

GAAIN Update

Arthur W. Toga



What is GAAIN?

A global cooperative of sharing, investigation and discovery for Alzheimer's Disease Research

- Data federation platform enabling access to data across distributed studies
- Global network of analysis and workflow software and tools
- Global network of grid resources



Era of Big Data

- Exponential growth in data generation on every level
- Data is dispersed globally
- Data is heterogeneous
- Data is collected in non-standardized ways
- Google doesn't do for searching databases, medical images or genetic data
- There is a huge amount of data out there that needs to be made accessible through a common framework



GAAIN Leadership

- *Alzheimer's Association*
- *Laboratory of Neuro Imaging (LONI)*
- *Istituto di Ricovero e Cura a Carattere Scientifico (N4U)*
- *Scientific Advisory Board*



Scientific Advisory Board

- Paul Aisen, UC San Diego
- Rhett Alden, GE Healthcare
- Neil Buckholtz, NIH
- Enrique Castro-Leon, Intel
- Alon Halevy, Google
- William Klunk, Univ of Pittsburgh



GAAIN Partners

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Australian Imaging, Biomarkers and Life Study (AIBL)
- Coalition Against Major Diseases (CAMD)
- European Medical Information Framework (EMIF)
- NeuGRID



LONI Infrastructure



Data Center

- 208 servers
- 3328 cores
- 26 TB memory
- 2.43 PB usable storage
- 3.3 PB raw storage
- 136 TB tape back-up
(expandable to double capacity)



GAAIN Challenges

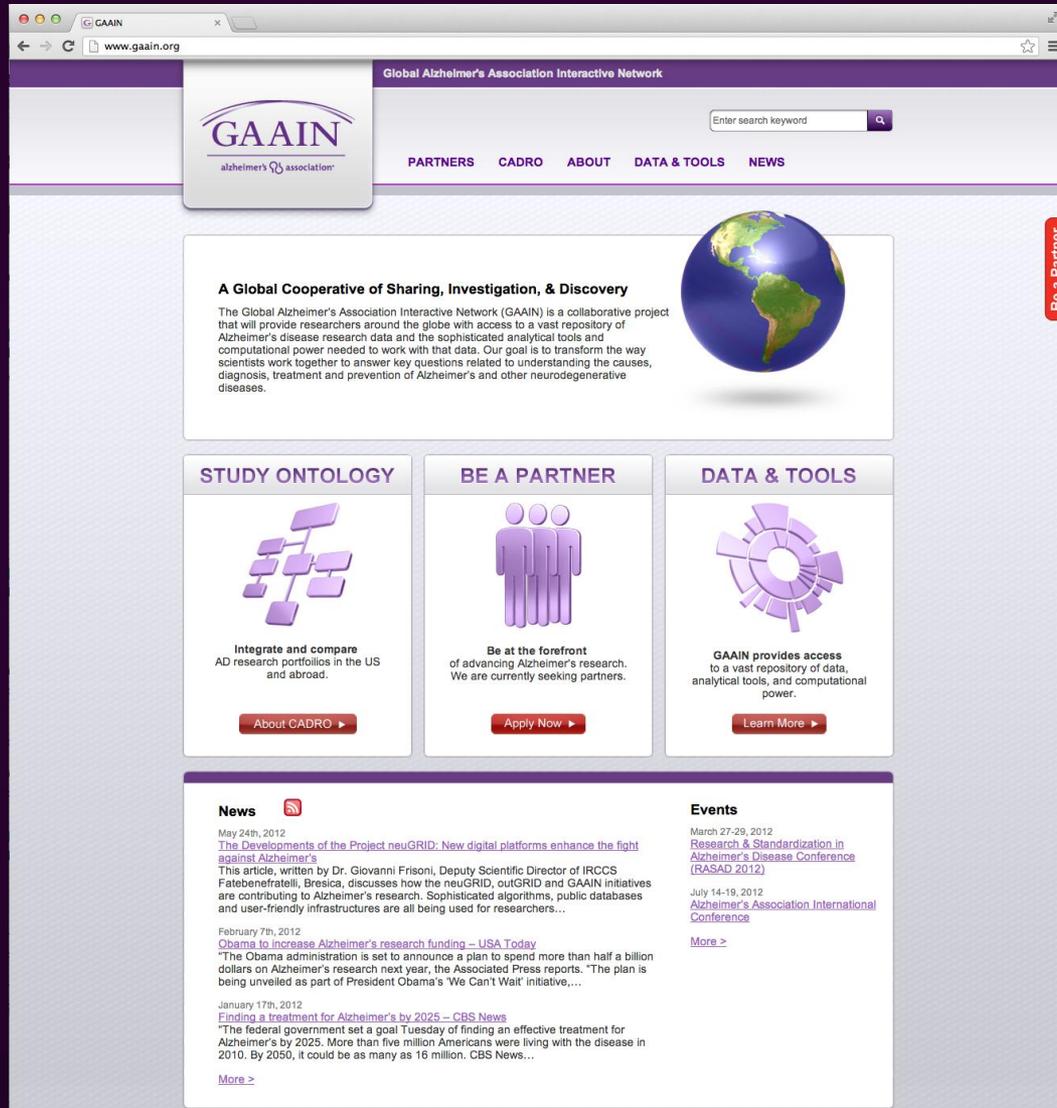
- Subject privacy protection across international boundaries
- Complexity from cross-disciplinary, multi-source data collection and analysis
- Creating robust and compelling tools for searching, sharing, visualizing and analyzing federated data
- Data “ownership” and security
- Data analysis across distributed GAAIN infrastructure
- Ontologies, terminologies, standards



