

Human neuroimaging as a “Big Data” science

John Darrell Van Horn · Arthur W. Toga

Published online: 10 October 2013
© Springer Science+Business Media New York 2013

Abstract The maturation of in vivo neuroimaging has led to incredible quantities of digital information about the human brain. While much is made of the data deluge in science, neuroimaging represents the leading edge of this onslaught of “big data”. A range of neuroimaging databasing approaches has streamlined the transmission, storage, and dissemination of data from such brain imaging studies. Yet few, if any, common solutions exist to support the science of neuroimaging. In this article, we discuss how modern neuroimaging research represents a multifactorial and broad ranging data challenge, involving the growing size of the data being acquired; sociological and logistical sharing issues; infrastructural challenges for multi-site, multi-datatype archiving; and the means by which to explore and mine these data. As neuroimaging advances further, e.g. aging, genetics, and age-related disease, new vision is needed to manage and process this information while marshalling of these resources into novel results. Thus, “big data” can become “big” brain science.

Keywords Neuroimaginggenetics · Big data · Informatics · Data sharing · Databases · Multi-site trials · International collaboration

Introduction

Images of the human brain, in form and function, seem to be everywhere these days—on television, in glossy magazines, and on internet blogs worldwide. This is due, in many

respects, to the incredible amount of information these images present and the sheer number of brain imaging research studies being performed to spy on the brain in action or at rest, to examine how it is built and wired, and what happens when things go wrong. Indeed, neuroimagers routinely collect more study data in a few days than was collected in over an entire year just a decade ago. These data are a rich source of information on detailed brain anatomy, the subtle variations in brain activity in response to cognitive stimuli, and complex patterns of inter-regional communication. Taken individually, these various data types would have once formed the basis for entire research programs. Now, with interests not only in multi-modal neuroimaging but the inclusion of co-occurring biological and clinical variable collection requiring linkage between geographically distributed researchers, neuroscience programs are rapidly becoming the brain-focused versions of projects more akin to those involving particle physics. The methods by which these data are obtained are themselves contributing to this growth, involving finer spatial and temporal resolution as MR physicists push the limits of what is possible and as brain scientists then rush to meet those limits. It is safe to say that human neuroimaging is now, officially, a “big data” science.

Such examples of large-data, their promise and challenges, have not gone unnoticed. In the US, The National Science Foundation, the National Institutes of Health, the Defense Department, the Energy Department, the Homeland Security Department as well as the U.S. Geological Survey have all made commitments toward “big data” programs. The Obama Administration itself has even gotten in on the act. In response to recommendations from the President’s Council of Advisors on Science and Technology, the White House sponsored a meeting bringing together a cross-agency committee to lay out specific actions agencies should take to coordinate and expand the government’s investment in “big data”, totaling \$200 million in support (see http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final.pdf).

J. D. Van Horn (✉) · A. W. Toga
The Institute for Neuroimaging and Informatics, Keck School of
Medicine of USC, University of Southern California, 2001 North
Soto Street—Room 102, MC 9232, Los Angeles, CA 90089-9235,
USA
e-mail: jvanhorn@usc.edu

A. W. Toga
e-mail: toga@usc.edu

Among the examples of “big data” featured at the meeting was—no surprise—human neuroimaging. Additionally, the recent announcement of the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative (<http://nih.gov/science/brain/index.htm>) forms part of a new Presidential focus aimed at revolutionizing understanding of the human brain. Initiatives surrounding large-scale brain mapping are also underway in Europe (<http://www.humanbrainproject.eu/>; Frisoni 2010) and examples of large-scale brain data sets have been on full display at recent annual meetings of the *Organization for Human Brain Mapping* (OHBM; <http://www.humanbrainmapping.org>) in Beijing, China in 2012 and Seattle, Washington in June 2013.

However, as the richness of brain data sets continues to grow and the push to place it in accessible repositories mounts, there are many issues to be considered on how to handle the data, move it from place to place, and how to store it, analyze it, and share it.

How big is “Big”?

While size is a relative term when it comes to data, medical imaging applied to the brain comes in a variety of forms, each generating differing types and amounts of information about neural structure and/or function. Moreover, in vivo neuroimaging is not unimodal but, rather, remarkably diverse—examining brain form, function, and connectivity, and rapidly improving its ability to resolve finer spatio-temporal scales. Further advancements in magnetic resonance scanner field and gradient strength are an ongoing area of research (Barry et al. 2011). A brief look at the history of some central parameters for specific forms of functional and structural neuroimaging is illustrative.

Since its inception, the collection of blood oxygenation level dependent (BOLD) functional time series has often been among the data types having the biggest storage footprint. While in the 1990’s, the earliest BOLD imaging studies used sampling intervals of one volume every 4 s for a modest number of slice planes needed to image the full brain. During the mid-2000’s, with advances in multi-channel coil technologies and the emergence of 3 T imaging systems, 2 s intervals were the most frequently used for the same in-plane image size but having finer slice resolution. More recently, improved sampling methods have made it possible to routinely obtain multiple BOLD image volumes of the whole head per second (Feinberg et al. 2010).

