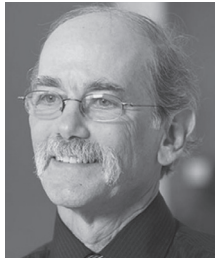# Data Sharing in Alzheimer's Disease Research

An Expert Interview with Arthur W Toga

University of Southern California (USC) Mark and Mary Stevens Neuroimaging and Informatics Institute, Los Angeles, CA, US

Arthur W Toga
Arthur W Toga is Director of the University of Southern California (USC) Mark and Mary Stevens Neuroimaging and Informatics Institute and Provost Professor with a primary appointment in the Department of Ophthalmology. He was trained in neuroscience and computer science and has written more than 1,000 papers, chapters and abstracts, including eight books. Recruited to USC in 2013, he also directs the Laboratory of Neuro Imaging. This 110-member laboratory includes graduate students from computer science, biostatistics and neuroscience. It is funded with grants from the National Institutes of Health (NIH) as well as industry partners. He has received numerous awards and honors in computer science, graphics and neuroscience. Prior to joining USC, he held several roles at UCLA, including Distinguished Professor of Neurology, the Geffen Chair of Informatics at the David Geffen School of Medicine, Associate Director of the Brain Mapping Division within the Neuropsychiatric Institute, and Associate Dean of the David Geffen School of Medicine. He is the founding Editor-in-Chief of the journal NeuroImage and holds the chairmanship of numerous committees within the NIH and a variety of international task forces.

There has been an explosion of data in the field of Alzheimer's disease (AD), not only from clinical studies but also studies that generate hypotheses and opportunities that may accelerate drug development. In order to make optimal use of these data, scientists must share findings across organizations and countries. While many researchers recognize the importance of data sharing, they often lack the capability to share it. The Global Alzheimer's Association Interactive Network (GAAIN) is a recent initiative that aims to create a global network of AD data, researchers, analytical tools and computational resources to enhance our understanding of this condition.[1] In an expert interview, Arthur W Toga discusses the need for data sharing and the GAAIN initiative.

Q. What are the major initiatives for data sharing in Alzheimer's disease research, and who can access these data?

There are several initiatives focused on data sharing, an area that has continued to evolve. Perhaps one of the most significant milestones occurred around 15 years ago with the launch of the Alzheimer's Disease Neuroimaging Initiative (ADNI) project in the US, which was funded by the National Institute of Health (NIH) and pharmaceutical companies. The primary goal of ADNI, which comprised around 60 acquisition sites throughout the country,[2] was to collect a range of data, including imaging, genetics, cognitive measures, and measurements from blood and cerebrospinal fluid. These data were used to identify biomarkers of AD. ADNI's government and industry leaders chose to immediately release all data collected to the entire scientific community, without any embargo. This was unique at the time and had a transformative effect on AD research. As the Informatics Core for ADNI, data comes to us at the Laboratory of Neuro Imaging (LONI). We are responsible for taking that data, describing it, putting it into databases and making it available for other people to use. The net result has been over 1,000 manuscripts, with many papers written by scientists not directly involved with the project. It showed how valuable well-characterized data can be and set the stage for future data sharing in the AD community. Since then, the project has received additional funding and remains highly productive and impactful.

Another example of successful multi-site data sharing is the National Alzheimer's Coordinating Center (NACC). NACC uses a different model, in which AD centers in the US deliver some of their data to NACC, which then coordinates, aggregates and releases data to those who request it.[3] It has a somewhat more rigorous approval process for those requesting data.

The data sharing efforts in the AD research community have continued to evolve. More recently, three initiatives provide good examples of the next stage. The Global Alzheimer's Association

Interactive Network (GAAIN – http://gaain.org/) launched in 2014. This system includes all kinds of data—clinical, imaging, genetics and bio-sample data—even those collected using different methods, i.e., not necessarily prospective but also retrospective data.[1] The Dementia Platform UK (DPUK) (https://portal.dementiasplatform.uk/) is similar to GAAIN, but with different functionality. The last one, which is being promoted in the US, is fairly new and is called the Accelerating Medicines Partnership (AMP) (www.nia.nih.gov/research/amp-ad). This effort focuses on proteomics, metabolomics and genetics data, which have specific requirements.[4]

All of these initiatives show that the AD community has been among the most aggressive in coordinating data sharing to maximize scientific utility and increase the pace of discovery—partly because the challenges are so great in this field, and also because the AD community was sociologically ready for such collaborative efforts.

Q. Why is data sharing not universal across Alzheimer's disease research?

There are a number of reasonable justifications—the same reasons that much biomedical research is not shared openly. One is a practical concern regarding patient anonymity: health information needs to be protected. In some cases, patients with AD can be easy to identify from a group because of their age or genetic profile.

