset basis
 - Finer participant/biosample data (finer demographics, collection details)
 - Technical specifics, such as dataset field lists
 - Record-level aggregation/statistics (non-intrusive)
 - Cohorts can build off the silver module to add further requirements.

The modelled approach provides a more streamlined development and lowers the burden of participation from cohorts, allowing EPND ample opportunities to receive valuable feedback and reinvent each model as cohorts experiment with this layered system.

DCAT

In abstract design terms, we decided to base the metadata structure on pre-existing, proven standards instead of starting from scratch [1]. Specifically, the model echoed the W3C Data Catalog (DCAT) vocabulary, a semantic web concept model focused on aiding the discoverability of datasets/services and facilitate federated search. This very same standard has been in use by the Alzheimer’s Disease

(AD) Data Initiative, in their AD Workbench web service, and is additionally a part of EPNDs FAIR Implementation Profile (FIP).

Our research into the DCAT vocabulary introduced a catalog-oriented structure in which the components – the “classes” – can be grouped and linked logically.

All semantic web data structures and fields are data-agnostic, meaning that the vocabulary is structured and worded in a way that it can be appropriate for any type of data, even if it is not related to medical contexts. That said, a new DCAT for Health data model will be released by W3C during 2024, and we will examine that to see if matching changes in our modelling may be required.

Most classes in the DCAT vocabulary are inheritors of the “dcat:Resource” class. This means that all classes act as “resource published or curated by a single agent”, and contain multiple fields that help describe themselves, such as a title, identifier, version, and many more. The dcat:Resource class is not intended to be used directly, only serving as a supplier for common fields among the other DCAT classes.

The “dcat:Dataset” class, which inherits from dcat:Resource, is defined as a collection of data that is published by an agent and made accessible – either by download or other means – in one or more representations. It additionally has some unique fields that give dataset-oriented context, such as spatial/temporal context, periodicity, and more.

dcat:Dataset classes can be grouped and aggregated by using “dcat:Catalog” instances. These classes serve to group the datasets into a “curated collection of metadata”, giving context to the associated datasets. The fields available to dcat:Catalog classes include some unique fields (“dcat:themeTaxonomy”, ...) and fields that it inherits from the dcat:Dataset and dcat:Resource classes. With this large collection of fields – coupled with the data-agnostic nature of the vocabulary – the dcat:Catalog class is quite versatile and can richly define metadata for datasets with little contextual restrictions.

By mirroring the general design structure of the DCAT vocabulary, an initial design for the metadata model was created. All versions of the dataset metadata model will always have bronze-level fields, these being fields present in the dcat:Resource and dcat:Dataset such as identifier, title, and others. The silver and gold metadata models, on the other hand, will introduce more granular fields that are context-specific to the EPND initiative. These subsequent models will include contextual information regarding, for instance, biosample collection and patient demographics.

Table 1: Abbreviated view of bronze-level fields, acting as a “base” for all metadata model tiers.

Text/Question	DCAT Property
Acronym / ID	Resource.identifier
Name	Resource.title

Description	Resource.description
Publisher	Resource.publisher
Contact Point	Resource.contact point
Date Published	Resource.release date
Last Update	Resource.update/modification date
Dataset version	Resource.version
Temporal Coverage	Dataset.temporal coverage
Spatial Coverage	Dataset.spatial/geographical coverage
Access rights	Resource.access rights
License	Resource.license
Main reference	Resource.is referenced by

Beacon-V2 API

In addition to developing a common metadata model for interaction across nodes, we have explored other avenues to support data discovery. Specifically, the GA4GH Beacon-V2 API standard was deemed promising for data syndication between the various instances, and so it was decided that this API would be used as an additional interface between EPND Catalogue and cohorts.

The Beacon-V2 specification we are in the process of creating will define a basic discovery model for datasets, and also concretely defines the behaviour and structure of its associated REST API, along with how the service presents itself to clients.

Results

Using the above methodologies and principles, a proof-of-concept solution was devised to validate the approaches. Via such testing, the design of the EPND Catalogue regarding federation and interoperability with Cohorts has been solidified and reiterated to its current state-of-the-art.

1) BMD Node — EPND SN

As an initial implementation, the BMD Software was used to validate the chosen modelling/federation approaches, along with the use of Beacon-V2 for discovery. This produced a demonstrator EPND Silver Node (EPND SN).

The EPND SN web service, employed pre-existing BMD capabilities regarding listing and displaying of datasets and questionnaires related to biostudies and neurodegenerative diseases. These functionalities were extended. Specifically, the EPND federation model was added to the service's

REST API to provide capabilities for looking up and displaying dataset metadata from additional remote servers using the then-new bronze metadata model.

In addition to federation support, the EPND SN service was endowed with rudimentary Beacon-V2 support, being able to provide service context information and accept queries from clients.

Developing the proof-of-concept served as an important discussion point to refine federation and associated methodologies.

Framework

The EPND SN web service was made possible by incorporating internally a web framework (Montra) developed in the Python programming language which enables the aggregation and sharing of biomedical data from one or multiple data sources [2, 3]. BMD Software already adopts the use of Montra for the EMIF-AD platform, serving it as a catalog that stores many neurodegenerative disease-oriented datasets from several cohorts around the globe [4].

In essence, a Montra instance is a “catalog” that houses one or more data sources – labelled “communities” – as shown in Figure 1. The communities can be of various types, such as cohorts, collections, and patient-centric documents.