Likewise, diffusion tensor imaging (DTI) has undergone its own steady progress both in the time to acquire data but also the degree of angular resolution by which to measure patterns of water molecule diffusion. Early diffusion imaging studies measured diffusion along 6 directions seeking to utilize diffusion signal to infer neural fiber orientation, quantifying this as

a diffusion tensor, and to then link these tensor orientation patterns to estimate neural fiber pathways (Basser et al. 2000). Interest in examining finer degrees of angular difference in fiber orientations and to detangle crossing fiber pathways led to using more diffusion directions through high angular diffusion approaches having 32, 64, or greater MR gradient directions (Zhan et al. 2010). More recent approaches have decomposed the spectrum of diffusion which is capable of then resolving upwards of 512 fiber directions in approximately the same time and volume sizes as the original DTI imaging sequences (Wedeen et al. 2008).

In each example, as the imaging technology has improved, been extended, or made faster, so too has the amount of brain data which can be obtained. Once these improved methodologies have proven robust and dependable, with analytic methods available with which to utilize them, researchers have been quick to adopt them—doubling or tripling the amount of data they can then gather per subject by doing so. Indeed, a simple examination of fMRI articles from representative issues of the field’s touchstone journal, *NeuroImage*, indicates that since 1995 the amount of data collected has doubled approximately every 26 months (Fig. 1). At this rate, by 2015 the amount of acquired neuroimaging data alone, discounting header information and before more files are generated during data processing and statistical analysis, may exceed an average of 20GB per published research study. This is likely to be an under-estimate for raw dataset sizes since, as noted above, advances in MRI physics are accelerating the pace at which data can be acquired per unit time.

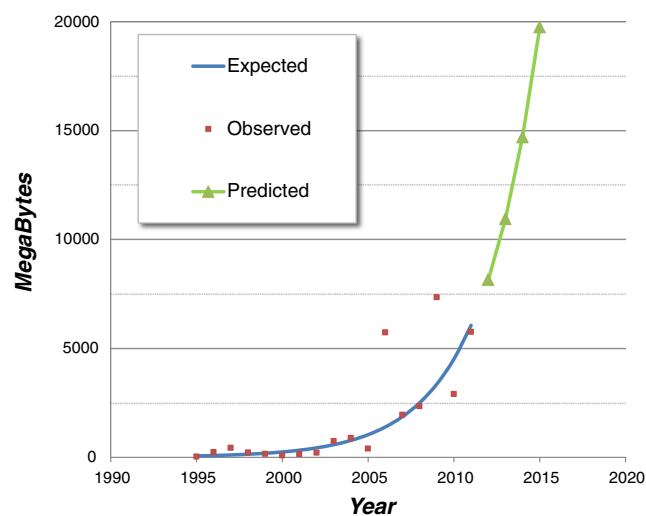


Fig. 1 The amount of acquired neuroimaging data reported from published articles in representative issues of the journal *NeuroImage* has doubled every 26 months and can expect to top 20GB of purely raw data on average per study in only a few years. Amassing, curating, storing, and sharing of such data from neuroimaging archives presents a growing big data challenge

Individually, the neuroimaging data sets from a single study themselves may not pose major difficulties for processing and analysis using existing algorithms and statistical methods. However, as data sets are amassed into large-scale databases, considerable challenges emerge. How to rectify between study differences in acquisition rates, resolution, scanning parameters, the presence of image artifacts, etc., take on importance—even the same make and model scanner installed at two different sites can generate sufficient differences in image quality to complicate combined analysis. Along with the increasing interest in gathering study data from across the lifespan or focused on specific patient groups, concurrently recorded electrophysiological time courses, and comprehensive phenomic meta-data on each study participant, wrangling, let alone interpreting, these data will not be for the faint of heart.

Big neuroimaging + big genetics = REALLY big data

With the ability to obtain genome-wide sets of single nucleotide polymorphism (SNP) information becoming routine and the costs of full genomic sequencing rapidly becoming affordable, interest in linking the influence of genes on the brain is heating up. Soon, exomic-level genetics will be readily at hand to replace SNP-based approaches. At several gigabytes per genome using Next Generation Sequencing (NGS) methods, for major brain imaging studies such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Weiner et al. 2012), with its initially available sample of 832 subjects, one can expect to store multiple petabytes of genetics data alone. More ADNI data are on their way, too. Coupled with the existing and ongoing collection of multimodal neuroimaging data types, this represents some really big data. As the bond between neuroimaging and genomics grows tighter, with both areas growing at incredible rates, disk storage, unique data compression techniques (like those proposed for genomics (Hsi-Yang Fritz et al. 2011); see also <http://www.sequencesqueeze.org/>) and data processing considerations rapidly become a scientific imperative.

Multisite consortia and data sharing

Along with advances in the ability to obtain data there has been an increase in the numbers of multisite consortia like ADNI for examining the healthy and diseased brain. The demand for multiscale data in the investigation of fundamental disease processes has been recognized for several years (Jiang et al. 2008), as well as the need for cooperation across centers and even disciplines to integrate and interpret the data (Van Horn and Toga 2009b). Examples of multisite neuroimaging efforts can be found in the ubiquitous application of neuroimaging in health but also in devastating illnesses such as Parkinson's

(Evangelou et al. 2009; see also <http://ppmi.loni.ucla.edu/>), psychiatric disorders (Schumann et al. 2010), and also the mapping of human brain connectivity (Toga et al. 2012). In addition to databases of aging and aging-related diseases, large-scale examples from the NIH-based National Database of Autism Research (NDAR; Hall et al. 2012) and the Federal Interagency Traumatic Brain Injury Research (FITBIR; Bushnik and Gordon 2012) exist to gather neuroimaging, genetic, and phenomic data on autism and brain injury, respectively. The various “grass roots” collections of resting-state fMRI data maintained as part of the “1000 Functional Connectomes” project (http://fcon_1000.projects.nitrc.org/) (see Biswal et al. 2010) and task-based OpenfMRI (<http://www.openfmri.org>) (Poldrack et al. 2013) are other notable examples.