Another reason for lack of data sharing is that data are often collected as part of clinical trials. If the US Food and Drug Administration (FDA) will be using the results to evaluate the effectiveness of drugs, sharing restrictions may exist. These restrictions are intended to prevent data tampering and to ensure that experiments are conducted correctly, analyzed in a statistically appropriate manner, and are generally free of bias. For that reason, some of the data cannot be shared because it introduces the possibility of manipulation. And in some cases, pharmaceutical companies hesitate to share data because of intellectual property concerns.

Other reasons for the unwillingness to share data are sociological. Scientists may not yet have written any publications and fear being scooped. While this rarely happens, it is a common fear. Investigators are by nature competitive, and competition breeds excellence. But that can sometimes run counter to open data sharing. In addition, people want credit for the work they have done—the data collection process can take years and researchers often invest a great deal of their career in a single project.

Finally, there are technical reasons. The data may not have been quality controlled or may have errors, which could lead to incorrect conclusions. In addition, researchers might not have the technology or resources, such as database systems and computational infrastructure, to share data. In the case of genetic or imaging studies, these data often require many terabytes of storage.

Q. Could you tell us a little about the aims and design of the Global Alzheimer's Association Interactive Network (GAAIN)?

GAAIN is an open-access federated data platform that allows Alzheimer's researchers to conduct preliminary analyses with thousands of subjects and connect with one another to share data. Its primary goal is to help researchers discover clinical, genetic, imaging and other data to fuel their AD analyses.

The ADNI Informatics Core developed the technology, crafted the policy and laid the groundwork for GAAIN. It became clear that there were many other data archives that could further amplify the impact of ADNI. We realized that if we could combine data collected across different studies and sufficiently harmonize it, we could achieve enormous statistical power. These data might be sensitive to smaller variations due to their vast quantity, enabling us to use the data in new and unexpected ways.

From the start, we knew it was essential to respect the sociological needs of all the players, i.e., people's willingness to share the data. We wanted to develop something that didn't trigger concerns about credit and losing control over their data, and felt that a centralized approach where we housed all data would not be acceptable to many investigators. Therefore, we developed a system that is federated—or distributed—the data still remain at their respective archives. The GAAIN system communicates with these local archives, and although the data is presented through GAAIN, it still resides in these archives. The user can examine these data and even perform analyses on them, but they don't get the data itself unless the owner gives permission.

Our technology is designed to be intuitive and interactive, while communicating with individual archives. This system makes it safe and easy for scientists to share data—an important aspect of GAAIN—because without partner collaboration and participation, we could not be successful. This model has proved effective in recruiting more than 40 partners sharing data from over 475,000 subjects.

Q. What have early experiences of GAAIN taught us in terms of using this approach for other neurological diseases?

First, the community must be ready. The community needs to acknowledge that data sharing can be beneficial to all, including the people that collect the data. The way in which the data are shared requires a critical mass. There must be enough investigators interested in doing this, enough data to make it worthwhile and a culture that is ready to embrace the practice. Funders and grantees must be willing to co-operate and, most importantly, commit. The grantees might be less enthusiastic than the funders. The technology is also important; it must be intuitive and present no barriers for users.

When researchers collaborate on this level, new and highly impactful insights become possible. An early example of this is a reexamination of the link between age, gender and Alzheimer's risk, coauthored by the GAAIN team and published last year in *JAMA Neurology*,[5] which has already helped us better understand the disease.

Q. How is GAAIN likely to evolve in the coming years?

The collection of free statistical tools available through GAAIN can be further expanded so that users can perform even more sophisticated statistical analyses. We will likely use GAAIN to link other similar efforts. In addition, we plan to migrate towards a more social network where scientists can communicate among themselves, comment on recent findings and publications and replicate analyses that may have been done in one paper using different data. This provides a powerful strategy to determine how consistent the results are in different data sets or using different methods. ❑

1.   Toga AW, Neu SC, Bhatt P, et al. The Global Alzheimer's Association Interactive Network. *Alzheimers Dement*. 2016;12:49–54.
2.   Hendrix JA, Finger B, Weiner MW, et al. The worldwide Alzheimer's Disease Neuroimaging Initiative: an update. *Alzheimers Dement*. 2015;11:850–9.
3.   Beekly DL, Ramos EM, van Belle G, et al. The National Alzheimer's Coordinating Center (NACC) Database: an Alzheimer disease database. *Alzheimer Dis Assoc Disord*. 2004;18:270–7.
4.   Hodes RJ, Buckholtz N. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. *Expert Opin Ther Targets*. 2016;20:389–91.
5.   Neu SC, Pa J, Kukull W, et al. Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. *JAMA Neurol*. 2017;74:1178–89.