Data Access Request

AMYPAD PNHS

1. Aims of this document

This document describes the data access request procedure for the Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) Prognostic and Natural History Study (PNHS).

All the researchers interested in using the integrated AMYPAD PNHS dataset should make a formal request to access the imaging, clinical, and biomarker data for scientific research investigation and/or educational activities[1].

In addition, the procedures described in this document relate only to the *raw*, *harmonized* and *derived* data. Access to the *source* data is not within the scope of this document, as accessing this data might require specific Data Transfer Agreements between the researcher and the Parent Cohorts.

2. Data locations

The AMYPAD PNHS is stored at two different locations:

- *Tabulated data* is stored in the Alzheimer's Disease Data Initiative (ADDI) Workbench (ADWB).
- *Neuroimaging data* is stored in the Health-RI XNAT imaging platform (https://xnat.bmia.nl/).

3. Data definitions

The following definitions have been adopted to distinguish between different types of data:

- *Source data:* refers to the data before its harmonization and/or any file format conversion. It represents, for example, the original files shared by the Parent Cohorts and the DICOM PET or MRI images.
- *Raw data:* unprocessed or minimally processed data. In the integrated tabulated dataset, it represents those variables directly copied from the source, such as age, years of education, weight, blood pressure, or the score of the different neuropsychological tests. For the neuroimaging data, it represents the PET and MRI in NIfTI format, right after conversion from DICOM.
- *Harmonized data:* refers to the processed data that has been harmonized across centres (e.g., categories and z-scores).
- *Derived data:* represent those metrics derived from the different neuroimaging processing pipelines. These metrics are stored in the tabulated database and as NIfTI images.

An overview of the available variables and their type can be found in the Appendix – List of Variables.

[1] Prospective data collected within the context of AMYPAD PNHS will be accessible to AMYPAD partners as provided in the AMYPAD Consortium Agreement. No formalities or disclosure of the study needs to be provided by the researcher, beyond assurance that applicable national and international laws on the use of research data are applied.

4. Data Access Request (DAR) Procedure

Those researchers interested in using the integrated AMYPAD PNHS dataset can request access to the imaging, clinical, and biomarker data for scientific research investigation and/or educational activities. The application can be performed via the FAIR Data Service of the Alzheimer's Disease Data Initiative (ADDI). In this platform, the user can perform the request to access the:

a. Harmonized and derived data
b. Raw data (which will include also access to the harmonized and derived information)

Moreover, the user needs to specify if the request is for:

1. Tabulated data only
2. Image data only (i.e. NIfTI files), which includes access to a basic set of demographic variables
3. Tabulated and Image data

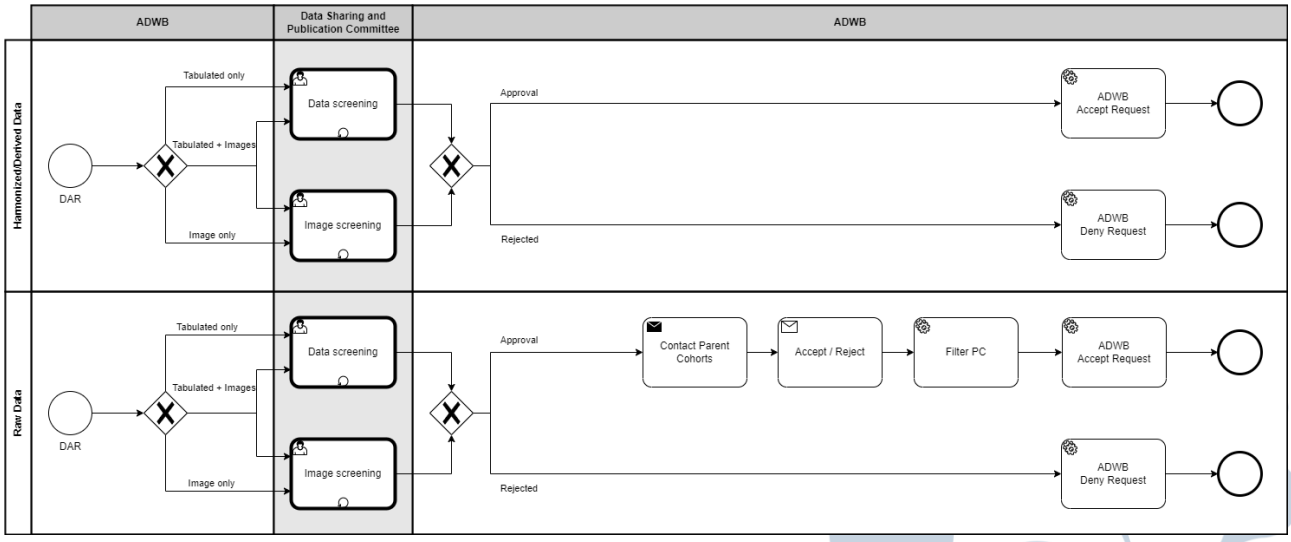
To complete the DAR, the user should fill out a form and provide a one-page proposal describing the study and the expected use of the data (see the section below).

