Published in final edited form as:

Alzheimers Dement. 2016 January ; 12(1): 49–54. doi:10.1016/j.jalz.2015.06.1896.

The Global Alzheimer's Association Interactive Network

Arthur W. Toga^{*}, Scott C. Neu, Priya Bhatt, Karen L. Crawford, and Naveen Ashish

Laboratory of Neuro Imaging, Institute for Neuroimaging and Informatics, Keck School of Medicine, University of Southern California, Los Angeles, California 90095, USA

Abstract

INTRODUCTION—The Global Alzheimer's Association Interactive Network (GAAIN) is consolidating the efforts of independent Alzheimer's disease data repositories around the world with the goals of revealing more insights into the causes of Alzheimer's disease, improving treatments, and designing preventative measures that delay the onset of physical symptoms.

METHODS—We developed a system for federating these repositories that is reliant upon the tenets that (a) its participants require incentives to join, (b) joining the network is not disruptive to existing repository systems, and (c) the data ownership rights of its members are protected.

RESULTS—We are currently in various phases of recruitment with over 55 data repositories in North America, Europe, Asia and Australia and can presently query 250,000+ subjects using GAAIN's search interfaces.

DISCUSSION—GAAIN's data sharing philosophy, which guided our architectural choices, is conducive to motivating membership in a voluntary data sharing network.

Keywords

GAAIN; data sharing; federated data repositories; Alzheimer's disease

1. Introduction

The current upsurge of collaborative initiatives that utilize big data has opened up new avenues in the exploration of large data sets for accelerating research on Alzheimer's disease [1]. For example, studies are now underway looking for people with genetic resistance to diseases, including Alzheimer's disease [2]. By understanding how the protective mutations work, it may be possible to develop drugs that can provide disease resistance to everyone. Finding these rare individuals requires intensive searching; these outliers cannot be found without first accumulating very large sets of population data. Also, bringing together data of different scopes, including those on genetic, imaging, and phenotypic scales, has the potential to provide new understandings and lead to new treatments and cures by revealing relationships (such as new biomarkers) that are not apparent at any single scale. Simply stated, the statistical significance of research findings is primarily dependent upon the amount of data available to study. As the questions become more complicated, the data that

^{*}To whom correspondence should be addressed; Arthur W. Toga, Laboratory of Neuro Imaging, The Institute for Neuroimaging and Informatics, Keck School of Medicine of USC, University of Southern California, 2001 North Soto Street - Room 102, Los Angeles, CA 90032, Phone: (323) 44-BRAIN (442-7246), toga@loni.usc.edu.

can provide answers becomes more difficult to find and it becomes necessary to aggregate more and more data collections together.

As a model of collaborative research, the Alzheimer's Disease Neuroimaging Initiative (ADNI) has been successful in standardizing its data acquisition protocols, allowing its results to be compared together across participating sites, and in making new data publicly available shortly after it is archived [3, 4]. ADNI is focused on developing clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease [5] and its success has led to the establishment of ADNI-like programs in Australia, Europe, Japan, Argentina, and Korea and to study other diseases such as Parkinson's disease [6]. The ADNI data sharing model was a break in what was the normal culture of hoarding data for years, setting aside concerns that others might publish results first in order to speed up hypothesis testing and develop methods to determine treatment effects in trials [7]. There were approximately 200 papers published as a direct result of ADNI during the first 6 years of its funding [8] and over 400 papers published to date¹. As one would expect, access to the data is not entirely without restrictions. ADNI's data use agreement² requires that all scientific investigators cite ADNI in manuscripts that use its data and that redistribution of the data in any manner is prohibited that all scientific investigators cite ADNI in manuscripts that use its data and that redistribution of the data in any manner is prohibited.

