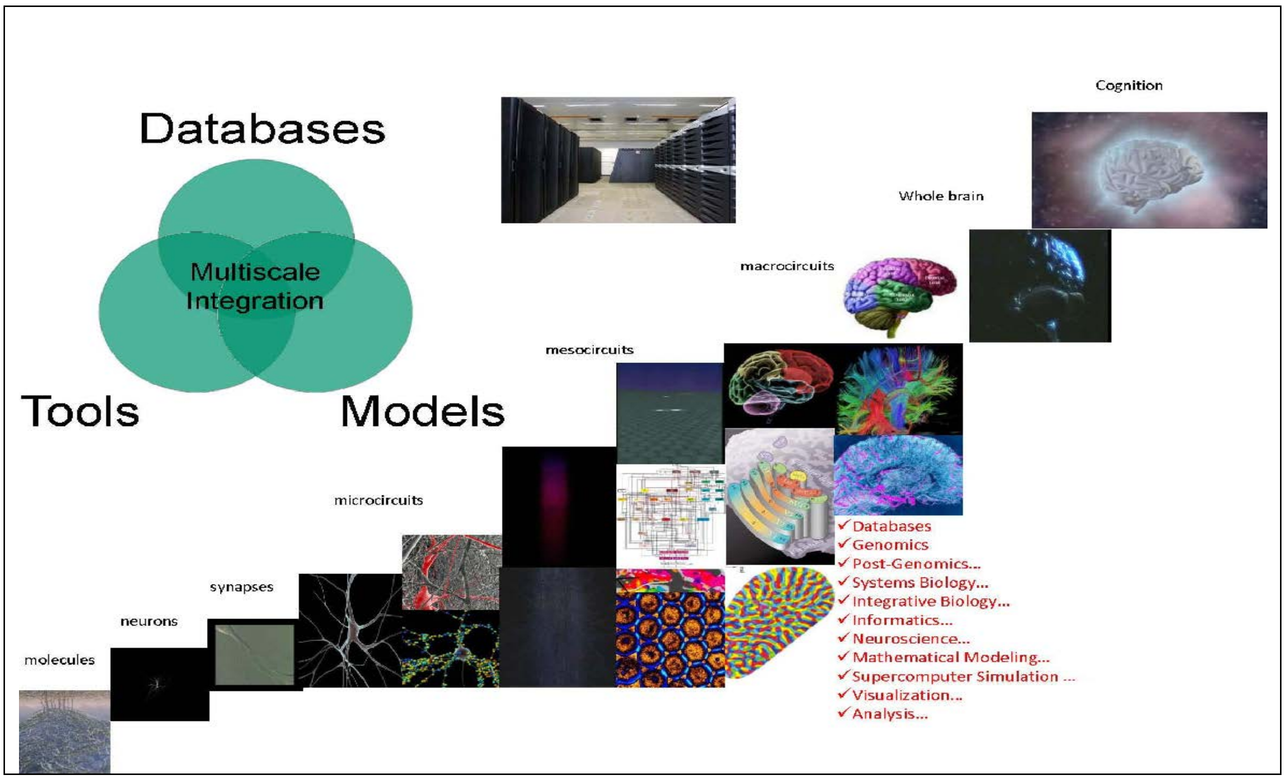
Global Alzheimer's Association Interactive Network

GAAIN
alzheimer's association

Transforming the way researchers approach the study of Alzheimer's disease

Arthur W. Toga

December 2, 2013



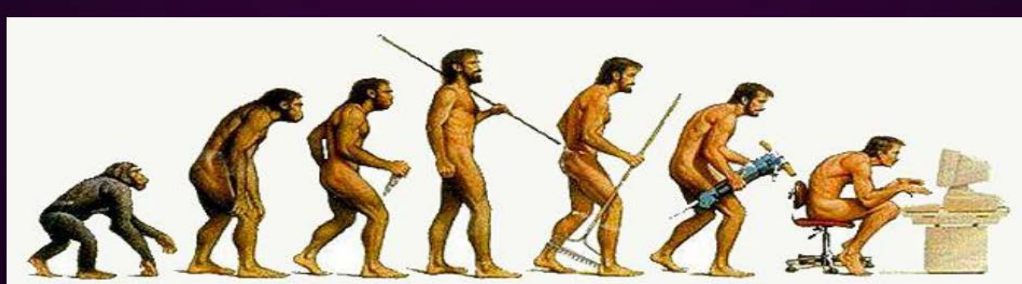
Problem: Storage

- Kryder's Law: Storage medium density is increasing faster than that of integrated circuits predicted by Moore
- Data growth is outpacing storage growth
  - Many researchers do not have sufficient local storage and/or computational resources

Niagara Falls into a teacup

Problem: Bandwidth

- No longer feasible to move ALL the data to the researcher
  - 2009 example of homing pigeon outpacing internet data transfer
    - http://hothardware.com/News/Homing-Pigeon-Faster-Than-Internet-in-Data-Transfer/
    - The time for the pigeon included detaching the memory card and downloading to a computer.