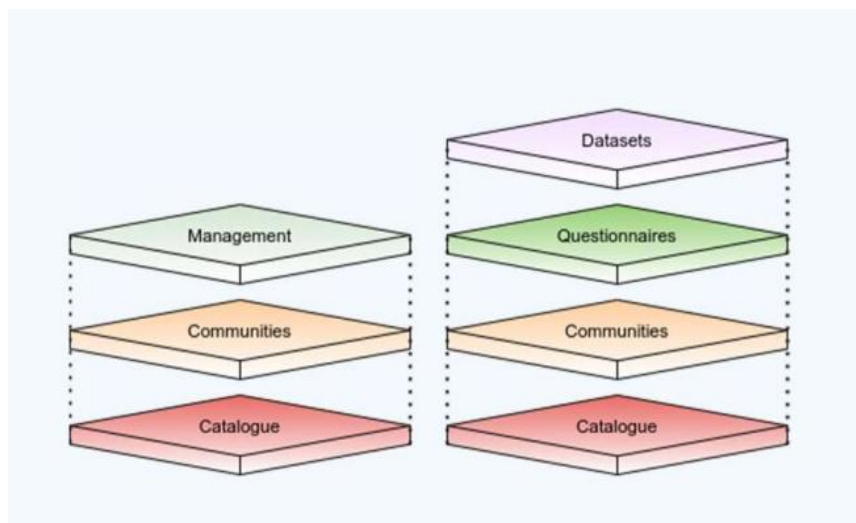


Figure. 1: Overview of Montra’s component structure. The catalogue hosts one or more communities (the data sources), and the communities have one or more questionnaires (data sources), containing datasets that abide the questionnaire’s structure. Additionally, each community can be managed by one or more users by defining proper role-based access rules.

Templates must be defined by maintainers of a Montra instance to define structures for one or more types of cohort datasets. The templates (named “questionnaires”) of the datasets are concretely defined in tabular files (in particular, Excel files) by administrators or maintainers (Figure 2). The files can be imported into Montra, and from there on, cohorts can add dataset entries that follow the template’s rules: allowed fields, order, structure and inter-field dependencies. The dynamic nature of templating datasets also allows maintainers to perform modifications to the structure and apply them without incurring data loss.

B	D	E	F	M
Text/Question	Data type	Value list	Help text/Description	Include in Advanced Search
Database General Information			The database view provides general information about the database, for example, name, institution, data held and main responsibilities.	Yes
Acronym	open-button			Yes
Name	open			Yes
Institution Name	open			Yes
In the last 3 years, the main financial support has been obtained through;	choice-multiple-freeform	Funding by Own Institution{...} National/Regional/Municipal Public Funding{...} European Public Funding{...} Funding From Industry / Contract Research{...} Funding From Public/ Private Partnership{...} Funds From Patients Organisations, Charity and Foundations{...} Other, Please Specify{...}	In the past three years (2010-2012), what has been the source of funding for the data base. For each source of funding, please provide a rough estimate of the percentage of the total funding of the data base that came from that funding source. If you have a source of funding not specified in the list, please describe that source together with the percentage of total funding.	Yes
Database population	comment			Yes
Location	location		Introduce the cities, part of the country, or country from the subjects of cohort	Yes
Population Size (N)	numeric		Introduce the total involved subjects in your cohort	Yes

1.01. Acronym

1.02. Name

1.03. Institution name

1.11. In the last 3 years, the main financial support has been obtained through;

☐ Funding by Own Institution

☒ National/Regional/Municipal Public Funding

☒ European Public Funding

☒ Funding From Industry / Contract Research

☒ Funding From Public/ Private Partnership

☒ Funds From Patients Organisations, Charity and Foundations

☐ Other, Please Specify

1.12. Database population

1.12.01. Location

1.12.02. Population Size

Figure 2: A template, or “questionnaire”, is defined in a tabular file for Montra. Datasets of a community’s questionnaire will have the same fields and the same field structure/order.

By concretely defining role-based access control (RBAC) rules, administrators can operate the visibility and operability of the data amongst users of the catalog. Users can be categorized into different access control groups (two examples are a “default” user group for ability to read datasets, and an “editors” user group for creating/managing datasets). The authorization rules can also be user-centric, allowing a very granular control over the handling of datasets.

Federation + REST API docs

The demonstrator EPND SN service exposes its dataset metadata by means of a REST API, which can also query other federation network instances for metadata regarding their own unique datasets. The datasets themselves expose silver-level metadata, giving some minimal data and more contextualized information regarding datasets.

Regarding the REST API, users must employ programmatic HTTP GET calls to retrieve federated datasets. Users that make use of the REST API must log into the EPND SN website beforehand and retrieve an API token, which is a small hash value that must be supplied along with the API request.

Users in possession of the API token may use the following federation endpoints:

- GET https://epnd.bmd-software.com/api/dcat/list/{community_slug}/{questionnaire_slug}
List datasets – Get a list of all datasets for a given community’s questionnaire. Users must specify both the community “slug” (the unique identifier) and the questionnaire “slug” in the URL itself.
- GET https://epnd.bmd-software.com/api/dcat/{fingerprint_hash}
Get dataset – Retrieve a single dataset. Users must only specify the dataset’s “fingerprint hash” (unique identifier) in the URL itself.

The returned datasets (see Figure 3) are structured in a JSON object, the “list datasets” endpoint will return a JSON array with all datasets contained, while the “get dataset” endpoint only retrieves the singular dataset in JSON format.

```

HTTP 200 OK
Allow: GET, POST, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "creator": null,
  "consent": {
    "Results_individual_subject_2": "#Available for sharing unrestricted"
  },
  "sampleTypes": null,
  "dataUseConditions": {
    "permission_recontact_subjects": null
  },
  "numberOfBiosamples": null,
  "logo": null,
  "anonymity": "pseudonymised",
  "diseases": null,
  "spatialCoverage": "https://www.iso.org/obp/ui/#iso:code:3166:GB",
  "temporalCoverage": "2012-01-01",
  "title": "Alzheimer's Disease Neuroimaging Initiative",
  "languages": null,
  "biobank": null,
  "federatedAccessLevel": "level_?",
  "description": null,
  "keyValuePairs": [

```

Figure 3: An example of a federated dataset being returned via REST API call.

Within the returned federated datasets, silver-level data is fetched for the given fields if possible. If there are fields that are not filled out, these will simply have a “null” value, and end-users must handle these situations accordingly.

The REST API documentation regarding federation, among other categories, can be found in the following URL (and see Figure 4):

- <https://epnd.bmd-software.com/api/doc>
User must be registered and logged in, in order to view this page.