Early Progress

- Initial set of partners on-board
 - Website deployed
 - Additional partner applications received via online application form
 - Initial set of standard terminologies selected
 - Subset of data exchanged
 - Mapping tool prototype in development
 - Federated database infrastructure in development
 - On-going communication with developers of CDISC terminologies
- 

Gaain.org | Homepage



The screenshot shows the Gaain.org homepage in a browser window. The browser's address bar displays "www.gaain.org". The page features a purple header with the GAAIN logo (Alzheimer's Association) and a search bar. Navigation links include PARTNERS, CADRO, ABOUT, DATA & TOOLS, and NEWS. The main content area is divided into several sections: a central introductory paragraph about GAAIN, a globe image, and three columns: "STUDY ONTOLOGY" (with a tree diagram), "BE A PARTNER" (with a group of people icon), and "DATA & TOOLS" (with a gear icon). At the bottom, there are "News" and "Events" sections with recent updates and links.

Global Alzheimer's Association Interactive Network

GAAIN
alzheimer's association

Enter search keyword

PARTNERS CADRO ABOUT DATA & TOOLS NEWS

A Global Cooperative of Sharing, Investigation, & Discovery

The Global Alzheimer's Association Interactive Network (GAAIN) is a collaborative project that will provide researchers around the globe with access to a vast repository of Alzheimer's disease research data and the sophisticated analytical tools and computational power needed to work with that data. Our goal is to transform the way scientists work together to answer key questions related to understanding the causes, diagnosis, treatment and prevention of Alzheimer's and other neurodegenerative diseases.

STUDY ONTOLOGY

Integrate and compare
AD research portfolios in the US and abroad.

About CADRO ▶

BE A PARTNER

Be at the forefront
of advancing Alzheimer's research.
We are currently seeking partners.

Apply Now ▶

DATA & TOOLS

GAAIN provides access
to a vast repository of data,
analytical tools, and computational
power.

Learn More ▶

News

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, Brescia, 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..."