Problem: Data Analysis

- Requires expertise across domains to understand data and know what questions may be asked
- Requires extensive computational resources –processes can take days even with parallel processing systems
- Volume and complexity make it difficult to visualize data
- Difficult to combine data across domains

Neuroimaging Study Size (Typical)

Year	Size	Equivalent to
1998	54MB	20 copies of War and Peace
2005	67MB	24 copies of War and Peace
2012	531MB	193 copies of War and Peace

Image Data Expansion

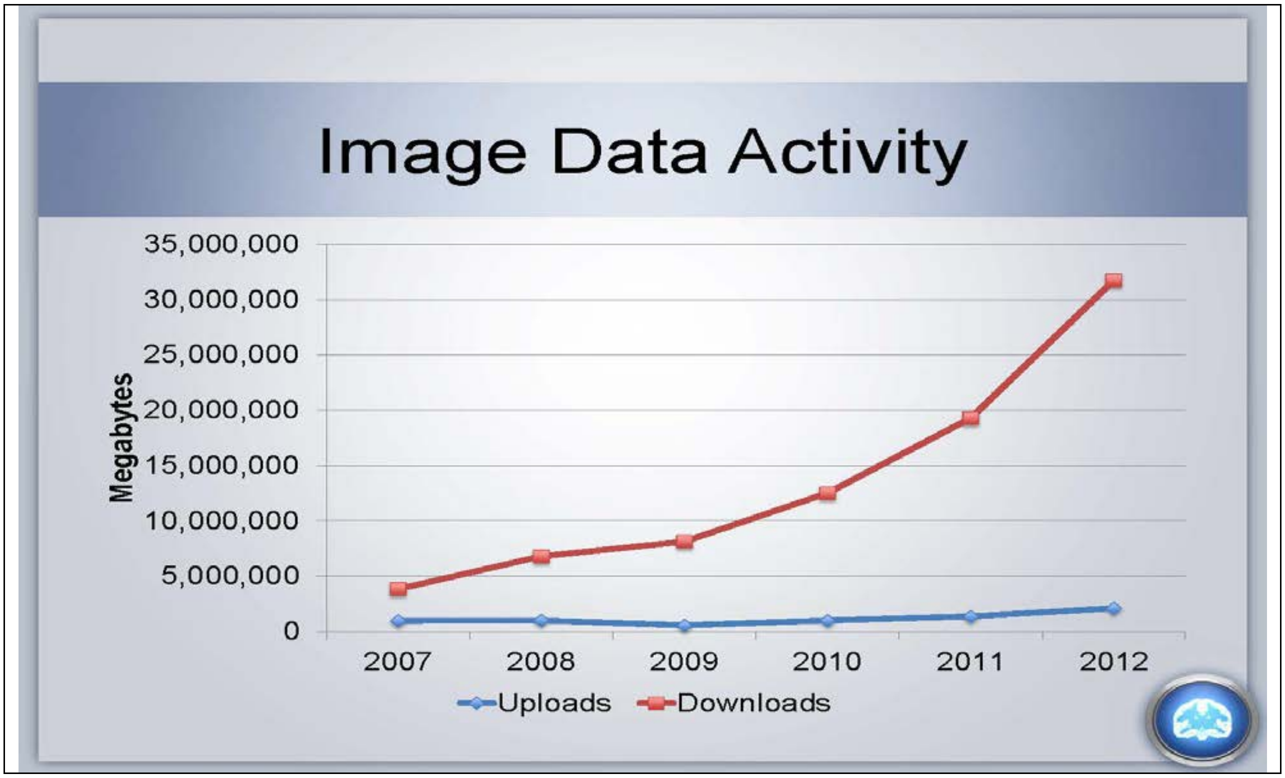
Each neuroimaging scan can spawn many derived image leading to exponential growth

Typical Example:
One 22MB structural scan→
Five preprocessed images (176 MB) →
Eleven postprocessed images (222 MB)

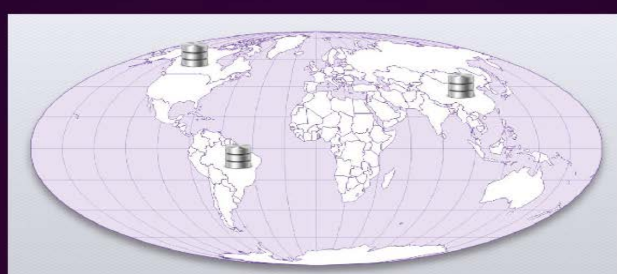
22MB of raw data produces 420MB data for one scan

Genetic Data

- Circa 2010 GWAS Data (per sample)
  - 620,000+ rows of data
  - ~81MB
- 2012: Whole Genome Sequencing (per sample)
  - Standard output from Illumina – multiple files and formats
  - ~250GB per sample
- Example
  - 800 subjects x 250GB = 195TB
  - Time to transfer 195TB:
    - High speed internet (90 Mbit/s): 26 days
    - DSL (45 Mbit/s): 59 days
    - Dial-up (56 kbit/s): 100+ years!