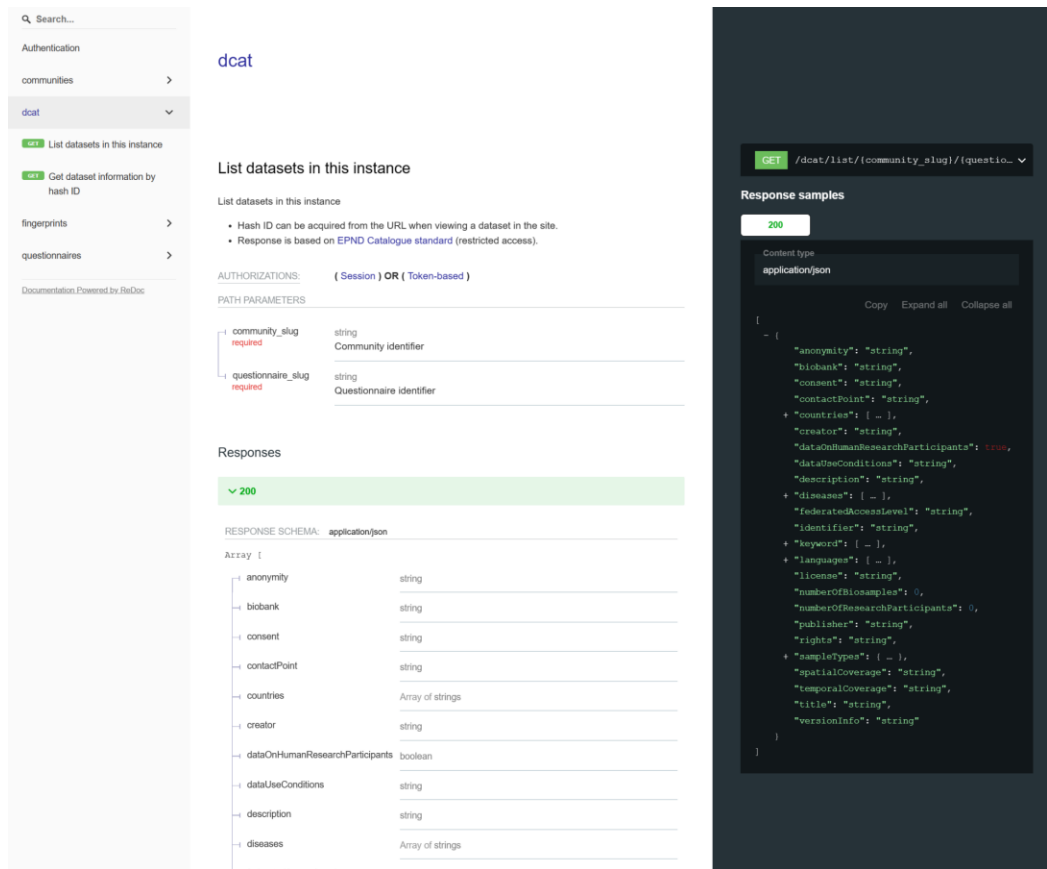
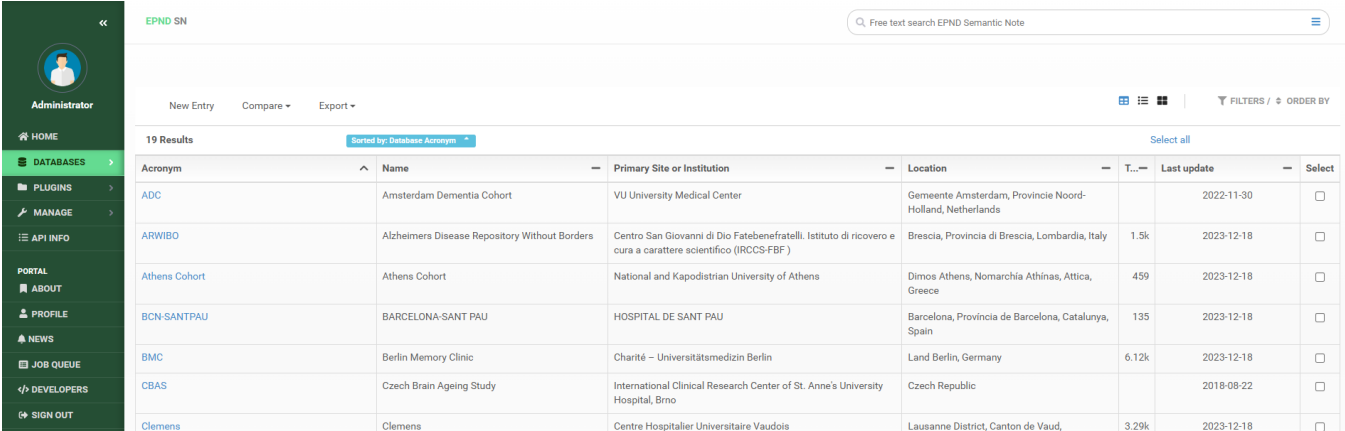


Fig. 4: EPND SN API documentation page.

System Presentation

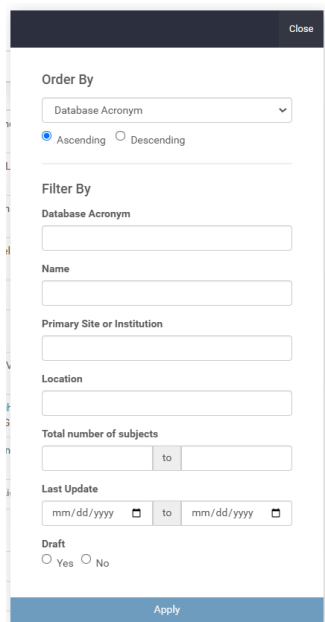
In terms of presentation, the online EPND SN catalogue has a similar layout and structure to EMIF-AD. The users must firstly register into EPND SN in order to use it. After registration, users are free to log in to access the data, and by default will only have the default RBAC settings (can only read datasets).

Upon logging in, users have access to a page containing the datasets in a table view (Figure 5). The table presents some of the datasets' fields, and users are free to filter and order the data by utilizing the "Filters / Order By" button in the top right corner (Figure 6).