What is more, pressures to share these data as openly as possible have put the onus on both the data collectors and database curators to store data efficiently and safely while providing data efficiently to anyone who would like to use it. However, inherent in multilaboratory projects are sociologic, legal, and often ethical concerns that must be resolved satisfactorily before they can work effectively or be widely accepted by the scientific community (Beaulieu 2001). While this often involves some expectations on the part of data collectors, etc., the exchange, storage, and computation on brain imaging data has many advantages including 1) maximizing the cost of their collection over the widest possible set of researchers, 2) allowing their use in new methods development, 3) promoting data re-analysis and re-purposing, 3) new means for collaboration, 4) hypothesis generation, and 5) clever forms of visualization. In other words, archived data can be subjected to novel approaches which can highlight relationships not envisioned by the original data collectors and shed new light on important mechanisms which are then worthy of additional study (Van Horn and Ishai 2007).

Examples of big neuroimaging databases

Despite the ever present challenges of archiving massive quantities of large brain imaging data, prominent examples for storing, sorting, and mining data exist. Key examples include XNAT Central (<https://central.xnat.org/>) (see Marcus et al. 2007) which contains data from several thousand subjects and is playing a key role in the data management and informatics of the Human Connectome Project (HCP) (Marcus et al. 2013); the SumsDB effort (<http://sumsdb.wustl.edu/sums/index.jsp>) for cortical surface-based atlasing of neuroimaging results (Van Essen 2005); and the NIH MRI Study of Normal Human Brain Development (http://www.bic.mcgill.ca/nihpd/info/data_access.html) (Evans 2006) containing multisite neuroimaging data of children from ages 6 to 18 years old accompanied by comprehensive neurological

assessments and neuropsychological testing results (Almli et al. 2007). Such resources represent leading efforts among a growing set of online repositories for the storage and sharing of raw and processed neuroimaging data and results.

Speaking from our own experience, the Laboratory of Neuro Imaging (LONI), formerly based at the University of California Los Angeles (UCLA) and now based at the University of Southern California (USC), has been serving as a repository for single and multi-site neuroimaging research studies for many years (Toga 2002b). The LONI Image and Data Archive (IDA) provides a comprehensive and interactive environment for safely archiving, querying, visualizing and sharing neuroimaging clinical and neurocognitive data. It presents a user-friendly environment to users for archiving, searching, sharing, tracking and disseminating neuroimaging data. All data is stored on redundant servers with daily and weekly on- and off-site backups. The IDA stores data from ADNI and the Michael J. Fox Foundation, among other prominent multi-site neuroimaging initiatives.

Archiving data in the IDA is straight-forward, secure, and requires no specialized hardware or software. The IDA automatically facilitates the de-identification and pooling of data from multiple institutions, protecting data from unauthorized access while providing the ability to share data among collaborative investigators. Integration of the LONI Debabeler file format translation engine (Neu et al. 2005) allows users to upload and download image data in a number of common neuroimaging file formats (e.g. DICOM, Analyze, NIFTI, MINC). Once archived, data can be downloaded and/or streamed into automated tools for processing and analysis (Dinov et al. 2009, 2010a).

Presently, the IDA repository contains over 14,000 imaging series for thousands of subjects, with growth averaging ~400 new image series per month, encompassing 130 TB of storage. This includes structural, functional, diffusion, and other MRI-based data types, in addition to studies of a range of PET ligands. Since the inception of the IDA, data drawn from it has been utilized in literally hundreds of research articles published in the peer-reviewed literature.

All in all, resources such as these are amassing brain imaging data at already impressive scales. However, will such archives be ready for still more data as investigators acquire more information through scanning as well as other measures of brain form and function? The next wave for neuroimaging genetic examinations of the brain will certainly test the computational and storage infrastructures of these and other databasing efforts.

The role of cyberinfrastructure

Individual desktop computers are now no longer suitable for analyzing potentially petabytes worth of brain and genomics

data at a time. What is needed is to combine effort from across multiple, distributed processing elements and leverage the combined power they possess toward massive-scale analyses of neuroscientific resources. The motivating interest in emerging forms of computing for biomedicine is the coordination of resource sharing and problem solving in dynamic, multi-institutional, spatially dispersed, virtual organizations who can gather and exchange data. While the National Science Foundation (NSF) has made major investments in the computer architecture needed for physics, weather, and geological data (e.g. XSEDE, <https://www.xsede.org/>, and the Open Science Grid, <https://www.opensciencegrid.org>), the NIH has invested warily (e.g. BIRN, <http://www.birncommunity.org>), despite strongly encouraging data re-use from online data repositories. More than just the sharing of files, this would involve direct access to computer resources, software, data, and services, increasingly required by a range of collaborative problem-solving and resource-brokering strategies emerging from industry, science, and engineering. The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC; <http://www.nitrc.org>) and the International Neuroinformatics Coordinating Facility (INCF; <http://incf.org>) have begun to deploy local clusters with Amazon EC2 server technology toward this goal but a larger effort will be required involving dedicated processing centers or distributed grids of linked computer centers. Resource availability can be open or highly controlled with providers and consumers defining clearly and carefully just what is shared, who is allowed to use resources, and the conditions under which it occurs (Van Horn and Toga 2009a). It shouldn't come as a surprise that, like so many other forms of data-rich science, big computing is an ideal framework for big neuroimaging.