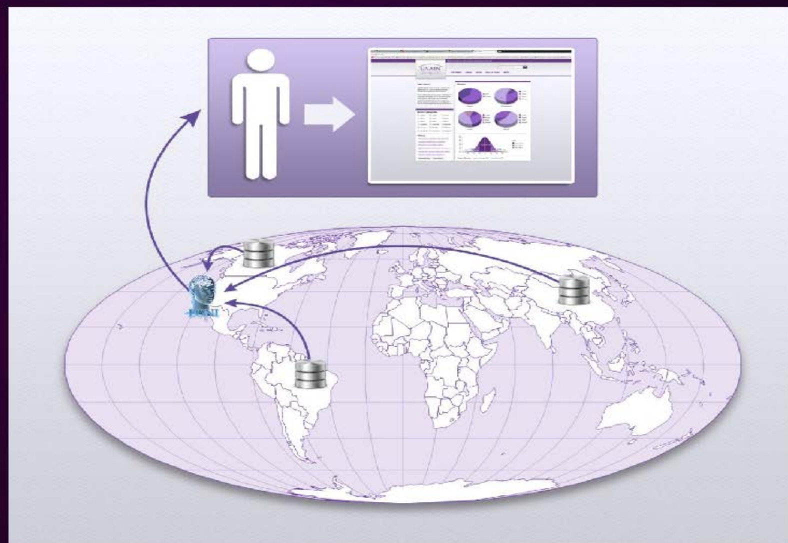
Research efforts in Alzheimer's disease



Research efforts could be vastly expanded in scope and capabilities if data were linked to a global infrastructure that would enable scientists to access and utilize vast, interlinked repositories of data on thousands of subjects at risk for or already suffering from the ravages of Alzheimer's disease.

GAAIN is the first Global Big Data Network for Alzheimer's Disease

Collaborative effort to provide researchers around the globe with access to a vast repository of Alzheimer's disease research data



Supercomputers and High Availability Storage

Data Resources

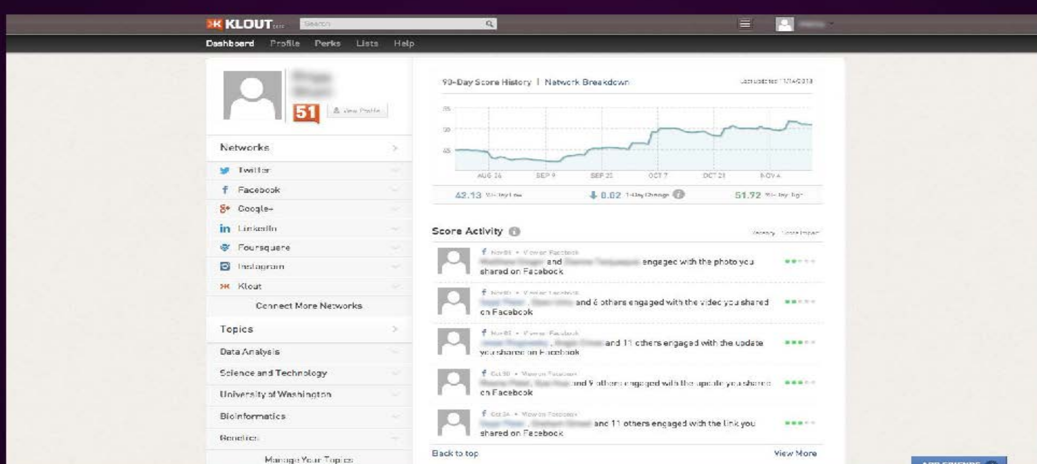
- Storage
  - Fault-tolerant storage area network
  - 400 megabytes per second data throughput
  - Near 24/7 availability
- Protection
  - Daily & weekly on-site backup
  - Monthly off-site backup

- New Data Center
- New 820+ node center:
  - > 7400 total cores
  - > 40 TB memory
  - > 4 PB storage
- PB storage

Aggregating accounts into one hub

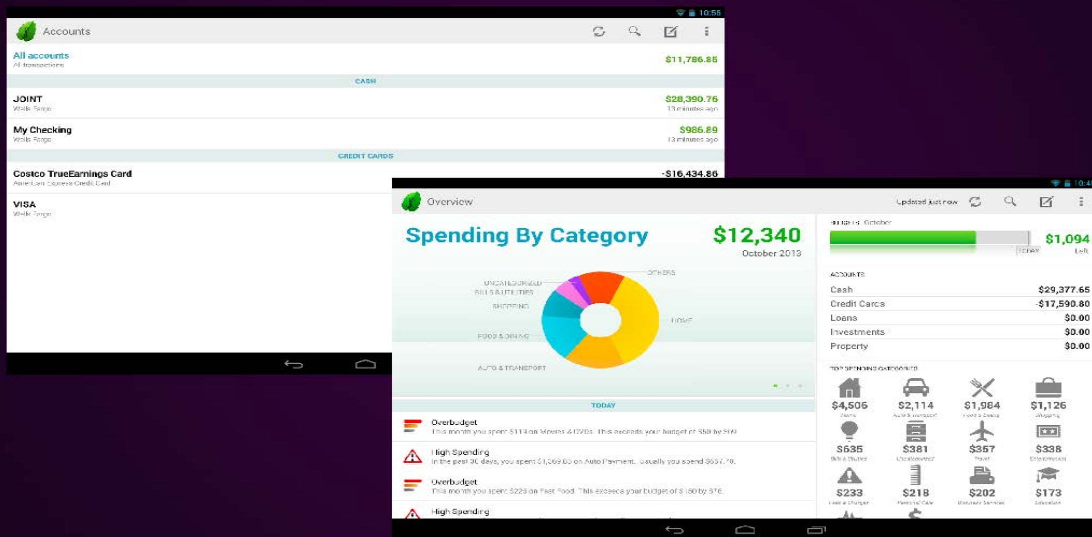
- A single location to obtain data from a variety of sources and accounts
- Users can apply to partnering consortiums via GAAIN after surfing through meta-data
- Users' active accounts with partnering consortiums are also active through the GAAIN portal