Acronym	Name	Primary Site or Institution	Location	T...	Last update	Select
ADC	Amsterdam Dementia Cohort	VU University Medical Center	Gemeente Amsterdam, Provincie Noord-Holland, Netherlands		2022-11-30	☐
ARWIBO	Alzheimers Disease Repository Without Borders	Centro San Giovanni di Dio Fatebenefratelli, Istituto di ricovero e cura a carattere scientifico (IRCCS-FBF)	Brescia, Provincia di Brescia, Lombardia, Italy	1.5k	2023-12-18	☐
Athens Cohort	Athens Cohort	National and Kapodistrian University of Athens	Dimos Athens, Nomarchia Athinas, Attica, Greece	459	2023-12-18	☐
BCN-SANTPAU	BARCELONA-SANT PAU	HOSPITAL DE SANT PAU	Barcelona, Provincia de Barcelona, Catalunya, Spain	135	2023-12-18	☐
BMC	Berlin Memory Clinic	Charité – Universitätsmedizin Berlin	Land Berlin, Germany	6.12k	2023-12-18	☐
CBAS	Czech Brain Ageing Study	International Clinical Research Center of St. Anne's University Hospital, Brno	Czech Republic		2018-08-22	☐
Clemens	Clemens	Centre Hospitalier Universitaire Vaudois	Lausanne District, Canton de Vaud,	3.29k	2023-12-18	☐

Figure 5: EPND SN front page, table view of datasets.



Close

Order By

Database Acronym

☒ Ascending ☐ Descending

Filter By

Database Acronym

Name

Primary Site or Institution

Location

Total number of subjects

to

Last Update

mm/dd/yyyy to mm/dd/yyyy

Draft

☐ Yes ☐ No

Apply

Figure 6: Filter/Sort form for datasets.

Additionally, the dataset table view can be swapped with either a list view or a grid view (see Figures 7 and 8). These modes are more user-friendly, and the datasets, should the questionnaire allow it, can have a dataset image be displayed if present.

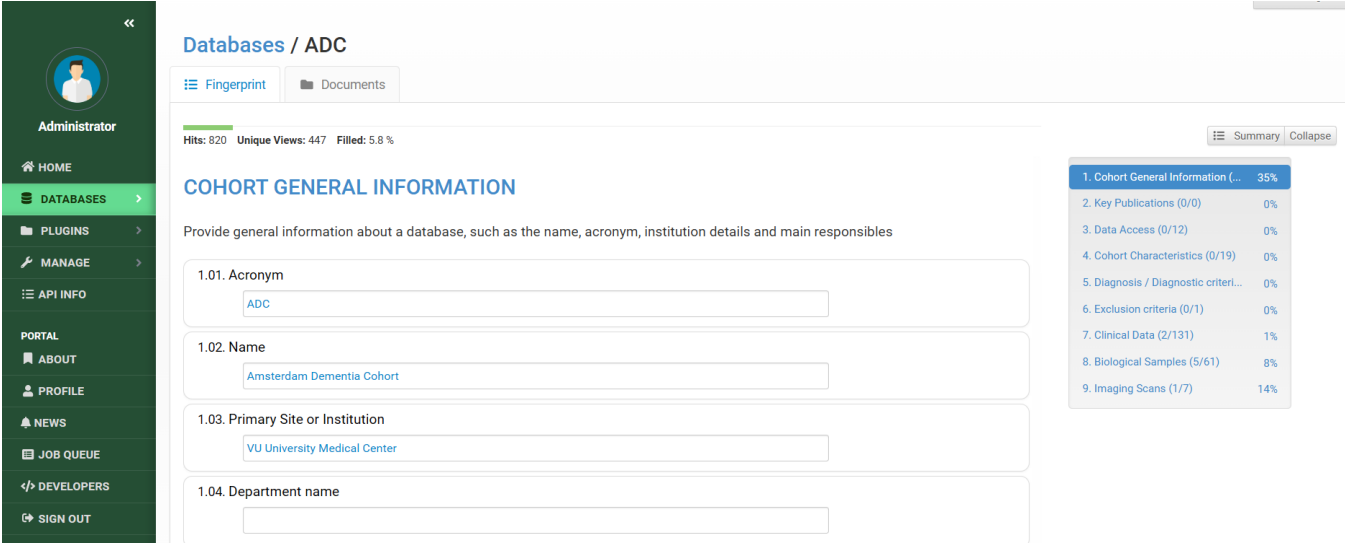
19 Results		Sorted by: Database Acronym
	ADC	Amsterdam Dementia Cohort Primary Site or Institution: VU University Medical Center Location: Gemeente Amsterdam, Provincie Noord-Holland, Netherlands
	ARWIBO	Alzheimers Disease Repository Without Borders Primary Site or Institution: Centro San Giovanni di Dio Fatebenefratelli. Istituto di ricovero e cura a carattere scientifico (IRCCS-FBF) Location: Brescia, Provincia di Brescia, Lombardia, Italy

Figure 7: List view for datasets.

19 Results		Sorted by: Database Acronym
	ADC	Amsterdam Dementia Cohort VU University Medical Center Gemeente Amsterdam, Provincie Noord-Holland, Netherlands
	ARWIBO	Alzheimers Disease Repository Without Borders Centro San Giovanni di Dio Fatebenefratelli. Istituto di ricovero e cura a carattere scientifico (IRCCS-FBF) Brescia, Provincia di Brescia, Lombardia, Italy
		Nat Dimo

Figure 8: Grid view for datasets.