Models for data management and availability

The organizing, annotating, archiving, and distributing of neuroimaging and biomedical data in useable and structured frameworks have become critical elements across a range of neuroscientific efforts (Toga 2002a). With all this brain data flying around needing someplace to land, several well known efforts to construct and populate large brain anatomy (Mazziotta et al. 2001), function (Van Horn et al. 2005), and genetics (Saykin et al. 2010) databases have arisen over the years in order to make such data open to a still broader audience of researchers. In fact, for many recent large-scale neuroimaging projects, such as the Human Connectome Project (HCP; www.humanconnectomeproject.org), their existence is predicated on the expectation that the data obtained will be well organized and available to the community for examination and study.

As one might imagine, there are as many data organizational models as there are laboratories gathering the data.

Because of the volume of all the study data itself, it is not uncommon for the imaging data to live apart from the meta-data describing it, e.g. on its own disk partition or even another computer system altogether. This information, in turn, can itself live apart from other data acquired (e.g. demographics, genetics, phenomics). Organizing and linking all of these distinct and possibly physically separated data types can be a major data management challenge.

What is more, multiple models exist for making data available to others (Van Horn et al. 2001). In the simplest form, anonymous FTP sites can be used as a place to post datasets where they may or may not be secure and the intentions of those accessing the data uncertain. Databases containing summary results are enormously valuable from a meta-analytic perspective (Laird et al. 2009), however, the original neuroimaging data themselves are likely unavailable. The use of results-only resources may also require care in due to the potential for factors such as publication bias (Jennings and Van Horn 2012). Federated archiving approaches seek to allow data to remain at the sites that collected it but to permit users at a consortia of participating institutions to search and access each other's data (Helmer et al. 2011). Centralized approaches also exist in which a single center is the focal point for databasing and accompanying informatics for multi-site efforts or multiple-multi-site efforts (Neu et al. 2012). These and other models have their own unique challenges for data storage, management, and availability which can be expected to become more complex as study data sizes grow. Contributions of raw and processed study data to centralized or federated repositories can provide maximal information and utility to subsequent users (Van Horn and Gazzaniga 2005), where the submission of data can be voluntary, as a condition of publication, or as an agreed upon aspect of multi-site data consortia (Mazziotta et al. 1995; Toga and Crawford 2010; Van Horn and Gazzaniga 2012).

Standards are frequently non-standard

With a diversity of storage models also comes a variety of ways in which study data is managed and meta-data is internally represented. Having common formatting standards for meta-data representation and organization has been an important topic for neuroinformatics over the past decade (Koslow 2000; Helmer et al. 2011). The creation of data standards enables both intra- and inter-disciplinary interaction (Van Horn and Ball 2008), encourages the development of novel software tools for helping to understand relationships within and among data elements, and encourages new investigation of database contents (Neu et al. 2012). Recent efforts from international data sharing working groups (Poline et al. 2012) have made considerable headway in helping to develop new frameworks for organizing study meta-data drawing from the

best parts of extant meta-data frameworks. Best practices for fMRI results reporting have been recommended (Poldrack et al. 2008) though may naturally need to mature further as emerging analytic approaches come into favor. On the other hand, standards designers need to avoid the feature creep that often arises which can result in frameworks which are overly rigid and cannot adapt as new information becomes available and need to be included. Nevertheless, when fully mature, these modern schema will significantly improve the description of neuroimaging data sets, encode the provenance associated with data processing (Mackenzie-Graham et al. 2008; McClatchey et al. 2013), and help to populate large-scale archives prospectively, thereby encouraging common analysis frameworks. Such practical standards will be essential for multi-site trials and major neuroimaging initiatives, where data sharing has been expressly mandated by funding agencies.

Factors governing the utility of “Big data” resources

However, simply having “big data” neuroimaging and making it available online is not a means to an end but only the next step in making that data available for others to explore and mine. Several factors contribute to a database's use and utility, including whether it actually contains viable data accompanied by a detailed description of their acquisition (Van Horn and Toga 2009a); whether the database is well-organized and the user interface is easy to navigate; whether the data are derived versions of raw data or the raw data itself; the manner in which the database addresses the sociological and regulatory issues that can be associated with data sharing; whether the data is fully anonymized and open to anyone who wants it, only open to members of a multi-site consortium, or if it is necessary to get IRB approval before access can be granted; whether it has a policy in place to ensure that requesting authors give proper attribution to the original collectors of the data; and the efficiency of secure data transactions. The systems must provide flexible methods for data description and relationships among various meta-data characteristics (Bug et al. 2008). Moreover, those that have been specifically designed to serve a large and diverse audience with a variety of needs, represent the types of databases that can have the greatest benefit to scientists looking to study the disease, assess new methods, examine previously published data, or with interests in exploring new ideas (Van Horn and Gazzaniga 2005; Keator et al. 2013).