Example: Klout.com



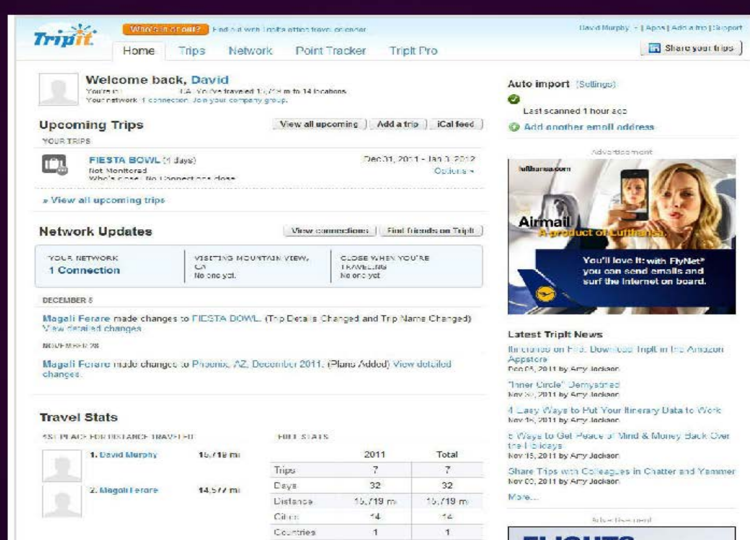
Klout.com collects data of user's presence in social media (ie: Facebook, Twitter, LinkedIn).

Example: Mint.com



Mint.com combines a user's financial information from a variety of sources (ie: bank accounts, credit cards, loans).

Example: Tripit.com

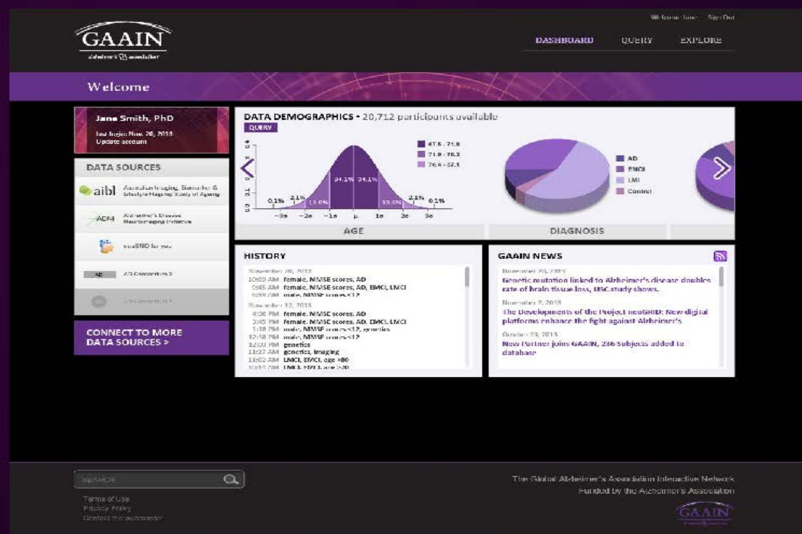


Tripit.com aggregates a user's travel and booking information (ie: airline tickets, vacation rentals).