[More >](#)

Events

March 27-29, 2012
[Research & Standardization in Alzheimer's Disease Conference \(RASAD 2012\)](#)

July 14-19, 2012
[Alzheimer's Association International Conference](#)

[More >](#)

Be a Partner

Partner Application

The screenshot shows a web browser window displaying the GAAIN website. The browser's address bar shows the URL <https://www.google.com>. The website's header includes the GAAIN logo (Global Alzheimer's Association Interactive Network) and a search bar. The main navigation menu contains links for PARTNERS, REGISTRY, ABOUT, DATA & TOOLS, and NEWS. The content area is divided into two columns. The left column, titled 'PARTNERS', describes the network's purpose and lists benefits of partnership, such as access to data, tools, and a global network of investigators. It includes a 'Download Application Form (PDF)' button and an illustration of stylized human figures. The right column, titled 'Are you a GAAIN Partner?', lists expectations for new partners, including having extensive research data, interest in collaborative efforts, and willingness to share data and test new tools. A vertical red sidebar on the right contains the text 'Study Registry' and 'Be a Partner'.

Home > Partners

PARTNERS

Network Partners in GAAIN will become part of a global effort to share and collaborate. GAAIN will link investigators, data, tools and infrastructure, creating a powerful and efficient environment for the study of Alzheimer's Disease.

Benefits of GAAIN Network Partnership include:

- Prestigious membership in a global network of investigators committed to advancing a sustainable and federated Alzheimer's Disease research infrastructure
- Access to a vast network of data and computer resources
- Access to a growing set of analytic tools spanning clinical, biologic, imaging and genetic data
- Data federation tools that support ontologies and metadata standards
- Online analytic tools allowing you to compare your data and analysis results with others.

 [Download Application Form \(PDF\)](#)



Are you a GAAIN Partner?

Our expectations for new GAAIN partners include:

- Extensive, accessible and well organized set of Alzheimer's Disease research data from imaging, behavior or genetics (300+ research subjects)
- Experience and interest in collaborative efforts
- Ability to share de-identified data*
- Reliable computer resources and high speed internet access to support data federation
- Technical expertise sufficient to test and implement data federation tools
- Willingness to beta test new tools
- Proponent of ethical data sharing*

* Only data collected in a manner that allows sharing in de-identified form may be part of the GAAIN effort. Strong and responsive support from your institution's ethics review committee is a plus.

[Study Registry](#)