The concerns of the research community with regards to data ownership should not be underestimated. Fifteen years ago brain researchers openly resisted [9] when an editor of the Journal of Cognitive Neuroscience declared that submission of the functional magnetic resonance imaging (fMRI) files associated with their studies would be a mandatory condition for publishing their articles [10]. Worried that other journal editors would also adopt this requirement, researchers argued against the practicality of data curation and it became clear that the lack of rewards for sharing data would prevent voluntary adherence to the data submission requirement [11]. Ultimately the journals chose not to force authors to submit their data along with their publications. There are however many journals that currently require supporting data to be made available in a public archive (e.g., Nature³, Science⁴, and Cell⁵) but have done so with mixed results [12, 13, 14, 15].

In addition to its data sharing policy⁶ that requires studies funded over \$500,000 to make their final research data publicly available, the National Institutes of Health (NIH) has adopted the requirement that every new research project it funds to study traumatic brain injury (TBI) must upload all its acquired data to the Federal Interagency Traumatic Brain Injury Research⁷ (FITBIR) informatics system on a quarterly basis. FITBIR is under development by the NIH and the Department of Defense (DOD) to promote collaboration on and advance the study of TBI. Participation is not voluntary and it is intended that FITBIR will serve as a centralized information system for TBI research. The NIH also supports the

¹<http://www.adni-info.org/scientists/adniscientistshome/adnipublications.aspx>

²http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Data_Use_Agreement.pdf

³<http://www.nature.com/authors/gta.pdf>

⁴http://www.sciencemag.org/site/feature/contribinfo/prep/prep_online.xhtml

⁵<http://www.cell.com/cancer-cell/authors>

⁶http://grants.nih.gov/grants/policy/data_sharing/

⁷<https://fitbir.nih.gov/jsp/about/index.jsp>

National Database for Autism Research (NDAR) [16] to promote data sharing and collaboration to advance research on autism spectrum disorders. Similar to FITBIR, many NIH awardees in the autism field are expected to submit their data to NDAR. However, unlike FITBIR, NDAR is federated with 4 private databases that keep data in their respective locations while enabling users to search across all the databases. This federation strategy was easier to implement rather than trying to persuade the private databases to donate their data to NDAR [16]. Researchers who do not have access to NDAR or its federated databases are allowed to see if there is enough data there to warrant applying for access. Both NDAR and FITBIR prohibit any researcher from distributing data without the permission of their data access committees.

The ADNI, NDAR, and FITBIR initiatives all demonstrate that significant collaboration of research data can occur when funding agencies mandate that all participants copy their data into a central archive. Without such a mandate, there are very few incentives to promote this type of data sharing. What is also apparent is the limited reach of these collaborative efforts. They prohibit inter-initiative collaborations by placing restrictions on the redistribution of their data. We believe that data federation on a larger scale must be accomplished with the “carrot,” and not the “stick.” There must be sufficient incentives to promote participation in a network that aggregates these kinds of collaborations.

The primary objective of the Global Alzheimer’s Association Interactive Network (GAAIN⁸) is to establish a virtual community for sharing Alzheimer’s disease-related data stored in independently operated repositories around the world. The information in these data repositories is currently not shared between them and their combined potential remains untapped. Aggregating these data together may reveal more insights into the causes of Alzheimer’s disease, improve treatments, and design preventative measures that delay the onset of physical symptoms. But this effort is not without its challenges; the needs and concerns of all those involved must be properly addressed. To this end our focus has first been on sociological issues and then on technical ones. In what follows, we discuss our philosophy and federation policies and then we report on our progress after our first year of work.

2. Methods

From our correspondence with organizations that manage their own data repositories we have found that, unlike individual research scientists who independently study and collect data on particular cohorts, they have different resources available to them and therefore have different concerns regarding data sharing. Sharing data can be complicated because each data repository may be subject to local policies, ethical considerations, and legal obligations; often an Institutional Review Board places significant limits on the way that data can be shared and even stronger limits on its re-distribution [16]. Additionally, most of these organizations have invested considerable time and resources in building their data repositories and are receptive to joining a data sharing network only if it is minimally disruptive to their systems and only if their data repositories continue to manage their data.

⁸<http://www.gaain.org>

Design and policy choices that motivate membership are critical because participation in GAAIN is voluntary.