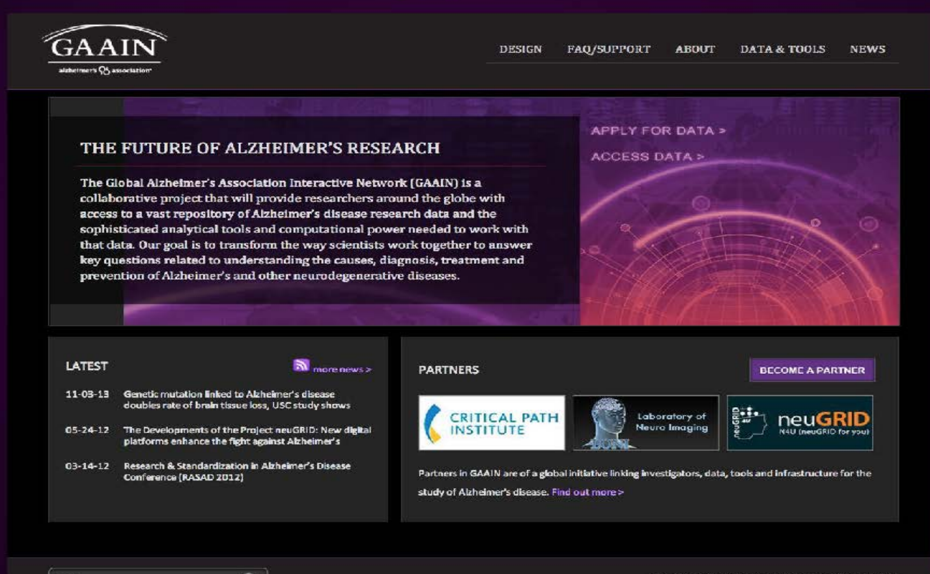
GAAIN Aggregator and Personal Dashboard

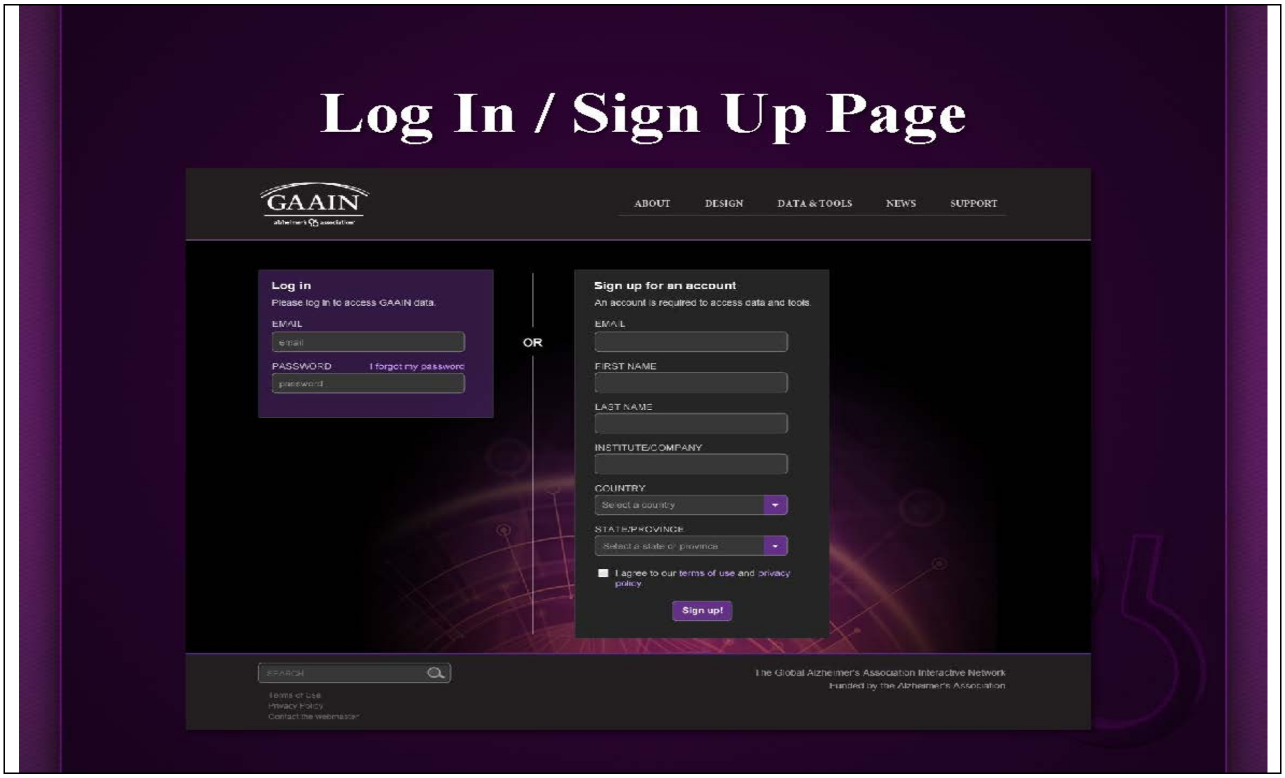
GAAIN recognizes a user's existing accounts for partnering data sources and allows him/her to analyze the data with our tools and/or apply for additional consortiums

The dashboard indicates which data sources are unavailable to the user (ie: the user must apply for access, data source is currently offline)