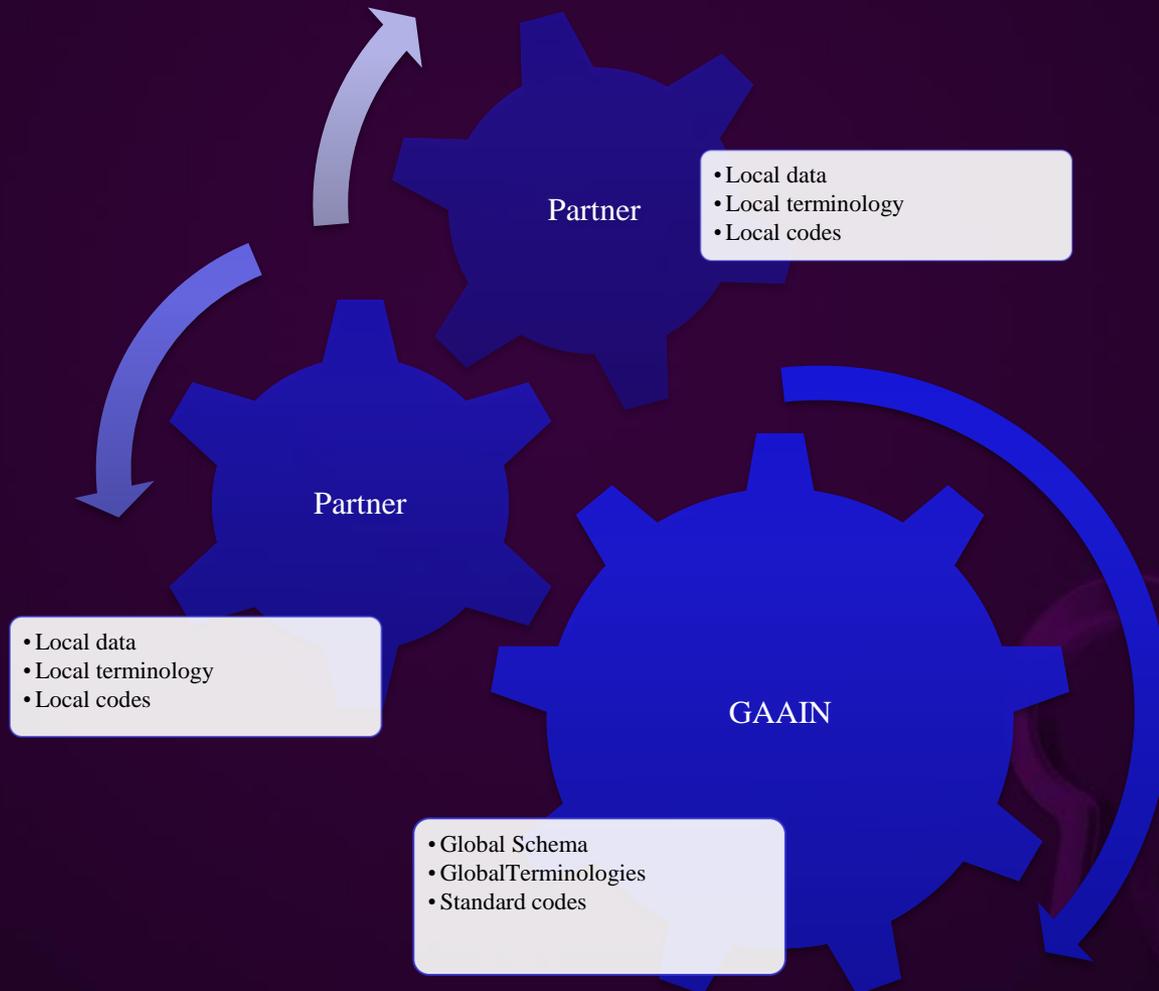
[Be a Partner](#)

Standard Terminologies

- Standard terminologies enable mapping of heterogeneous data into common terms
 - Supports searching across databases
 - Supports data federation
- CDISC Terminologies
 - CDISC AD v1.1 standard for Alzheimer's Disease
 - On-going discussions with CDISC on the neuroimaging standard currently under development



Mapping Data



Mapping Tool Prototype

1 : Select x

Table: Trash

<input checked="" type="checkbox"/>	subject_identifier	education_level	marital_status	primary_occupation	Occupation [primary_occupation]	New
▶	12762	3	married	housewife	housewife	
▶	12761	8	widowed	dealer	dealer	
▶	12760	4	widowed	labourer	employee	
▶	12758	5	widowed	dealer	dealer	
▶	12755	11	married	employee	employee	

Sort: ▼

Group:

Filter: Maps: Occupation: Description of new map

Map Name: Occupation x

Map Description: Description of new map

Occupation [0] Double-click to add = IF THEN Trash

ELSE

1 : Insert x

Table: Trash

<input checked="" type="checkbox"/>	subject_reference_id	project_code	site_id	subject_sex	subject_education
Set:	<input type="text" value="subject_identifier"/>	<input type="text" value="GAAIN"/>	<input type="text" value="34"/>	<input type="text" value="X"/>	<input type="text" value="education_level"/>
▶	12762	GAAIN	34	X	3
	TR1121	XSB	1	M	0
	TR2104	XSB	1	M	0
	TR2103	XSB	1	M	0
	TR2102	XSB	1	F	0
	TR1120	XSB	1	M	0

Sort: ▼

Group:

Maps:

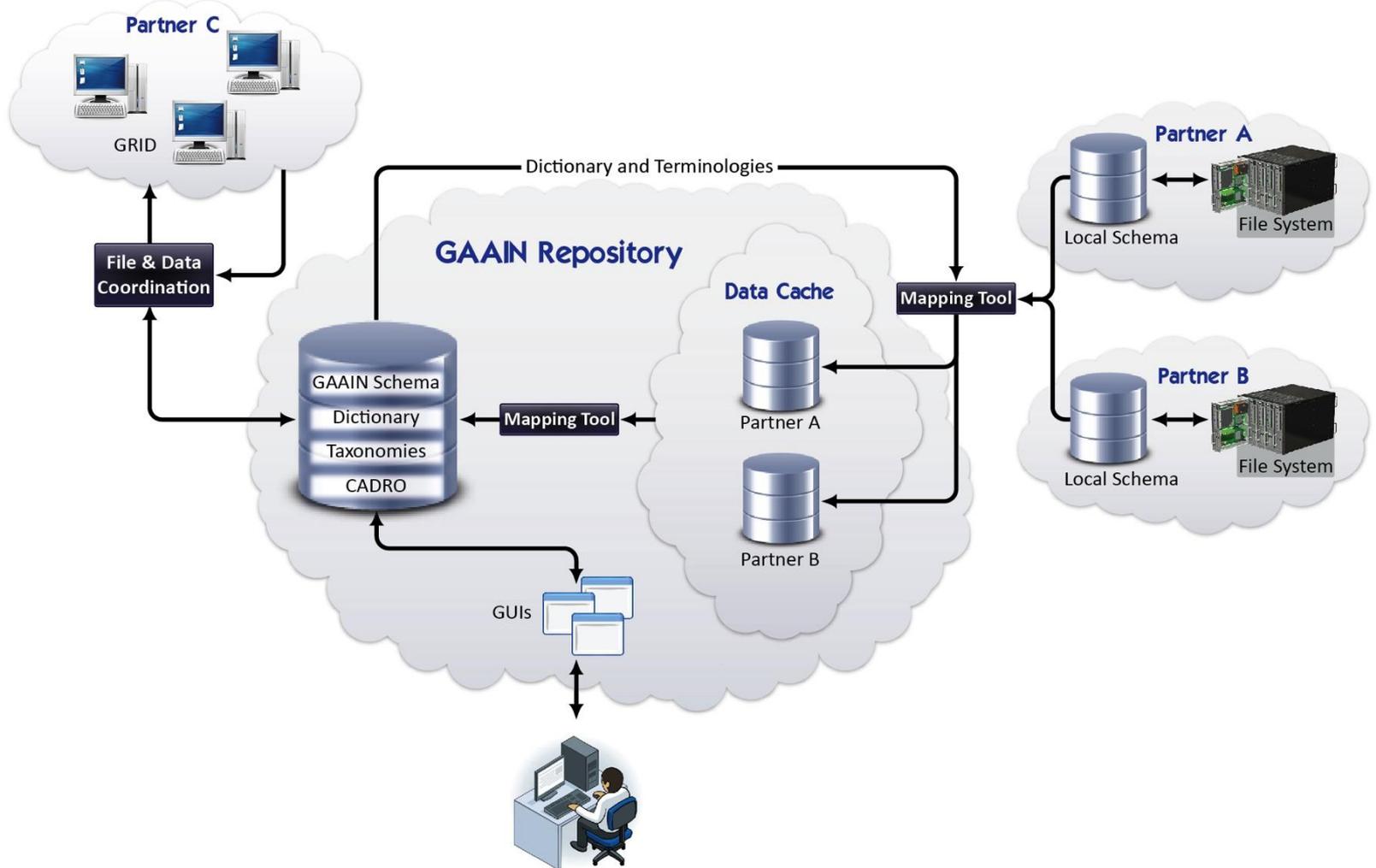
Each Insert: ▼

Each Error: ▼

Each Row: ▼

2 :

Data Federation



GAAIN Client Prototype

Global Alzheimer's Association Interactive Network Client

Configuration

Share Data

Register



Go Offline

Status

GAAIN Home

Edit Configuration

Admin Port ?