GAAIN provides incentives to join its federation by advertising the data collected by its data partners. The identities of all its data partners, including their logos and URL links that forward investigators to their web sites, are displayed on each GAAIN search page. This not only increases the public visibility of each data repository, but can also help partners comply with data sharing requirements of their funding agencies. Further, GAAIN addresses the data ownership and interdependence concerns of its data partners through the following policies: (a) GAAIN data partners retain complete control of their data and continue to use their existing web application pages, (b) GAAIN connections do not interfere with or consume the resources of data partner repository systems, (c) unless otherwise requested, GAAIN will not store data from any data partner on a GAAIN server computer disk, and (d) all communications between GAAIN and its data partners are conducted securely.

In accordance with the above policies, GAAIN enables investigators to search over all data in its network using graphs. The reason for using graphs is two- fold. First, correlations and trends in data can be visualized without revealing the actual data. Each graph can plot only two subject attributes at a time and there are data points from multiple subjects in each graph. It is not possible to infer all the subject data because subjects cannot be traced across different graphs. This prohibits unauthorized investigators from downloading the data without the permission of GAAIN's data partners. Second, data visualization through graphs is sufficient to motivate analysis. GAAIN encourages future study of the data in its network by directing investigators to appropriate data partners. GAAIN does not perform analyses of its own and therefore has no reason to copy and use the actual data values for its own purposes. It should be noted that these graphs are no different than plots and charts published in research papers in that they must maintain subject confidentiality. Nothing should lead someone to believe that his or her identity has been exposed by a data point (e.g., an outlier) in the search results. Subject identities are known only to the data partners and GAAIN explicitly requires data partners to review their data before making it searchable.

The GAAIN Interrogator is the main search interface for GAAIN (Fig. 1) and it displays its search results using graphs. Investigators search the data in GAAIN by defining and visually comparing the attributes of two cohorts of subjects. As the cohort definitions are incrementally adjusted by the investigator, the Interrogator graphs automatically update and the investigator utilizes this feedback to make further adjustments. In this way the investigator effectively "interrogates" the data in GAAIN. With each adjustment to the cohort definitions, the Interrogator displays the total number of subjects in each investigator-defined cohort from each data partner. Clicking the apply link adjacent to a data partner's totals forwards the investigator to the data partner's application web page where the investigator can apply for access to the data.

Since all subjects across different research studies typically do not have the same data collected for them, GAAIN also provides investigators an additional search interface so they can get the actual numbers of subjects available for their work. The GAAIN Scoreboard

(Fig. 1) displays subject tallies of data (such as gender, diagnosis, or race) and allows investigators to refine the display by selecting combinations of attributes as well. In the latter case, a subject is counted only if there is data collected for that subject for all selected attributes (e.g., gender, diagnosis, and race). Periodically, GAAIN will check the network connection status of each data partner and if a data partner does not have a connection then the Scoreboard displays red “X”s instead of the subject counts for that data partner. Data partners are therefore encouraged to maintain their connections to the network in order to avoid public ignominy.

Each data partner is required to sign a non-legally binding Memorandum of Understanding (MOU) prior to joining GAAIN that formalizes GAAIN’s data sharing policies and other terms and conditions of GAAIN participation. Data partners must ensure that no subject-identifying information can be extracted from their data and they must maintain their connections to the network. As incentives for joining, the MOU also obligates GAAIN to display data partners’ logos on GAAIN web pages along with brief descriptions of their studies and URL links to their web sites.

Investigators may register with GAAIN and access the GAAIN Interrogator without first being approved by a data sharing committee. Investigators are only required to have valid email addresses and to agree to cite GAAIN’s data partners in all publications and presentations using their data. We require investigators to register with GAAIN so that we can track usage and report usage statistics to our data partners, and we plan to display these statistics on GAAIN’s web site. The GAAIN Scoreboard has been made publicly accessible in order to generate interest in GAAIN and to motivate potential investigators to register with GAAIN. Both the GAAIN Scoreboard and Interrogator have undergone extensive testing and have been modified in accordance with feedback we received from users, our data partners, and from GAAIN’s advisory board.