Gaain.org | Homepage



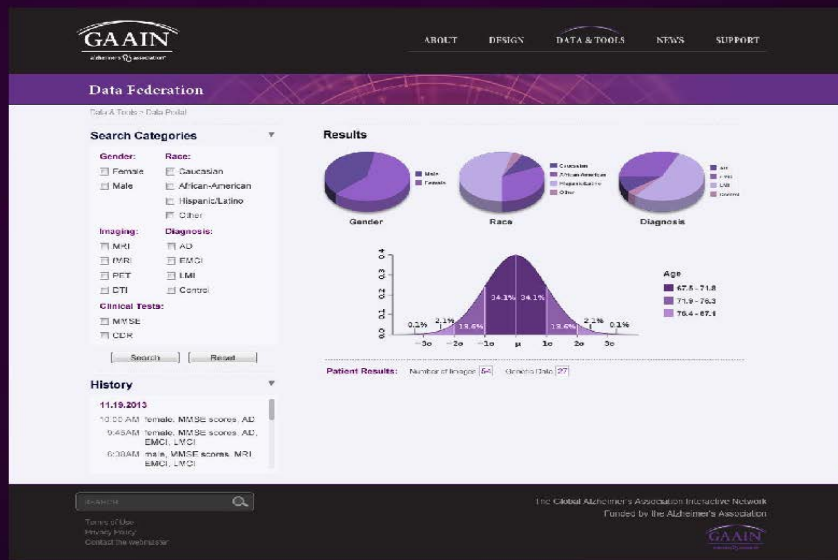


Log In / Sign Up Page
GAAIN
ABOUT
DESIGN
DATA & TOOLS
NEWS
SUPPORT
Log in
Please log in to access GAAIN data.
EMAIL
email
PASSWORD
I forgot my password
password
OR
Sign up for an account
An account is required to access data and tools.
EMAIL
FIRST NAME
LAST NAME
INSTITUTE/COMPANY
COUNTRY
Select a country
STATE/PROVINCE
Select a state or province
I agree to our terms of use and privacy policy.
Sign up!
SEARCH
Terms of Use
Privacy Policy
Contact the webmaster
The Global Alzheimer's Association Interactive Network
Funded by the Alzheimer's Association

News and Updates Page
GAAIN
DESIGN
FAQ/SUPPORT
ABOUT
DATA & TOOLS
NEWS
GAAIN News & Events
November 3rd, 2013
Genetic mutation linked to Alzheimer's disease doubles rate of brain tissue loss, USC study shows
Professor Paul Thompson's group at the Keck School of Medicine of USC have mapped the effects of that genetic mutation, showing for the first time how the Alzheimer's risk factor affects the living human brain. Their work was published in...
May 24th, 2012
The Developments of the Project neuGRID: New digital platforms enhance the fight against Alzheimer's
This article, written by Dr. Giovanni Frisoni, Deputy Scientific Director of IRCCS Fatebenefratelli, Bresica, discusses how the neuGRID, outGRID and GAAIN initiatives are contributing to Alzheimer's research. Sophisticated algorithms, public databases and user-friendly infrastructures are all being used for researchers...
February 7th, 2012
Obama to increase Alzheimer's research funding – USA Today
"The Obama administration is set to announce a plan to spend more than half a billion dollars on Alzheimer's research next year, the Associated Press reports. "The plan is being unveiled as part of President Obama's 'We Can't Wait' initiative,...
January 17th, 2012
Finding a treatment for Alzheimer's by 2025 – CBS News
"The federal government set a goal Tuesday of finding an effective treatment for Alzheimer's by 2025. More than five million Americans were living with the disease in 2010. By 2050, it could be as many as 16 million. CBS News...
November 7th, 2011
Partnering for Cures 2011 – Global Alzheimers Association Interactive Network
A video presentation from the Alzheimer's Association about GAAIN from November 7, 2011.
News Categories
All News & Events
Events
News

One-stop Data Access

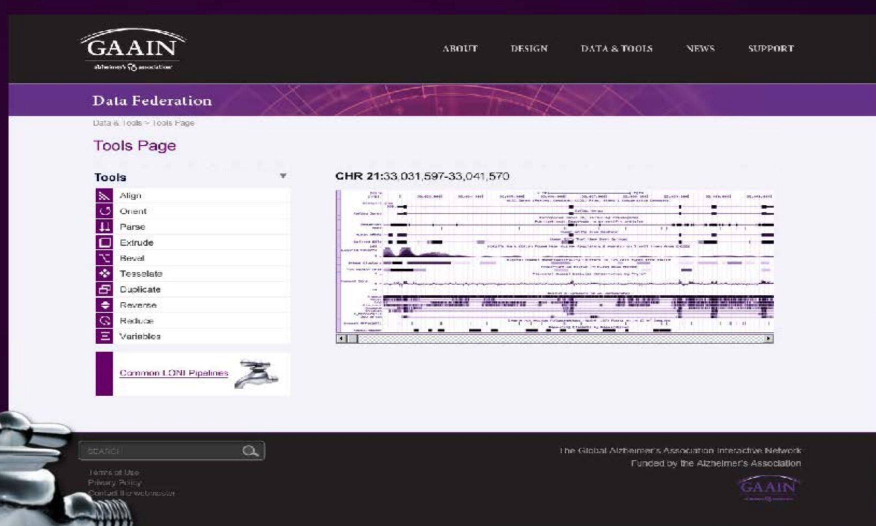
Data from thousands of subjects, including clinical, genetic and imaging data types from our partners