Data Directory ?

Save

Cancel

Inventory Search

Global Alzheimer's Association Interactive Network [Login](#)

Search

[PARTNERS](#) [CADRO](#) [ABOUT](#) [DATA & TOOLS](#) [NEWS](#)

INVENTORY	ADNI	AIBL	DIAN	NAU	Partner X
	access	access	access	access	<input type="button" value="apply"/>
Neuropsychological Assessments ▲	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
MMSE: Mini-mental State Examination	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
BDI: Beck Depression Inventory	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ADAS: Alzheimer's Disease Assessment Scale	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Clinical Dementia Rating	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Functional Assessments ▼	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Biological Markers ▼	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Physical Examinations ▼	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Neural Imaging ▼	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Genetics ▼	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Ontology-Directed Search

The screenshot displays the GAIN website interface. At the top, the header includes the GAIN logo (Global Alzheimer's Association Interactive Network) and a search bar. The main navigation menu consists of links for PARTNERS, CADRO, ABOUT, DATA & TOOLS, and NEWS. Below the navigation, there are three buttons: Display Results, Clear, and Export. The left sidebar is a hierarchical menu with the following sections:

- SUBJECT CHARACTERISTICS -
 - Socio-demographics ▾
 - Family History ▾
- EXAMINATIONS AND ASSESSMENTS -
 - Neuropsychological Assessments ▲
 - MMSE: Mini-mental State Examination
 - BDI: Beck Depression Inventory
 - ADAS: Alzheimer's Disease Assessment Scale
 - Clinical Dementia Rating
 - Functional Assessments ▾
- BIOLOGICAL MARKERS +
- PHYSICAL EXAMINATIONS +
- NEUROIMAGING +

The main content area features a large grey box containing three filter buttons: Socio-demographics, Functional Assessments, and Neuroimaging.

Genetic Data

- Circa 2010 GWAS Data (per sample)
 - 620,000+ rows of data
 - ~81MB
- 2012: Full Genome Sequencing (per sample)
 - Standard output from Illumina – multiple files and formats
 - ~250GB per sample
- Example (ADNI)
 - 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!



ADNI Genetics Data



ADNI @LONI

PROJECTS

SEARCH

ARCHIVE

DOWNLOAD

EXPLORE

MANAGE

LONI Home

Study Data

Image Collections

Genetic Data

Download Genetic Data

Reminder: The ADNI Data Use agreement prohibits unauthorized sharing of these data, posting to public databases and any attempt data to identify individuals using these data. By downloading these data you acknowledge to our terms & conditions. [Link to Data Use Agreement].

ADNI WGS

Documentation

Indels Data

SNPs Data

ALL

ADNI1 GWAS

ALL

ADNI WGS: ALL

ADNI Whole Genome Sequencing (WGS) samples were genotyped using the Illumina Omni 2.5M BeadChip and assembly performed using CASAVA-1.9.0a1_110909.90.

WGS Documentation

DNA Source Reference for WGS Samples	Version:1	.xlsx format
WGS Methods	Version:1	.pdf format

