

AD Data Initiative's Product Suite Powering Novel Digital Biomarker Study



The [SpeechDx](#) project, convened by the Alzheimer's Drug Discovery Foundation's Diagnostics Accelerator, is designed to help solve the longstanding problem of diagnosing Alzheimer's disease and related dementias (ADRD) early enough in the process of disease progression to enable effective interventions. The three-year observational longitudinal study with over 2,000 participants across nine global sites aims to create a gold-standard dataset for the development of prognostic biomarkers based on the subtle changes in speech patterns that may present before signs of cognitive decline become noticeable.

SpeechDx leverages the AD Data Initiative's product suite, which consists of data sharing and analysis tools to empower researchers to work with data across multiple platforms. Specifically, SpeechDx uses the AD Curation Studio, which is managed by [Aridhia](#), to harmonize its digital speech and clinical data and make it available to permissioned researchers around the world.

The AD Data Initiative's partnership with SpeechDx has also provided an opportunity to expand the product suite to include digital voice ingestion tools through the [Global Research and Imaging Platform \(GRIP\)](#).

The infrastructure that has been set up to support SpeechDx is detailed in **Figure 1**.

Data Collection

There are two main types of data in the SpeechDx study:

1. Digital voice data, which is gathered through the SpeechDx app pre-installed on Samsung Galaxy Tab Lite tablets at the location(s) of participants' choice. The SpeechDx app contains a battery of open-source speech tasks chosen to elicit more naturalistic speech (e.g., picture description and storytelling). It was built by the Global Research and Imaging Platform (GRIP), a recently launched open-access research ecosystem with modular workflows, a platform for secure collaboration, and customizable, community-driven tools.
2. Clinical data, including cerebrospinal fluid and plasma, MRI and PET scans, and neuropsychological testing data, gathered during periodic visits to clinical trial sites.

The digital voice data is directed to a secure Azure cloud-hosted server where a dashboard has been set up to help clinical managers track participant flow, including session completion and successful data upload. It is then manually QC'ed, spliced to remove personally identifying information, and transcribed, before being uploaded to the AD Curation Studio.

The clinical data is uploaded and harmonized in the AD Curation Studio after being anonymized at the clinical trial sites.

As soon as the data enters the AD Curation Studio, it is encrypted at all times—at rest and in transit.

Data Curation

The digital voice data are received in the AD Curation Studio, cleaned, described, and organized. A defined set of clinical variables that have been gathered from each site are then harmonized prior to matching them with the respective participants' digital voice data to create the aggregated SpeechDx dataset, unlocking a new pathway for early diagnostic and prognostic discovery.

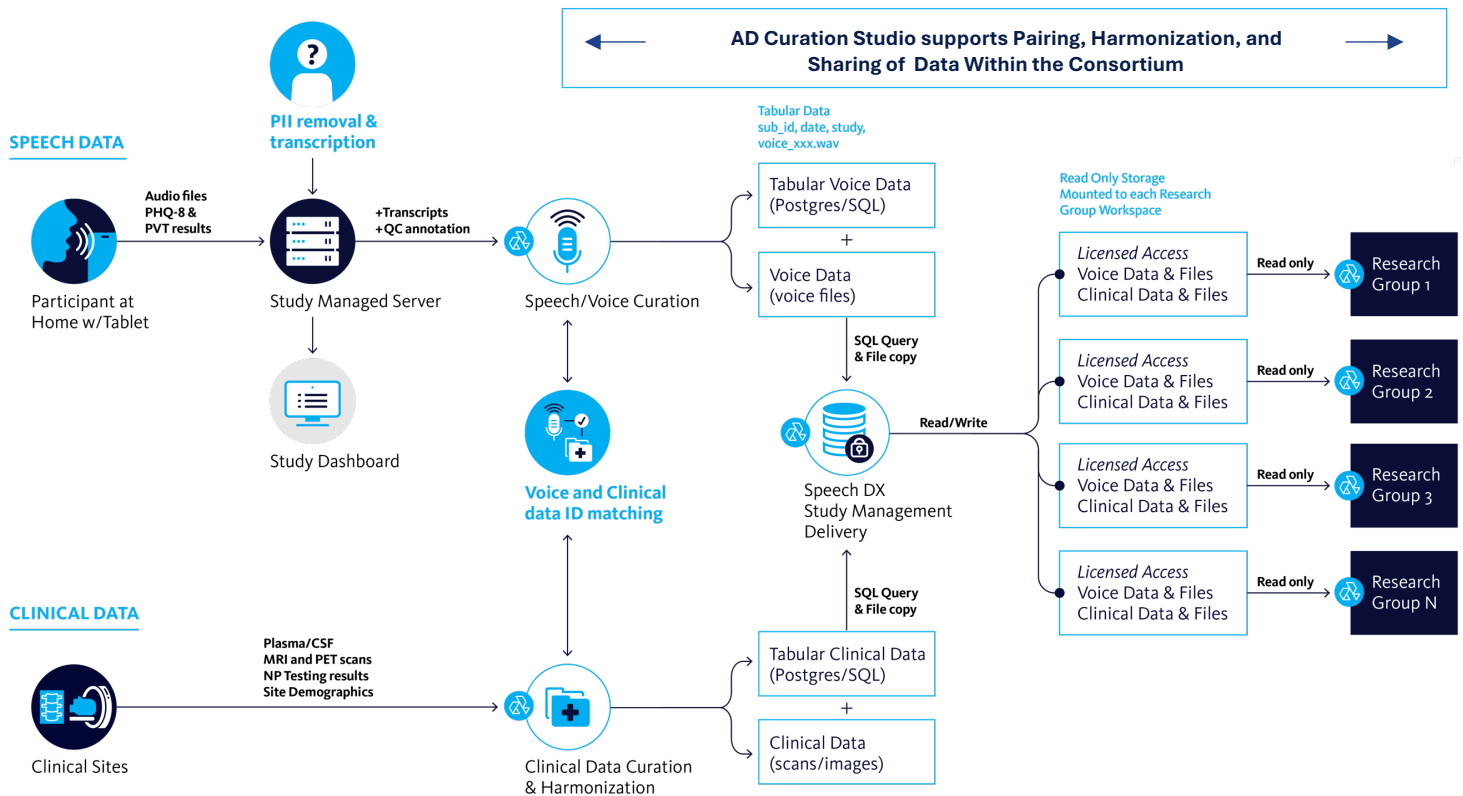
Data Sharing

The analysis-ready data are made available to permissioned research groups that have subscribed access to conduct analyses on the data in secure private workspaces.

- The data is stored as read-only; research groups cannot download the data without permission.
- Research groups can use their own code and models to conduct their analyses and the code, models, and results remain private.

Figure 1.

SPEECH AND CLINICAL DATA ARE PAIRED, HARMONIZED, AND SHARED WITHIN THE CONSORTIUM VIA THE AD DATA INITIATIVE ECOSYSTEM. (All data in AD Curation Studio are encrypted at a rest and in transit.)



If you are interested in discussing whether the AD Data Initiative product suite might help strengthen your research initiatives, please contact us at info@alzheimersdata.org.