3. Results

We are currently in various phases of partner recruitment with over 55 data repositories in North America, Europe, Asia and Australia. Appendix A lists many important Alzheimer’s and aging data repositories that have joined GAAIN or are in ongoing discussions to join GAAIN. The on-boarding process of these groups tends to be lengthy, averaging at least three months per data partner. This is partially because our data partners often require approvals from internal executive committees that meet infrequently and their approval processes may require them to review detailed written proposals. After project leaders have learned about GAAIN’s approach, they usually involve their technical support personnel who have questions of their own. It has thus been useful to conduct a series of web conference meetings during the recruitment phase where we introduce GAAIN’s policies, demonstrate its operations, and then walk through technical requirements.

GAAIN has defined clear boundaries that delineate the responsibilities assigned to the data partner and to GAAIN for the purposes of mediating data. Because data partners import data into GAAIN clients after exporting CSV (comma-separated values) files from their databases, they decide what data to share and how often the data values are updated [17].

Although GAAIN recommends semi-annual updates, data partners are free to schedule their updates at their own frequency. There is no prescribed format for these CSV files and establishing the data mapping (transforming the nomenclature and conventions used by the data partner into those used in GAAIN) is the responsibility of GAAIN. This is another incentive to join GAAIN since it removes the anxiety felt by most data partners, that they will be burdened with the work required to map their data. GAAIN allocates its resources to map data and is in the best position to construct the mappings because it best understands the global schema that all data is mapped to. It should be noted that it is crucial to involve data partners during the mapping process as they best understand the data that is being mapped. We have found our data partners to be very responsive to inquiries of their data because they are interested in having their data mapped and displayed properly in GAAIN's public interfaces. Most of our data partners have been willing to send us small data sets that are representative of the data they are sharing. This expedites the construction of the mappings because it provides examples of the actual data values. We currently write our mappings in computer code, but we plan on developing a mapping tool in the future to accelerate the process and to support changes to existing mappings.

For our first year deliverable, we defined a small number (24) of attributes in our data sharing schema and that allowed us to quickly complete a full software development life cycle and start recruiting data partners. Even though we mapped such a small number of attributes, we still encountered some of the difficulties commonly encountered in data standardization. For example, logical memory was measured in most projects using the Wechsler Memory Scale [18] except in the Alzheimer's Disease Repository Without Borders (ARWIBO)⁹ project which used an Italian version [19, 20]. Amyloid- β 42 ($A\beta_{42}$), total tau (t-tau), and phosphorylated tau (p-tau) cerebrospinal fluid (CSF) biomarker levels were measured using different assays in ADNI [21] and in the Integrated Neurodegenerative Disease Database (INDD) [22]. Also, we found the protocol for hippocampal volume estimation in ADNI [23] to be different than the one used in the Layton Aging & Alzheimer's Disease Center (LAADC) [24]. Over the next year, we plan to add many more attributes to our schema and to allow for different methods of measurement. Rather than require data partners to conform to a new standardized ontology, we intend instead to map their attributes to a set of attributes that are common across data partners [25] and then incorporate attributes that are specific to each data partner without modification. We also encountered cases where subjects appeared across data repositories; for example, some subjects in the National Alzheimer's Coordinating Center (NACC) were also subjects in LAADC. Similarly to NDAR and FITBIR, we are currently in the process of evaluating unique identifier [26, 16, 27] for each subject so that we can identify the same subject between two data partners.

4. Discussion

Most of the Alzheimer's disease and aging data repositories we have approached have shown considerable interest in joining GAAIN. We believe this is in large part due to the opportunities for groups to increase their public visibility while retaining control of their

⁹<https://www.neugrid4you.eu>

data. Since participation in GAAIN is voluntary and its success depends upon the size of its membership, the relationships between GAAIN and its data partners are mutually beneficial.