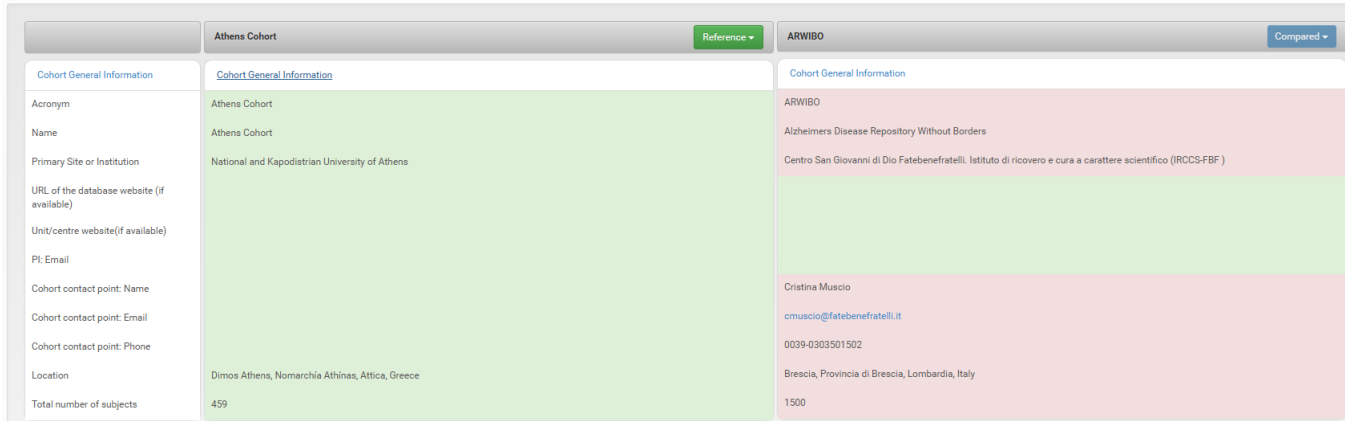
By selecting a dataset, the user may view its structured data. The fields are organized based on the questionnaire structure: the first section is displayed, and only one can be viewed at a time. The current section can be as depicted in Figures 9 and 10.



Summary	
1. Cohort General Information (...)	35%
2. Key Publications (0/0)	0%
3. Data Access (0/12)	0%
4. Cohort Characteristics (0/19)	0%
5. Diagnosis / Diagnostic criteri...	0%
6. Exclusion criteria (0/1)	0%
7. Clinical Data (2/131)	1%
8. Biological Samples (5/61)	8%
9. Imaging Scans (1/7)	14%

Figure 9: Dataset metadata.

In the dataset list view, users can also select two or more datasets and compare their values. Selecting the datasets and pressing “Compare” → “Datasets” leads to a page where the datasets can be viewed side-by-side, with any differences being highlighted.



	Athena Cohort	Reference	ARWIBO	Compared
Cohort General Information	Cohort General Information			
Acronym	Athena Cohort		ARWIBO	
Name	Athena Cohort		Alzheimers Disease Repository Without Borders	
Primary Site or Institution	National and Kapodistrian University of Athens		Centro San Giovanni di Dio Fatebenefratelli. Istituto di ricovero e cura a carattere scientifico (IRCCS-FBF)	
URL of the database website (if available)				
Unit/centre website(if available)				
PI: Email				
Cohort contact point: Name			Cristina Muscio	
Cohort contact point: Email			cmuscio@fatebenefratelli.it	
Cohort contact point: Phone			0039-0303501502	
Location	Dimos Athens, Nomarchia Athinas, Attica, Greece		Brescia, Provincia di Brescia, Lombardia, Italy	
Total number of subjects	459		1500	

Figure 10: Dataset comparison page.

Users with the adequate authorization can access the “API Info” page (see Figure 11), where they can see their API token. The API token can be used to programmatically access the datasets in their federated form.

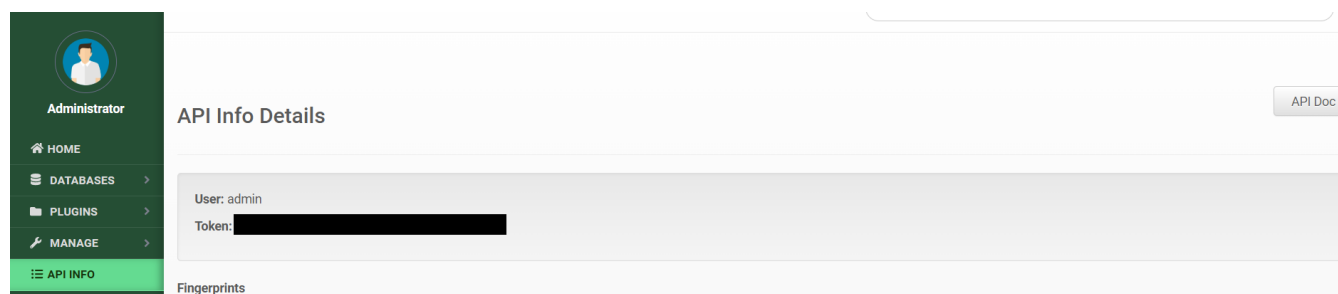


Figure 11: API details page, containing the API token needed for the federation endpoints.

2) EPND Catalogue

The [EPND Catalogue](#) (a component of the EPND Hub which includes all the software services), provides the central portal whereby EPND users can find, request and access cohort, data and samples. Its development is ongoing, with input from WP4 and WP2, and in co-ordination with WP6 for managing communications regarding the releases. To date it has been released in phases, as follows:

1. Bronze Catalogue – tabular view of minimal study information about EPND cohorts and Contact Cohort option with associated terms and conditions.
2. User Management – user registration and login with Multi factor authentication
3. Bronze Admin tools – self-service administration capability for onboarding Bronze cohorts with the role of catalogue administrators to review submissions for approval for public sharing.
4. Silver metadata model – developed in collaboration between WP1 and WP4. Please see the general modelling principles above, and Figure 12 below. This introduced the notion of datasets and biosample collections, access and use conditions to the studies/cohorts.
5. Silver Catalogue – This list view of the catalogue provides improved filtering, search capabilities and additional metadata based on the Silver metadata model
6. Silver Admin tools – self-service administration for cohort owners to create, edit studies, datasets and biosample collections
7. Single Sign-On (SSO) between the Hub and the AD Workbench – This allows EPND users to request access to AD Workbench hosted datasets from EPND.
8. DUC/CCE – Hub-based support for the Digital Use Conditions and Common Conditions of use Elements (DUC/CCE) framework [5, 6], which was initially developed by an international consortium (led by ULEIC) and provides external software and training (<https://duc.le.ac.uk/>)