Mining data and digging for gold

The successes of molecular biological (Huang et al. 2007), systems biological (Hood et al. 2004), and astrophysics (Gray

et al. 2002) infrastructures for archiving and mining data are well known. With the accumulation of neurobiological data into large databases and the availability of compute cluster enabled means for large-scale data processing, a new form of discovery-oriented neuroscience is on the horizon. Researchers are increasingly mining vast and disparate collections of data and hunting for unseen patterns that might provide clues to underlying biological mechanisms (Phan et al. 2002; Arnone et al. 2009; Frazier and Hardan 2009) and trends in how studies are conducted (Jennings and Van Horn 2012). The terabytes-worth of data provide input for informatics-driven pattern-seeking and other relevant algorithms (Jones and Swindells 2002; Ma et al. 2002; Schutte et al. 2002) which can provide further understanding into complex brain processes.

Departing from a purely repository-based approach, a collection of research groups from around the world has adopted what might be considered more of a social networking strategy for data sharing. The *Enhancing Neuroimaging Genetics through Meta-Analysis* (ENIGMA; enigma.ionu.edu) network brings together interested researchers in imaging genomics, to understand brain structure and function, based on MRI, DTI, fMRI and genome-wide association study (GWAS) data. Among the network's goals: to ensure promising findings are replicated via member collaborations, in order to satisfy the mandates of most journals; gain increased statistical power; and to share ideas, algorithms, data, and information on promising findings or methods. They have provided the means for imaging, genetics, as well as computer science, mathematics and statistics experts to make significant collaborative contributions to these fields from which, without the needed infrastructure, most of their expertise would have been excluded (Stein et al. 2012). These successes are not necessarily unique, but building upon them and extending them to a wider set of scientific research arenas is an ever-present theme (Persson 2000; Brookes 2001; Altman 2003).

New computer science designed with big data in mind

While it might be tempting to think that once all the data has been archived, indexed, and is ready to go, that all one would need is to start analyzing it and answers to all our questions about the brain will be revealed. Even examining the contents of an archive to know what data is available to be analyzed requires new, cleverly designed and user friendly software tools and novel approaches for exploratory inspection. Such tools are only now beginning to appear (Bowman et al. 2012) and their further development will be essential for dealing with existing as well as the expected size of neuroimaging data sets.

Once a selection of data worthy of further analysis has been identified, a new concern is realized—it becomes clear that

many software packages for neuroimaging data analysis are ill-suited toward very large data sets involving potentially thousands of subjects. Algorithm optimization is not often considered for when data sets are small or modest in size but as data sets grow memory management is an important factor. New mathematics and informatics approaches will be needed to more completely model multi-modal brain imaging data in the context of cortical anatomy, white matter connectivity, and functional activity. These will need to work fast, be accurate, and be interoperable with other tools so that data processing can be automated as much as possible. Interactive workflow environments for automated data analysis will also be critical for ongoing or retrospective research studies involving complex computations on large multi-dimensional datasets (Dinov et al. 2010b; Gorgolewski et al. 2011). Yet, few tools, if any, now exist which enable the joint analysis of both genes and brain imaging data which would be capable of efficiently obtaining results while also achieving the requisite degree of statistical power. Moving forward, software engineers will need to create brilliant and innovative ways to tackle the massive amounts of brain and genomic data. These continuous interactions between neuroimagers, geneticists, software creators, and other biomedical scientists will be essential to develop these new, memory-efficient software algorithms and computational tools.

Big data in the era of ADNI2, the HCP, the HBP, and BRAIN

As the NIH and other major funding agencies spearhead big picture multi-center brain science efforts, the needs for big data solutions will grow. The HCP consortia are now releasing large datasets which even the best neuroimaging researchers are struggling to analyze. More of that data is on the way. ADNI2, the premier neuroimaging data collection effort for understanding the aging brain, will be providing imaging, phenomic, and exomic information which we have yet to get our heads around how best to analyze. The HBP, involving data collection sites throughout Europe, will likely eclipse that! And while the basis for infrastructure support is also changing, the simple thought that data and processing will simply be done “in the cloud” is somewhat naïve. The “cloud” must exist somewhere and system failures have been known to happen. Though commercial solutions such as those offered by Google, Amazon, and Microsoft can provide an attractive solution for individual centers, dedicated neuroscience computational databasing and data processing resources remain essential. As the BRAIN initiative takes shape, it is clear that demonstrable access to data, processing workflow methods, running on remote CPU clusters dedicated to this purpose is the best way to ensure that we can keep up in the extraction of real findings as opposed to a collection of results.

Conclusions

Neuroimaging research, by its very nature, is data intensive, multimodal, and collaborative—factors which have been instrumental in its success and growth. Indeed, we contend that neuroimaging is an emerging example of discovery-oriented science, wherein patterns of brain structure and activity present across multiple subjects and dozens of studies can be systematically extracted, examined, and result in new knowledge. Yet the infrastructure needed for supporting this advancing form of brain research where data is king is still maturing. The rapid processing of large quantities of data in this way will lead to new scientific outcomes and patterns of results not envisioned during the examination of each study individually. Patterns may suggest fundamental mechanisms. Confirmed mechanisms add to the knowledgebase of neurobiological science and provide the basis for further experimentation and the generation of still more valuable data that can be included in still greater analyses. Greater knowledge about fundamental brain processes then suggests new and testable hypotheses that lead to novel experimentation, the data from which should then be contributed back into a publicly available archive—continuing a healthy and helpful cycle.