It is our future goal to offer a data homogenization service to the scientific community. When investigators download data from multiple repositories, they must learn the different terminologies and nomenclature used by each data repository before they can transform the data into a single set of unified values. This process normally takes a great deal of time and is repetitive when multiple investigators download the same data.

Because GAAIN has already mapped partner data into its data sharing schema, investigators could save resources and time if they could download the mapped data. We are working on extending the GAAIN system so that if an investigator has been approved by multiple data partners and if the data partners give their consent to the service, the investigator will be able to download all the data in homogenized formats.

For our next year deliverable, we plan to incorporate analysis tools and computational resources into GAAIN so that we can extract information from the neuroimaging and genetic data files of our data partners. The derived data will be stored in GAAIN and integrated together with investigator search results. This, along with the addition of many more attributes, will greatly extend the search capabilities and utility of GAAIN. The Alzheimer's Association is committed to the continued development and support of GAAIN.

Acknowledgments

This work was supported by the Global Alzheimer's Association Interactive Network (GAAIN) initiative of the Alzheimer's Association and by National Institutes of Health grants 5P41 EB015922-16 and 1U54EB020406-01.

Appendix A. Alzheimer's and Aging Data Repositories

AgedBrainSYSBIO

Aims to address the basis of brain aging by studying the pathways involved in this process and by identifying the interactions through which the aging phenotype develops in normal and in disease conditions. The project aims at identifying and validating new molecular targets and biomarkers associated with late-onset Alzheimer's Disease.

AlzGene

Provides a comprehensive, unbiased and regularly updated field synopsis of genetic association studies performed in Alzheimer's disease.

Alzheimer's Disease Neuroimaging Initiative

A longitudinal study that assesses clinical, imaging, genetic and biospecimen biomarkers through the process of normal aging to early mild cognitive impairment (EMCI), to late mild cognitive impairment (LMCI), to dementia.

Alzheimer's Preventative Initiative

An international collaborative formed to launch a new era of Alzheimer's prevention research by evaluating the most promising therapies in cognitively normal people who, based on their age and genetic background, are at the highest imminent risk of developing Alzheimer's disease symptoms.

Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing

A study to discover which biomarkers, cognitive characteristics, and health and lifestyle factors determine subsequent development of symptomatic Alzheimer's Disease (AD).

Biomarkers of Cognitive Decline Among Normal Individuals

Identifies biomarkers associated with progression from normal cognitive status to cognitive impairment or dementia, with a particular focus on Alzheimer's Disease.

Brain Health Registry

Promotes healthy brain function through the prevention of brain diseases, brain disorders and brain injuries that affect brain function in adults. This is the first neuroscience project to leverage online possibilities in this way and on this large scale.

Comparative Effectiveness Research Trial of Alzheimer's Disease Drug

A study to learn about drug adherence, how patients take medications, and the side effects patients experience when taking these medications.

Canadian Longitudinal Study on Aging

National, long-term study that follows approximately 50,000 men and women between the ages of 45 and 85 for at least 20 years. The study collects information on the changing biological, medical, psychological, social, lifestyle and economic aspects of people's lives.

Dallas Lifespan Brain Study

Focuses on the study of 350 healthy adults, 50 from each decade from 20 to 89, to thoroughly characterize cognition, brain structure and function across the adult lifespan.

Dominantly Inherited Alzheimer's Network

An international research partnership of leading scientists determined to understand a rare form of Alzheimer's disease that is caused by a gene mutation.

European Alzheimer's Disease Initiative

The first extension of ADNI where 50 European clinical and research sites carried out a variety of multinational, multi-center investigations in Alzheimer's treatment, diagnosis and centralized data storage.

European Medical Information Framework

Aims to discover and validate biomarkers of Alzheimer's disease onset in the preclinical and prodromal phase as well as for disease progression and identify high-risk individuals for therapeutic trials for prevention.

Framingham Heart Study

Examines lifetime risk estimates for AD, environmental risk factors for AD, circulating and imaging markers of aging-related brain injury, and explores the genetics underlying AD.