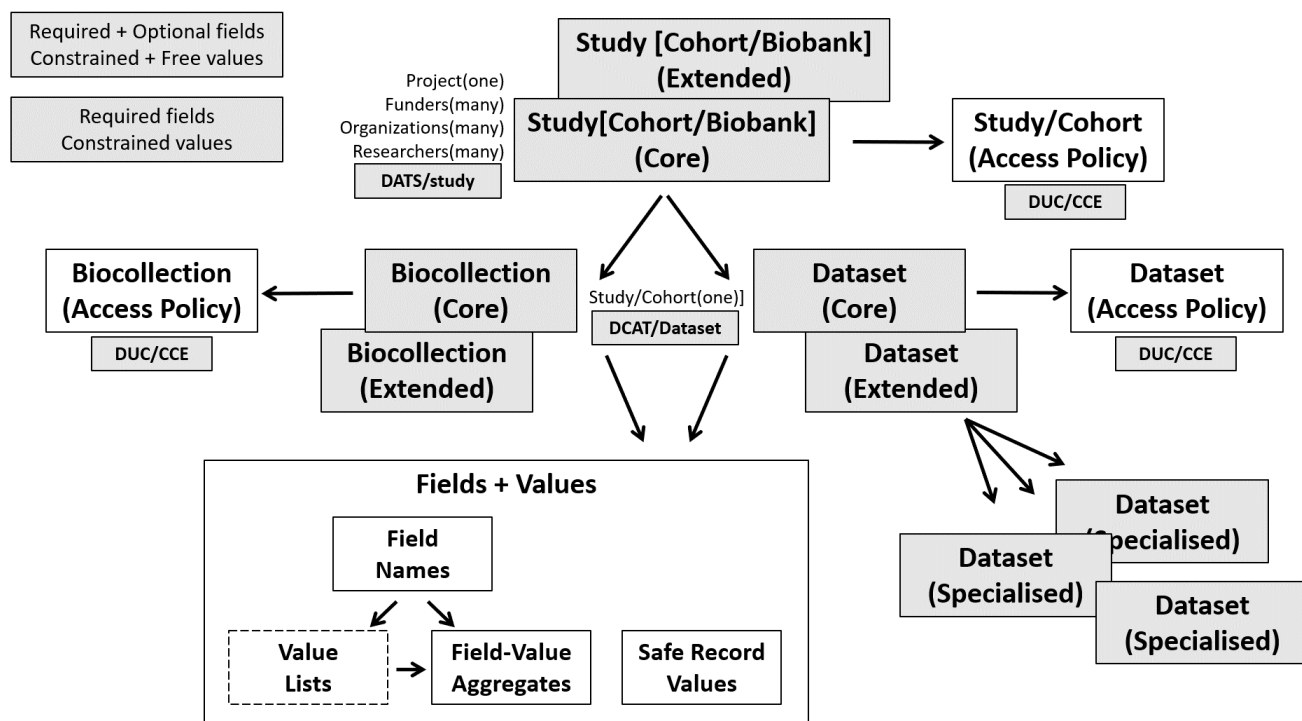


Figure 12: Silver Metadata Model

Figures 13 - 20 below provide explanatory screenshots of the EPND Catalogue and admin tools.

epnd European Platform
for Neurodegenerative
Diseases

Register

Register as an EPND Catalogue user.

Email *

vijay.sureshkumar@outlook.com

Password *

Confirm password *

First name *

Last name *

Area of expertise

Job title

Organization name *

Organization city

Organization country *

Organization URL

How did you hear about the EPND? *

☐ Social media (Twitter, LinkedIn)

☐ Web Search

☐ Colleague/Professor

☐ Conference

☐ News media

☐ Email/newsletter

☐ Other

With which sector are you affiliated? *

☐ Academia

☐ Industry

Figure 13: Register with EPND

 European Platform
for Neurodegenerative
Diseases

Sign in

Sign in here to access the Study Administration Dashboard.

Email

Password

☐ Remember me [Forgot Password?](#)

[Sign In](#)

New user? [Register as an EPND Catalogue user!](#)

Figure 14: Login to EPND

×

I would like to contact cohort: Aiginition Longitudinal Biomarker Investigation Of Neurodegeneration

☐ I acknowledge EPND will use the personal data provided hereunder in accordance with the [Privacy Statement](#).

☐ I acknowledge EPND may disclose personal data (e.g. names, affiliations, email addresses) to interested third parties including EPND partners, and may display this data in the EPND platform for the purposes of facilitating scientific research and collaboration.

[Contact](#)

Please, check Acknowledge privacy and Acknowledge disclose personal data terms.

Figure 15: Terms and Conditions


**European Platform
for Neurodegenerative
Diseases**

[About](#)
[Platform](#)
[News & Resources](#)
[My Dashboard](#)

Filter
Reset Filter

Use spaces to combine keywords for more precise results. Example: 'alzheimer France'. Use the filter to further specify your search.

Participant Disease Area(s)

- ☐ Cognitively normal (control group)
- ☒ Alzheimer's disease (AD)
- ☒ Parkinson's disease (PD)
- ☐ Dementia with Lewy bodies (DLB)
- ☐ Isolated REM sleep behavior disorder (IRBD)
- ☐ Cerebral amyloid angiopathy (CAA)
- ☐ Frontotemporal dementia (FTD)
- ☐ Amyotrophic lateral sclerosis (ALS)
- ☐ Progressive supranuclear palsy (PSP)
- ☐ Corticobasal degeneration (CBD)
- ☐ Multiple system atrophy (MSA)
- ☐ Huntington's disease (HD)
- ☐ Ataxia

Biosample Collection Type

- ☒ CSF
- ☐ Serum
- ☐ Plasma
- ☐ DNA
- ☐ Saliva
- ☐ Faeces
- ☐ Urine

Imaging Data Type

- ☒ MRI
- ☐ PET-Amyloid
- ☐ PET-Tau

The EPND Catalogue

Welcome to the next iteration of the EPND Catalogue. EPND aims to accelerate neurodegeneration research by providing a growing catalogue of international studies across the neurodegenerative disease spectrum. Studies include metadata on participants, biosample collections, imaging and cognitive data. Search and filter through the catalogue to find and connect with studies relevant to your research.

Showing 8 of 72 Studies (35697 of 237389 Participants*)

*Selecting multiple filters within a single category will display studies with all of the selected characteristics (e.g. Serum AND CSF). Selecting single filters across categories will restrict studies to those with all of the selected characteristics (e.g. AD AND Serum AND MRI). The number of filtered participants represents the disease areas only. The actual number of biosamples or images might be lower depending on the protocols. Please contact the studies for precise counts.