The next steps for the development of resources supporting “big data” brain imaging at the Exabyte scale will require the further creation of new tools and services for data discovery, integration, analysis, and visualization. Components for discovering data residing in database architectures must be developed (sort of a PubMed for data discovery). Present examples exist such as the “EB-eye” resource for genomics (<http://www.ebi.ac.uk/ebisearch/>) or the Neuroscience Information Framework (NIF; <http://neuinfo.org>) for neuroscience terminologies. Such meta-resources will need to include contextual information that allows data to be accessed, understood, reused, and the results reproduced. Integrating a broader spectrum of neuroscience data and providing tools for interrogating and visualizing those data will enable investigators to more easily and interactively investigate broader scientific questions.

Beyond just neuroimaging data, architectures for peta-scale biomedical data must be flexible enough to allow integration of additional clinical and biochemical data and analysis results into the database and employ tools for interactively interrogating and graphically visualizing database contents (Bowman et al. 2012). Frameworks for storing and making available big neuroimaging data, their standards, and the infrastructure for so doing must be enriched with modern data processing workflow design and execution systems in place to permit exchange of processing methodologies between labs, among consortia members, or by independent researchers. Comprehensive mechanisms to gather, organize, and distribute data, results, and information between and among project participants, but also to the scientific community at large, are worth being

examined, developed, and deployed. For the “big data” science of human brain imaging, now is the time to begin.

Acknowledgments This article was supported by a P41 (RR013642) award from the National Institute for General Medicine (NIGMS), through the National Institutes of Health Roadmap for Medical Research (U54 RR021813, The Center for Computational Biology—CCB), and through an National Institute of Mental Health ARRA award (RC1 MH088194). We express our gratitude the faculty and staff of the Laboratory of Neuro Imaging (LONI) at the University of California Los Angeles.

References

- Almli, C. R., Rivkin, M. J., & McKinstry, R. C. (2007). The NIH MRI study of normal brain development (Objective-2): newborns, infants, toddlers, and preschoolers. *NeuroImage*, *35*(1), 308–325.
- Altman, R. B. (2003). The expanding scope of bioinformatics: sequence analysis and beyond. *Heredity*, *90*(5), 345.
- Arnone, D., Cavanagh, J., Gerber, D., Lawrie, S. M., Ebmeier, K. P., & McIntosh, A. M. (2009). Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. *The British Journal of Psychiatry*, *195*(3), 194–201.
- Barry, R. L., Strother, S. C., Gatenby, J. C., & Gore, J. C. (2011). Data-driven optimization and evaluation of 2D EPI and 3D PRESTO for BOLD fMRI at 7 Tesla: I. Focal coverage. *NeuroImage*, *55*(3), 1034–1043. PMID: 3049844.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, *44*(4), 625–632.
- Beaulieu, A. (2001). Voxels in the brain: neuroscience, informatics and changing notions of objectivity. *Social Studies of Science*, *31*(5), 635–680.
- Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kottler, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., McMahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., G. J. S., Sorg, C., Teng, G. J., Vejjola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences USA*.
- Bowman, I., Joshi, S. H., & Van Horn, J. (2012). Visual systems for interactive exploration and mining of large-scale neuroimaging data archives. *Frontiers in Neuroinformatics*, *6*.
- Brookes, A. J. (2001). Rethinking genetic strategies to study complex diseases. *Trends in Molecular Medicine*, *7*(11), 512–516.
- Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., Larson, S. D., Rubin, D., Shepherd, G. M., Turner, J. A., & Martone, M. E. (2008). The NIFSTD and BIRN Lex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics*, *6*(3), 175–194.
- Bushnik, T., & Gordon, W. (2012). Updates from the Third Federal Interagency Conference on traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *27*(3), 222–223.
- Dinov, I., Van Horn, J., Lozev, K., Magsipoc, R., Petrosyan, P., Liu, Z., MacKenzie-Graham, A., Eggert, P., Parker, D., & Toga, A. (2009). Efficient, distributed and interactive neuroimaging data analysis