French National Alzheimer's Database

Registers all medical acts performed by memory units and independent specialists where subject data is collected at several hundred memory centers all over France.

Fundaci3 ACE

A private organization dedicated to the diagnosis, treatment, research and support for people with Alzheimer's disease and other dementia.

Integrated Neurodegenerative Disease Database

Integrated database of multiple, aging related neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and frontotemporal lobar degeneration.

Layton Aging and Alzheimer's Disease Center

Primary emphasis is on studies of preclinical dementia, as well as early dementia. Well-characterized patients, clinical, MRI and genetic data, as well as biological specimens are made available to investigators and research groups worldwide.

National Alzheimer's Coordinating Center

Maintains a cumulative database including clinical evaluations, neuropathy data when available, and MRI imaging contributed by the 34 past and present Alzheimer's Disease Centers.

NIA Genetics of Alzheimer's Disease Data Storage Site

A national genetics data repository that facilitates access of genotypic data to qualified investigators for the study of the genetics of late-onset Alzheimer's Disease.

neuGRID for you

Web portal aimed to help neuroscientists do high-throughput imaging research and provide clinical neurologists automated diagnostic imaging markers of neurodegenerative diseases

for individual patient diagnosis in the fields of Alzheimer's disease, psychiatric diseases, and white matter diseases.

Oxford Project To Investigate Memory and Aging

Database of neuropsychological assessments, brain scans, blood samples, cerebrospinal fluid samples (CSF), physical examination data and histopathological information following brain donation.

Swedish Dementia Registry

National quality registry on dementia disorders financed by the Swedish Association of Local Authorities and Regions and the Swedish Brain Power network.

Translational Genomics Research Institute

Focused on making genomic discoveries in diseases and disorders in the areas of oncology, neurogenomics and metabolic disease. The institute has undertaken a large collaborative effort to identify every genetic variation that results in increased risk for Alzheimer's disease.

Texas Alzheimer's Research Care Consortium

Collaboration between six of Texas' leading medical research institutions to improve early diagnosis, treatment, and prevention of Alzheimer's disease.

The Three City Study (3C)

An observational study aiming to examine the relation between vascular diseases and dementia in adults 65 years and older in three French cities.

Wisconsin Longitudinal Study

Has followed the life course of 10,317 Wisconsin high school graduates of 1957 and randomly selected siblings through repeated surveys and has detailed records of educational, social, psychological, economic and mental and physical health characteristics in a homogeneous population.

Wisconsin Registry for Alzheimer's Prevention

An observational study that is tracking the characteristics and habits of two important groups of volunteers: people who have one or both parents with Alzheimer's disease and people whose parents lived to old age with no signs of Alzheimer's disease or other serious memory problems.

Women's Healthy Aging Project

A prospective, longitudinal, epidemiological study of 438 Australian women that has spanned two decades.