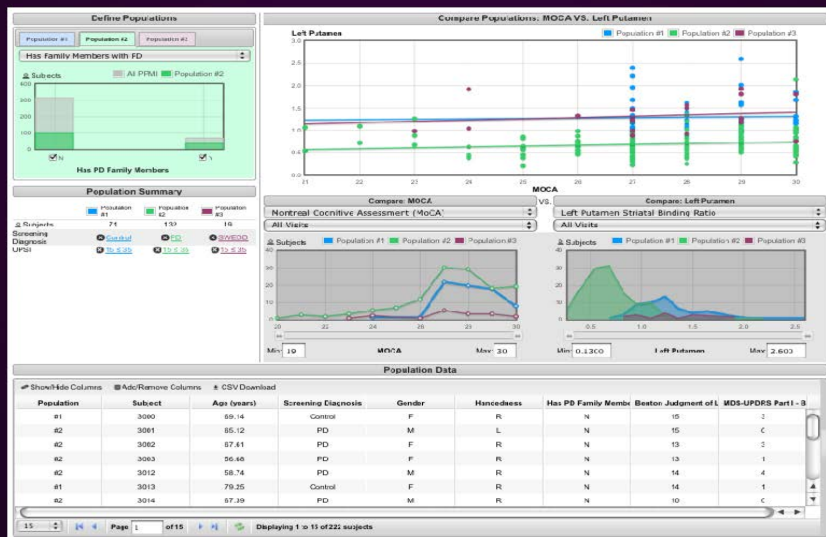
Comprehensive Analytical Tool Stack

Bank of sophisticated imaging and genetic analytical tools available

Tools are supported by the LONI Pipeline



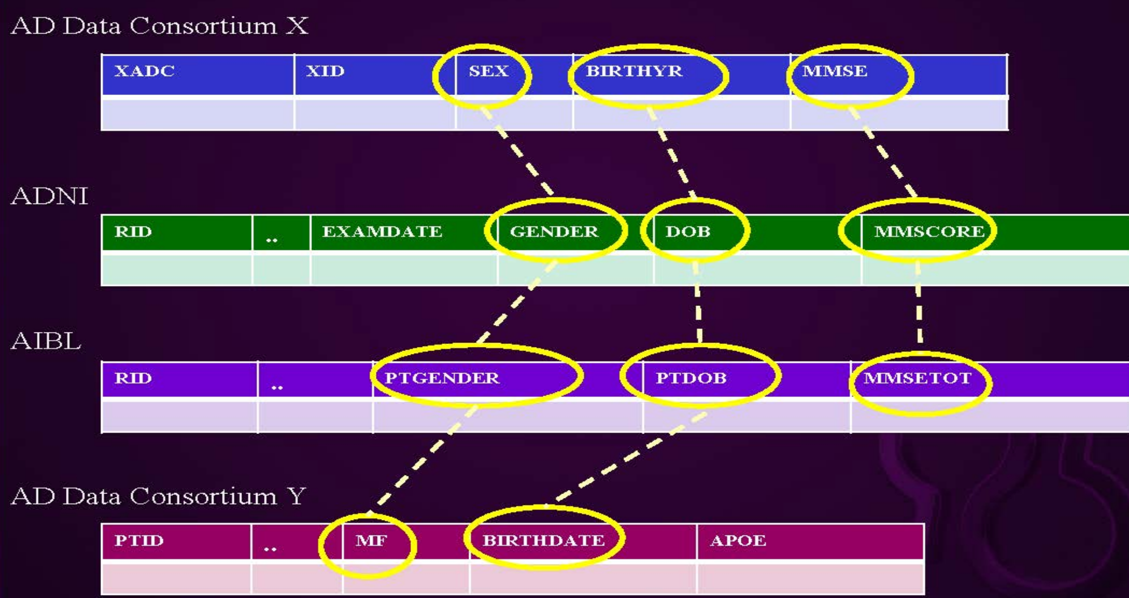
Interactive Filtering/Selecting UI



GAAIN Global Federation Version 1.0

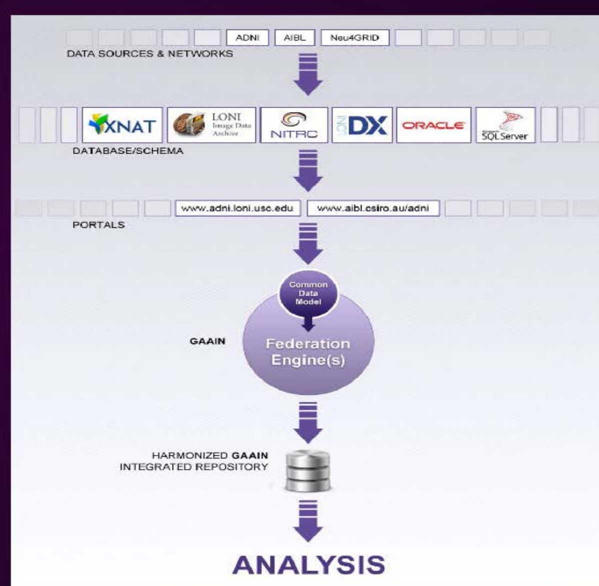
- Provide *federated* integrated access to multiple distributed Alzheimer's disease datasets
  - Stepwise model development
    - Phase I: Similar or identical data models
    - Phase II: Different data models but with same representation
      - Such as (all) relational
    - Phase III: Heterogeneous models
      - Relational versus XML …
  - Integration of data in varying data models
  - *"Syntactic and Semantic Heterogeneity"*
    - Simply put – data sources differ in how they represent the same thing!
- Mediator technology to combine these data
- Common Data Model based on and linked to CDISC