Sort by study name

Ongoing Single-site study since 2008

Biobank Alzheimer Center Limburg

 Maastricht University Medical Center (MUMC)
  Frans Verhey
  Netherlands
  EPND

Memory Outpatient Center of the MUMC for diagnosis of memory problems. Blood and/or cerebrospinal fluid are stored in a biobank. The biobank is a collection of stored body material, which is linked to relevant data, such as medical data, cognitive test performance or the brain scan. Patients participating in the BB-ACL study are called for repeated examination after diagnosis in order to map the course over time.

709 Patients

Ongoing Multi-site study since 2008

Danish Dementia Biobank

 Danish Dementia Research Centre
  Gunhild Waldemar
  Denmark
  EPND

Biofluid samples from patients who have undergone diagnostic assessment at the Memory Clinic, Rigshospitalet.

2140 Patients

Ongoing Single-site study since 2000

GEDOC - A database and biobank for clinical dementia research

 Karolinska Institutet, Stockholm
  Miia Kivipelto
  Sweden
  EPND

Observational memory clinic cohort including data and biosamples from patients remitted to the two Memory Clinics in Karolinska University Hospital in Stockholm. The cohort is suitable for biomarker research and clinical research about dementia. CSF, serum, plasma and for a subset also whole blood are banked. Database includes eg clinical measures, memory tests, ApoE, MRI.

Figure 16: EPND Catalogue

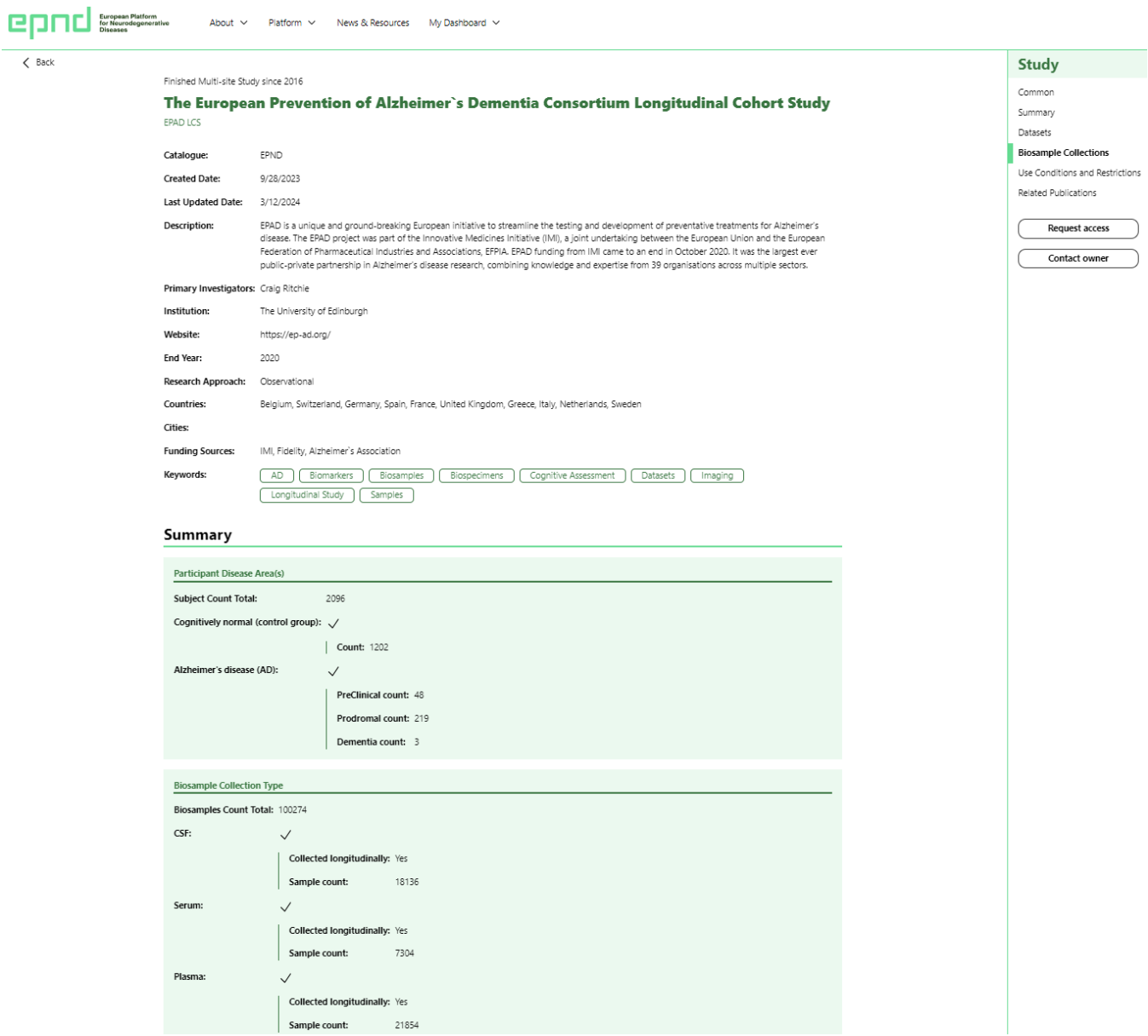


Figure 17: Study Details page

Dataset

EPAD LCS v.IMI

Study: The European Prevention of Alzheimer's Dementia Consortium Longitudinal Cohort Study

Catalogue: EPND

Created Date: 12/11/2023

Last Updated Date: 2/26/2024

Description: The latest and final EPAD dataset went into open access to all researchers from over the world in November 2020. This final dataset is called Version IMI (V.IMI) as it represents all the data collected and processed during the IMI period of EPAD. This contains the final longitudinal data including cognitive, clinical, biomarker and neuroimaging data sets from over 2,000 participants of the EPAD Longitudinal Cohort Study (LCS). Screening for the first participant in the LCS occurred in May 2016 and finished in early 2020. All 2,096 participants who consented and were included into the dataset are entered in this final V.IMI release.