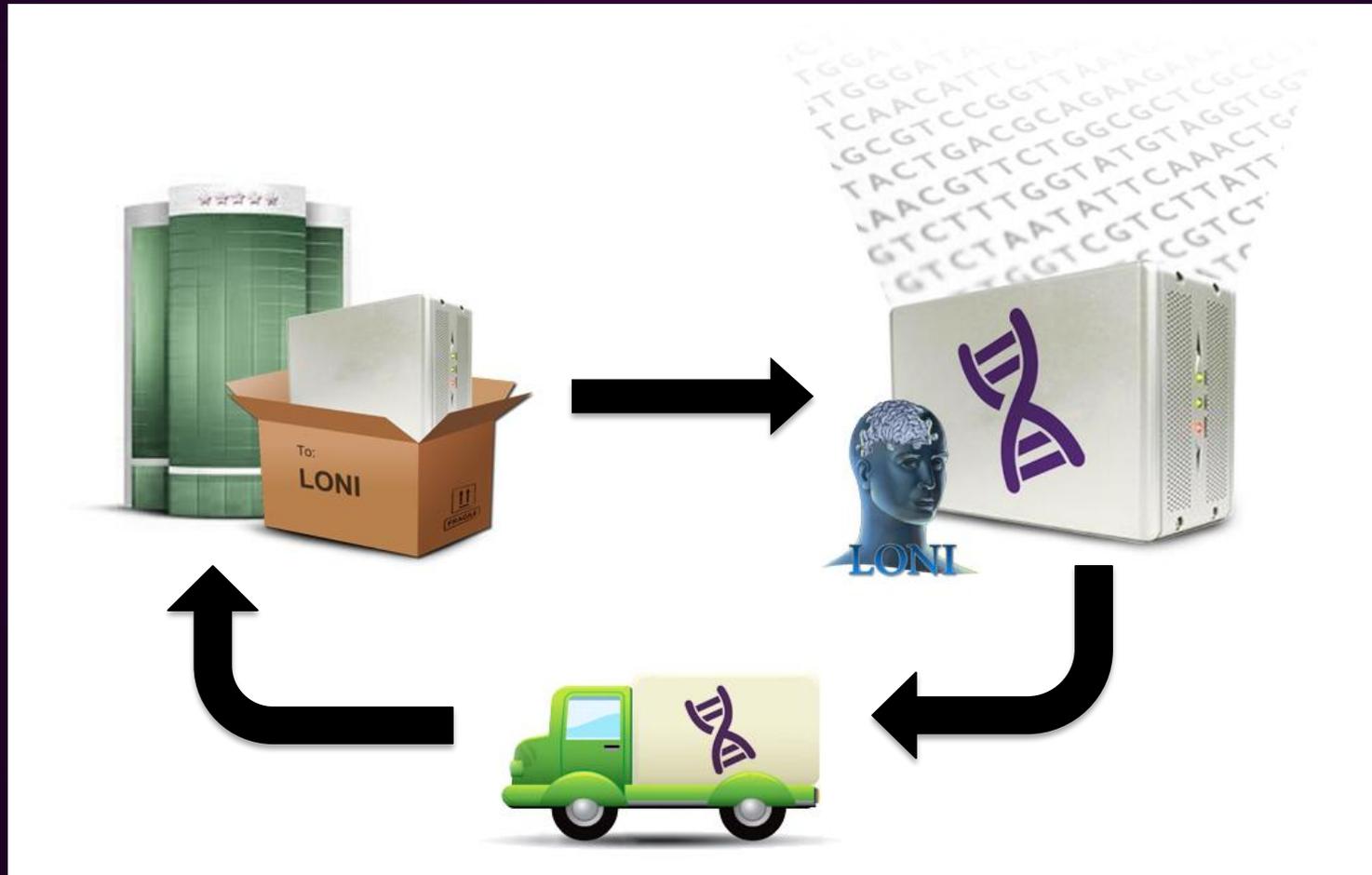
Indels Data

Indels Data (1 of 7)	.vcf format
Indels Data (2 of 7)	.vcf format
Indels Data (3 of 7)	.vcf format
Indels Data (4 of 7)	.vcf format
Indels Data (5 of 7)	.vcf format
Indels Data (6 of 7)	.vcf format
Indels Data (7 of 7)	.vcf format

SNPs Data

SNPs Data (1 of 28)	.vcf format
SNPs Data (1 of 28)	.vcf format

Lend Lease



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 Growth Process

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

ADNI Example:

One 22MB structural scan →

Five preprocessed images (176 MB) →

Eleven postprocessed images (222 MB)

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