Data Heterogeneity



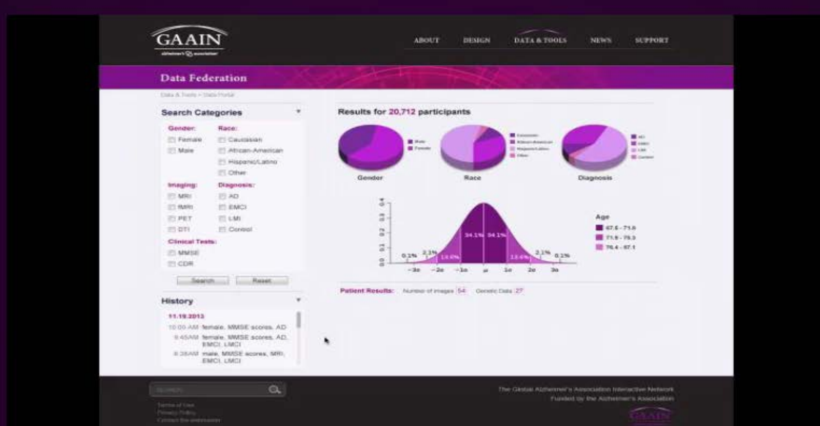
Federated Data Access via Mediator

- Mediation approach
- One-stop data access
- Actual integration of data – not just a clearinghouse
- Maintain autonomy of each source



A big solution for big data

- GAAIN serves as a benchmark for large data research efforts
- Provides seamless connections of a users' existing Alzheimer's disease consortium data accounts
- Allows researchers to narrow down a study population that relates to their work across multiple partner consortiums
- Provides tools capable of analyzing clinical, imaging and genetic data types via the LONI Pipeline



Global Partners and Affiliates

LONI

neuGRID

EMIF

ADNI

aibl

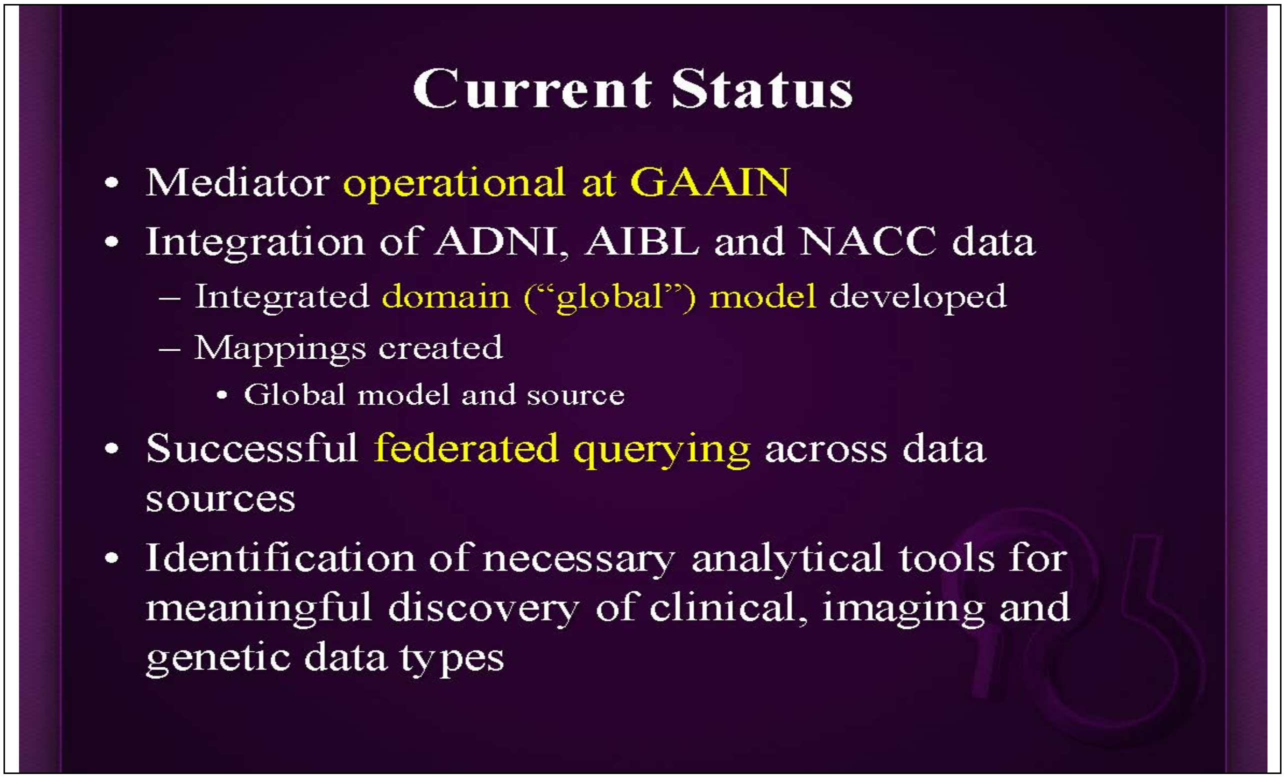
DIAN
Dominantly Inherited Alzheimer Network

CRITICAL PATH INSTITUTE

Common Representation Across Partner Data

- CDISC-CPATH Alzheimer's Therapeutic Area Standard
  - Domain Model
  - Common Data Elements
- CADRO* Ontology
  - Categories, Topics, Themes
  - Common Data Model linked to CDISC standards and CADRO

*Common Alzheimer Disease Research Ontology (CADRO) is a collaborative effort between the National Institute on Aging (NIA) and the Alzheimer's Association (AA)


Current Status
• Mediator operational at GAAIN
• Integration of ADNI, AIBL and NACC data
– Integrated domain ("global") model developed
– Mappings created
• Global model and source
• Successful federated querying across data sources
• Identification of necessary analytical tools for meaningful discovery of clinical, imaging and genetic data types