Primary Investigators: Craig Ritchie

Countries: Spain, France, United Kingdom, Greece, Ireland, Sweden

Funding Sources: IMI, Fidelity, Alzheimer's Association, Currently unsupported

Keywords: AD Biomarkers Biosamples Biospecimens Cognitive Assessment Datasets Imaging Longitudinal Study Samples

Summary

Coverage

Minimum Age: 50

Maximum Age: 90

Birth Sex Coverage: Female, Male

Ethnicity Coverage

Asian: ✓

Black or African American: ✓

White: ✓

Other, including mixed group/background: ✓

Dataset Information

Dataset Version: v.IMI

Number of Subjects: 2096

Diseases: AD

Demographics

Education: ✓

Clinical Information

Comorbidities: ✓

Medication use: ✓

Family history of neurodegenerative disorders: ✓

Dataset

Common

Summary

Use Conditions and Restrictions

Related Publications

Change History

Contact owner

Figure 18: Dataset Details Page


 European Platform
for Neurodegenerative
Diseases

[About](#)
[Platform](#)
[News & Resources](#)
[Cohort Administration](#)

Study Administration

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

Add Study
 Contact Admin

Pending Studies

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua.

Study Name	Proposed changes on	Last updated	Submitted by	Status	
Automated platform for the capture, storage and automated management of data collected during daily clinical activity (CLEMENS)	Sun, 15 Sep 2019 22:00:00 GMT	Sun, 15 Sep 2019 22:00:00 GMT	User Name	Pending	Manage
Automated platform for the capture, storage and automated management of data collected during daily clinical activity (CLEMENS)	Sun, 15 Sep 2019 22:00:00 GMT	Sun, 15 Sep 2019 22:00:00 GMT	User Name	Pending	Manage
Automated platform for the capture, storage and automated management of data collected during daily clinical activity (CLEMENS)	Sun, 15 Sep 2019 22:00:00 GMT	Sun, 15 Sep 2019 22:00:00 GMT	User Name	Pending	Manage

Approved Studies

Figure 19: Study Administration Page

 **European Platform for Neurodegenerative Diseases**

About ▼Platform ▼News & ResourcesMy Dashboard ▼

Back

Edit Study

Please complete the form below and submit for upload to the Study Catalogue.

Common

Name *

GEDOC - A database and biobank for clinical dementia research

Acronym

GEDOC

Description *

Observational memory clinic cohort including data and biosamples from patients remitted to the two Memory Clinics in Karolinska University Hospital in Stockholm. The cohort is suitable for biomarker research and clinical research about dementia. CSF, serum, plasma

Related Documents

No Files added yet.

Add File

Request Access URL

Primary Investigators *

Miia Kivipelto, miia.kivipelto@ki.se

Add contact

Secondary Investigators

Anna Matton/Johan Lundberg, gedoc.karolinska@regionstockholm.se

Add contact

Data Managers

gedoc.karolinska, gedoc.karolinska@regionstockholm.se

Edit Study

Common

Summary

Datasets

Biosample Collections

Use Conditions and Restrictions

Related Publications

Personal Data Usage Agreement

Save as draft

Publish to catalogue

Figure 20: Self Service Study Editing

Discussion

One of the main goals of EPND is to provide online users with a way to ‘discover’ specific assets (cohorts, datasets, biosamples, etc) of interest to them. This Deliverable relates to the first incarnation of such a capability. In particular, as per the title (“Online demonstration of resource-level discovery service, with at least 30 participating cohorts/biobanks”) this first report on this topic addresses

resource-level (i.e., metadata rather than data driven) approaches, for which we invested a lot of work in devising suitable metadata models. These were generally DCAT based, covered cohort, dataset and biosample objects, and were strategically split into bronze, silver and gold tiers to ease the burden of ‘onboarding’ for cohorts and biobanks.

Iterative testing and refinement of these models validated their design, and enabled a powerful EPND Hub/Catalogue to be build based upon them. Moreover, the models were proven to be acceptable to the community of contributors to the catalogue, in that now over 70 cohorts have been onboarded at the bronze level (vs the 30 scheduled for this Deliverable), with around 1/5 of these now already transitioning onto the Silver level.

The usefulness of the models for federated sharing of metadata was demonstrated by extension of the BMD Software to create a functioning EPND Silver Node (EPND SN). Furthermore, this node undertook the stretch goal of implementing a basic / metadata-level Beacon-V2 service for federated querying. A more complete design and implementation of a Beacon-based federated query network is now underway, and this will provide the basis of a follow-on Deliverable which extends the Catalogue discovery capabilities to the Record-level.

Conclusion

The provision of an effective asset discovery capability is a cornerstone requirement for EPND. This report is the first of several Deliverables and Milestones that relate to this objective, and the progress herein described shows that excellent progress is being made. It is not unreasonable to say that the central catalogue already provides a discovery capability that compares favourably with any other in this domain, and the metadata model it is based on provide a way forward for yet many further improvements. These improvements include the launch of a (Beacon-V2 based) federated discovery network, and the use of not only metadata but also data level elements. By progressing stepwise, the project will determine how elaborate these discovery systems should be, to optimally meet the goal of enabling users to find asset of interest before the next step where they reach out to discuss the assets in more detail with, and request access from, the asset controllers.

References

1. Pereira, A., Almeida, J.R., Lopes, R.P. and Oliveira, J.L., 2023. Querying semantic catalogues of biomedical databases. *Journal of Biomedical Informatics*, 137, p.104272.
2. Silva, L.B., Trifan, A. and Oliveira, J.L., 2018. Montra: An agile architecture for data publishing and discovery. *Computer methods and programs in biomedicine*, 160, pp.33-42.
3. Almeida, J.R. and Oliveira, J.L., 2024. MONTRA2: A web platform for profiling distributed databases in the health domain. *Informatics in Medicine Unlocked*, 45, p.101447.
4. Oliveira, J.L., Trifan, A. and Silva, L.A.B., 2019. EMIF Catalogue: a collaborative platform for sharing and reusing biomedical data. *International journal of medical informatics*, 126, pp.35-45.
5. In Press @ DOI: 10.1038/s41597-024-03280-6
6. In Press @ DOI: 10.1038/s41597-024-03279-z