After completion of the request, the AMYPAD *Data Sharing and Publication Committee* (DPC) will review the application and the research proposal. Incomplete applications or those without a clear focus will not receive approval. The results of the DPC review will be sent via the FAIR platform, and the approved application will be processed based on the requested data type:

a. *Harmonized and derived data* does not require further approval by the Parent Cohorts, and access to the data will be granted. This process will take up to one month.
b. *Raw data* requires specific approval by the Parent Cohorts and, therefore, a representative of each cohort will be contacted with a copy of the proposal. Each cohort will decide if they would grant or not access to their data. This process will take up to 2 months (one month for the assessment of the DPC and one month for the Parent Cohort).

Note: Data requests, independently of their type (i.e., raw, harmonized or derivative) that include only a single parent cohort will require specific approval by the Parent Cohort.

The results of the data access request will be sent to the researcher via the FAIR platform. For approved applications, the user will receive access to retrieve the data in the AD Workbench. In case neuroimaging data is also requested, information to access the XNAT will be also provided.



Data	Type	Location	Complexity	Approval by Parent Cohort
Tabulated data	Source	-	High	Out of the scope of the DAR
	Raw	ADWB	Medium	Yes
	Harmonized	ADWB	Low	No
	Derived	ADWB	Low	No
PET images	Source	-	High	Out of the scope of the DAR
	Raw	XNAT	Medium	Yes
	Derived	XNAT	Low	No
MR Images	Source	-	High	Out of the scope of the DAR
	Raw	XNAT	Medium	Yes
	Derived	XNAT	Low	No

The "complexity" to access the data tries to reflect a combination of potential risks:

a. data closer to the source data from the Parent Cohort,
b. a higher risk for subject identification,
c. stronger technical and theoretical knowledge of the underlying data, assumptions and limitations.

5. Data Use Agreement

Those researchers requesting access to the AMYPAD PNHS should accept the responsible use of the data under the terms described in the *AMYPAD PNHS Data Use Agreement*. Briefly, in this document, the following points are mentioned:

1. Amsterdam UMC – location VUMC remains as the **data controller** of the dataset, while the user becomes a **data processor** and, therefore, will process the data only on *behalf of the AMYPAD Consortium.*
2. There will be no attempt to establish or retrieve the identity of the study participants.
3. The data will not be redistributed with others unless they have independently applied and been granted access.
4. Secondary and derived data will be shared only at the group level.
5. The *AMYPAD PNHS Publication Policy* should be used when publicly presenting any results or algorithms.

Appendix – List of Variables

Domain	Concepts
Demographics	Age, sex, ethnicity, household, education, handedness
Family history of dementia	Father or mother with dementia and its age of onset
Genetics	APOE genotype
Vital signs	Resting heart rate, blood pressure, waist and hip circumference, weight and height, body mass index, waist-to-height ratio
Medical history	Medication for hypertensive disorder or cognitive impairment; history of cerebrovascular accident, diabetes, hypercholesterolemia, myocardial infarction, chronic ischaemic heart disease, chronic obstructive lung disease, depression, rheumatoid arthritis, malignant neoplasm, general anaesthesia, and head injury; Framingham risk score
Lifestyle	Tobacco and alcohol behaviour; physical, cognitive, and social activity; sleep quality; diet
Neuropsychiatric scales	Depression and anxiety
Neuropsychological tests	CDR, MMSE, IADL, memory function (immediate and delayed recall, recognition, and visual), language function (fluency and naming), attention (digit span forward and backwards, coding, TMT A), executive function (TMT B and B/A, letter fluency), visuospatial functioning
Cerebrospinal fluid	P-tau, T-tau, Amyloid beta 1-42 and 1-40
PET	SUVR* and Centiloid*, SUV, Visual read, pharmacokinetic modelling (Reference Logan and SRTM2)
MRI	Volumes, Visual read (GCA, PCA, MTA, Fazekas, White Matter Changes, lacunary infarcts, microhemorrhage, large vessel infarcts)

** These variables were calculated with the following reference regions: 1) the grey matter cerebellum, 2) the whole cerebellum, 3) the pons, and the 4) subcortical white matter*

In general terms, the variables related to clinical measurements and/or questionnaires are included in the AMYPAD PNHS integrated dataset as:

Content type	Level
Name of the scale, test, questionnaire or laboratory kit	Harmonized
Original numeric score or results	Raw
Dichotomic results (e.g., yes/no)	Harmonized
Categorical results (e.g., low/medium/high)	Harmonized
Percentage of Maximum Possible (POMP)	Harmonized
Z-score	Harmonized

A full list of the variables and their assigned level can be found in *AMYPAD PNHS List of Variables*.