EPAD
European Prevention of
Alzheimer's Dementia Consortium

ADataViewer: exploring semantically harmonized Alzheimer's disease cohort datasets

Yasamin Salimi, Daniel Domingo-Fernández, Carlos Bobis-Álvarez, Martin Hofmann-Apitius & Colin Birkenbihl for the Alzheimer's Disease Neuroimaging Initiative, the Japanese Alzheimer's Disease Neuroimaging Initiative, for the Aging Brain: Vasculature, Ischemia, and Behavior Study, the Alzheimer's Disease Repository Without Borders Investigators, for the European Prevention of Alzheimer's Disease (EPAD) Consortium

Background: Currently, Alzheimer's disease (AD) cohort datasets are difficult to find and lack across-cohort interoperability, and the actual content of publicly available datasets often only becomes clear to third-party researchers once data access has been granted. These aspects severely hinder the advancement of AD research through emerging data-driven approaches such as machine learning and artificial intelligence and bias current data-driven findings towards the few commonly used, well-explored AD cohorts. To achieve robust and generalizable results, validation across multiple datasets is crucial.

Methods: We accessed and systematically investigated the content of 20 major AD cohort datasets at the data level. Both, a medical professional and a data specialist, manually curated and semantically harmonized the acquired datasets. Finally, we developed a platform that displays vital information about the available datasets.

Results: Here, we present ADataViewer, an interactive platform that facilitates the exploration of 20 cohort datasets with respect to longitudinal follow-up, demographics, ethnoracial diversity, measured modalities, and statistical properties of individual variables. It allows researchers to quickly identify AD cohorts that meet user-specified requirements for discovery and validation studies regarding available variables, sample sizes, and longitudinal follow-up. Additionally, we publish the underlying variable mapping catalog that harmonizes 1196 unique variables across the 20 cohorts and paves the way for interoperable AD datasets.

Conclusions: In conclusion, ADataViewer facilitates fast, robust data-driven research by transparently displaying cohort dataset content and supporting researchers in selecting datasets that are suited for their envisioned study. The platform is available at https://adata.scai.fraunhofer.de/.

Alzheimer's Research & Therapy 14, Article number: 69 (2022)

Published **Online**

May 21, 2022

https://doi.org/10.1186/s13195-022-01009-4