References

1. Husain M. Big data: could it ever cure alzheimer's disease? *Brain*. 2014; 137(10):2623–2624. [PubMed: 25217787]
2. Kolata G. New research looks for cures in mutations. *The New York Times*. 2014:A1.
3. Toga AW, Crawford KL. The informatics core of the alzheimer's disease neuroimaging initiative. *Alzheimer's & Dementia*. 2010; 6(3):247–256.
4. Toga AW. The clinical value of large neuroimaging data sets in alzheimer's disease. *Neuroimaging clinics of North America*. 2012; 22(1):107–118. [PubMed: 22284737]
5. Weiner MW, Aisen PS, Jack CR Jr, Jagust WJ, Trojanowski JQ, Shaw L, Saykin AJ, Morris JC, Cairns N, Beckett LA, et al. The alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's & Dementia*. 2010; 6(3):202–211.
6. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, Coffey C, Kiebertz K, Flagg E, Chowdhury S, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*. 2011; 95(4):629–635. [PubMed: 21930184]
7. Burton A. Big science for a big problem: Adni enters its second phase. *The Lancet Neurology*. 2011; 10(3):206–207. [PubMed: 21349436]
8. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, et al. The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia*. 2013; 9(5):e111–e194.
9. Neuroscience N. A debate over fmri data sharing. *Nat Neurosci*. 2000; 3:845–846. [PubMed: 10966604]
10. Marshall E. A ruckus over releasing images of the human brain. *Science*. 2000; 289(5484):1458–1459. [PubMed: 10991728]
11. Koslow SH. Sharing primary data: a threat or asset to discovery? *Nature Reviews Neuroscience*. 2002; 3(4):311–313.
12. Ventura B. Mandatory submission of microarray data to public repositories: how is it working? *Physiological genomics*. 2005; 20(2):153–156. [PubMed: 15661852]
13. Piwowar HA. Who shares? who doesn't? factors associated with openly archiving raw research data. *PloS one*. 2011; 6(7):e18657. [PubMed: 21765886]
14. Brazma A. Minimum information about a microarray experiment (miame)—successes, failures, challenges. *The Scientific World Journal*. 2009; 9:420–423. [PubMed: 19484163]
15. Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B. The user's view on biodiversity data sharing? investigating facts of acceptance and requirements to realize a sustainable use of research data? *Ecological Informatics*. 2012; 11:25–33.
16. Hall D, Huerta MF, McAuliffe MJ, Farber GK. Sharing heterogeneous data: the national database for autism research. *Neuroinformatics*. 2012; 10(4):331–339. [PubMed: 22622767]
17. Neu SC, Crawford KL, Toga AW. Sharing data in the global alzheimer's association interactive network. *NeuroImage*. doi: dx.doi.org/10.1016/j.neuroimage.2015.05.082.
18. Wechsler D. A standardized memory scale for clinical use. *The Journal of Psychology*. 1945; 19(1):87–95.
19. G. italiano per lo studio neuropsicologico dell'invecchiamento. Spinnler H, Tognoni G. Standardizzazione e taratura italiana di test neuropsicologici. *Masson Italia Periodici*. 1987
20. Novelli G, Papagno C, Capitani E, Laiacina M, et al. Tre test clinici di memoria verbale a lungo termine: taratura su soggetti normali. *Archivio di psicologia, neurologia e psichiatria*.
21. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, et al. Cerebrospinal fluid biomarker signature in alzheimer's

- disease neuroimaging initiative subjects. *Annals of neurology*. 2009; 65(4):403–413. [PubMed: 19296504]
22. Toledo JB, Brettschneider J, Grossman M, Arnold SE, Hu WT, Xie SX, Lee VMY, Shaw LM, Trojanowski JQ. Csf biomarkers cutoffs: the importance of coincident neuropathological diseases. *Acta neuropathologica*. 2012; 124(1):23–35. [PubMed: 22526019]
23. Schuff N, Woerner N, Boreta L, Kornfield T, Shaw L, Trojanowski J, Thompson P, Jack C, Weiner M, et al. Mri of hippocampal volume loss in early alzheimer's disease in relation to apoe genotype and biomarkers. *Brain*. 2009; 132(4):1067–1077. [PubMed: 19251758]
24. Kaye JA, Swihart T, Howieson D, Dame A, Moore M, Karnos T, Camicioli R, Ball M, Oken B, Sexton G. Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology*. 1997; 48(5):1297–1304. [PubMed: 9153461]
25. Kuchinke W, Aerts J, Semler S, Ohmann C, et al. Cdisc standard-based electronic archiving of clinical trials. *Methods of information in medicine*. 2009; 48(5):408. [PubMed: 19621114]
26. Johnson SB, Whitney G, McAuliffe M, Wang H, McCreedy E, Rozenblit L, Evans CC. Using global unique identifiers to link autism collections. *Journal of the American Medical Informatics Association*. 2010; 17(6):689–695. [PubMed: 20962132]
27. Christen P. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer Science & Business Media. 2012

Figure 1. GAAIN search interfaces. The GAAIN Interrogator (top) supports data discovery through interactive graphs that refine search criteria and the GAAIN Scoreboard (bottom) provides a summary view of the data shared by its data partners.