- using the LONI pipeline. *Frontiers of Neuroinformatics*, 3(22), 1–10.
- Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J., Parker, D. S., Magsipoc, R., Leung, K., Gutman, B., Woods, R., Toga, A. (2010). Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE*, 5(9), PMID: 2946935.
- Dinov, I., Van Horn, J., Lozev, K., Magsipoc, R., Petrosyan, P., Liu, Z., MacKenzie-Graham, A., Eggert, P., Parker, D. S., & Toga, A. W. (2010b). Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. *Frontiers in Neuroinformatics*, 3(22), 1–10.
- Evangelou, E., Maraganore, D. M., Annesi, G., Brighina, L., Brice, A., Elbaz, A., Ferrarese, C., Hadjigeorgiou, G. M., Krueger, R., Lambert, J. C., Lesage, S., Markopoulou, K., Mellick, G. D., Meeus, B., Pedersen, N. L., Quattrone, A., Van Broeckhoven, C., Sharma, M., Silburn, P. A., Tan, E. K., Wirdefeldt, K., Ioannidis, J. P. (2009). Non-replication of association for six polymorphisms from meta-analysis of genome-wide association studies of Parkinson's disease: Large-scale collaborative study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*.
- Evans, A. C. (2006). The NIH MRI study of normal brain development. *NeuroImage*, 30(1), 184–202.
- Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS ONE*, 5(12), e15710.
- Frazier, T. W., & Hardan, A. Y. (2009). A meta-analysis of the corpus callosum in autism. *Biological Psychiatry*, 66(10), 935–941.
- Frisoni, G. B. (2010). Alzheimer's disease neuroimaging initiative in Europe. *Alzheimer's & Dementia*, 6(3), 280–285.
- Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*, 5, 13, PMID: 3159964.
- Gray, J., Szalay, S., Thaker, A. R., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C., vanden Berg, J. (2002). The SDSS SkyServer—public access to the Sloan Digital Sky Surver Data. *ACM SIGMOD*.
- Hall, D., Huerta, M. F., McAuliffe, M. J., & Farber, G. K. (2012). Sharing heterogeneous data: the national database for autism research. *Neuroinformatics*, 10(4), 331–339.
- Helmer, K. G., Ambite, J. L., Ames, J., Ananthakrishnan, R., Burns, G., Chervenak, A. L., Foster, I., Liming, L., Keator, D., Macciardi, F., Madduri, R., Navarro, J. P., Potkin, S., Rosen, B., Ruffins, S., Schuler, R., Turner, J. A., Toga, A., Williams, C., Kesselman, C. (2011). Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). *Journal of the American Medical Informatics Association*.
- Hood, L., Heath, J. R., Phelps, M. E., & Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306(5696), 640–643.
- Hsi-Yang Fritz, M., Leinonen, R., Cochrane, G., & Birney, E. (2011). Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Research*, 21(5), 734–740. PMID: 3083090.
- Huang, H., Hu, Z. Z., Arighi, C. N., & Wu, C. H. (2007). Integration of bioinformatics resources for functional analysis of gene expression and proteomic data. *Frontiers in Bioscience*, 12, 5071–5088.
- Jennings, R. G., & Van Horn, J. D. (2012). Publication bias in neuroimaging research: implications for meta-analyses. *Neuroinformatics*, 10(1), 67–80.
- Jiang, T., Liu, Y., Shi, F., Shu, N., Liu, B., Jiang, J., & Zhou, Y. (2008). Multimodal magnetic resonance imaging for brain disorders: advances and perspectives. *Brain Imaging and Behavior*, 2(4), 249–257.
- Jones, D. T., & Swindells, M. B. (2002). Getting the most from PSI-BLAST. *Trends in Biochemical Sciences*, 27(3), 161–164.
- Keator, D. B., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G., Gadde, S., Ashish, N., Burns, G. A., Nichols, B. N. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *Neuroimage*.
- Koslow, S. H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*, 3(4), 863–865.
- Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., Robinson, J. L., Lancaster, J. L., & Fox, P. T. (2009). ALE meta-analysis workflows via the brainmap database: progress towards a probabilistic functional brain atlas. *Front Neuroinform*, 3, 23. PMID: 2715269.
- Ma, B., Tromp, J., & Li, M. (2002). PatternHunter: faster and more sensitive homology search. *Bioinformatics*, 18(3), 440–445.
- Mackenzie-Graham, A. J., Van Horn, J. D., Woods, R. P., Crawford, K. L., & Toga, A. W. (2008). Provenance in neuroimaging. *NeuroImage*, 42(1), 178–195. PMID: 2664747.
- Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007). The extensible neuroimaging archive toolkit (XNAT): an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, 5, 11–34.
- Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., Barch, D. M., Archie, K. A., Burgess, G. C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., Coalson, T., Schindler, J., Elam, J. S., Curtiss, S. W., Van Essen, D. C. (2013). Human connectome project informatics: Quality control, database services, and data visualization. *Neuroimage*.
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). *NeuroImage*, 2(2), 89–101.
- Mazziotta, J., Toga, A. W., Evans, A., Fox, P., Lancaster, J., Ziles, K., Woods, R. P., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P. M., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L. M., Narr, K., Kabani, N., LeGoualher, G., Boomsma, D., Cannon, T., Kawashima, R., & Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 356, 1293–1322.
- McClatchey, R., Branson, A., Anjum, A., Bloodworth, P., Habib, I., Muir, K., Shamdasani, J., Soomro, K. (2013). Providing traceability for neuroimaging analyses. *International Journal of Medical Informatics*.
- Neu, S. C., Valentino, D. J., & Toga, A. W. (2005). The LONI Debabeler: a mediator for neuroimaging software. *NeuroImage*, 24(4), 1170–1179.
- Neu, S. C., Crawford, K. L., & Toga, A. W. (2012). Practical management of heterogeneous neuroimaging metadata by global neuroimaging data repositories. *Front Neuroinform*, 6, 8. PMID: 3311229.
- Persson, B. (2000). Bioinformatics in protein analysis. *Exs*, 88, 215–231.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16(2), 331–348.
- Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *NeuroImage*, 40(2), 409–414. PMID: 2287206.
- Poldrack, R. A., Barch, D. M., Mitchell, J. P., Wager, T. D., Wagner, A. D., Devlin, J. T., Cumba, C., Koyejo, O., & Milham, M. P. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Front Neuroinform*, 7, 12. PMID: 3703526.
- Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., Haselgrove, C., Helmer, K. G., Keator, D. B., Marcus, D. S., Poldrack, R. A., Schwartz, Y., Ashburner, J., & Kennedy, D. N.

- (2012). Data sharing in neuroimaging research. *Front Neuroinform*, 6, 9. PMID: 3319918.
- Saykin, A. J., Shen, L., Foroud, T. M., Potkin, S. G., Swaminathan, S., Kim, S., Risacher, S. L., Nho, K., Huentelman, M. J., Craig, D. W., Thompson, P. M., Stein, J. L., Moore, J. H., Farrer, L. A., Green, R. C., Bertram, L., Jack, C. R., & Weiner, M. W. (2010). ADNI biomarkers as quantitative phenotypes: genetics core aims, progress, and plans, Alzheimer's and Dementia. *Journal of the Alzheimer's Association*, 6(3), 265–273.
- Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Buchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J. L., Paus, T., Poline, J. B., Robbins, T. W., Rietschel, M., Reed, L., Smolka, M., Spanagel, R., Speiser, C., Stephens, D. N., Strohle, A., & Struve, M. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, 15(12), 1128–1139.
- Schutte, B. C., Mitros, J. P., Bartlett, J. A., Walters, J. D., Jia, H. P., Welsh, M. J., Casavant, T. L., & McCray, P. B., Jr. (2002). Discovery of five conserved beta-defensin gene clusters using a computational search strategy. *Proceedings of the National Academy of Sciences of the United States of America*, 99(4), 2129–2133.
- Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., Winkler, A. M., Toro, R., Appel, K., Bartecek, R., Bergmann, O., Bernard, M., Brown, A. A., Cannon, D. M., Chakravarty, M. M., Christoforou, A., Domin, M., Grimm, O., Hollinshead, M., Holmes, A. J., Homuth, G., Hottenga, J. J., Langan, C., Lopez, L. M., Hansell, N. K., Hwang, K. S., Kim, S., Laje, G., Lee, P. H., Liu, X., Loth, E., Lourdasamy, A., Mattingsdal, M., Mohnke, S., Maniega, S. M., Nho, K., Nugent, A. C., O'Brien, C., Papmeyer, M., Putz, B., Ramasamy, A., Rasmussen, J., Rijpkema, M., Risacher, S. L., Roddey, J. C., Rose, E. J., Ryten, M., Shen, L., Sprooten, E., Strengman, E., Teumer, A., Trabzuni, D., Turner, J., van Eijk, K., van Erp, T. G., van Tol, M. J., Wittfeld, K., Wolf, C., Woudstra, S., Aleman, A., Alhusaini, S., Almasy, L., Binder, E. B., Brohawn, D. G., Cantor, R. M., Carless, M. A., Corvin, A., Czisch, M., Curran, J. E., Davies, G., de Almeida, M. A., Delanty, N., Depondt, C., Duggirala, R., Dyer, T. D., Erk, S., Fagermess, J., Fox, P. T., Freimer, N. B., Gill, M., Goring, H. H., Hagler, D. J., Hoehn, D., Holsboer, F., Hoogman, M., Hosten, N., Jahanshad, N., Johnson, M. P., Kasperaviciute, D., Kent, J. W., Jr., Kochunov, P., Lancaster, J. L., Lawrie, S. M., Liewald, D. C., Mandl, R., Matarin, M., Mattheisen, M., Meisenzahl, E., Melle, I., Moses, E. K., Muhleisen, T. W., Nauck, M., Nothen, M. M., Olvera, R. L., Pandolfo, M., Pike, G. B., Puls, R., Reinvang, I., Renteria, M. E., Rietschel, M., Roffman, J. L., Royle, N. A., Rujescu, D., Savitz, J., Schnack, H. G., Schnell, K., Seiferth, N., Smith, C., Steen, V. M., Valdes Hernandez, M. C., Van den Heuvel, M., van der Wee, N. J., Van Haren, N. E., Veltman, J. A., Volzke, H., Walker, R., Westlye, L. T., Whelan, C. D., Agartz, I., Boomsma, D. I., Cavalleri, G. L., Dale, A. M., Djurovic, S., Drevets, W. C., Hagoort, P., Hall, J., Heinz, A., Jack, C. R., Jr., Foroud, T. M., Le Hellard, S., Macciardi, F., Montgomery, G. W., Poline, J. B., Porteous, D. J., Sisodiya, S. M., Starr, J. M., Sussmann, J., Toga, A. W., Veltman, D. J., Walter, H., Weiner, M. W., Bis, J. C., Ikram, M. A., Smith, A. V., Gudnason, V., Tzourio, C., Vernooij, M. W., Launer, L. J., Decarli, C., Seshadri, S., Andreassen, O. A., Apostolova, L. G., Bastin, M. E., Blangero, J., Brunner, H. G., Buckner, R. L., Cichon, S., Coppola, G., de Zubicaray, G. I., Deary, I. J., Donohoe, G., de Geus, E. J., Espeseth, T., Fernandez, G., Glahn, D. C., Grabe, H. J., Hardy, J., Hulshoff Pol, H. E., Jenkinson, M., Kahn, R. S., McDonald, C., McIntosh, A. M., McMahon, F. J., McMahon, K. L., Meyer-Lindenberg, A., Morris, D. W., Muller-Myhsok, B., Nichols, T. E., Ophoff, R. A., Paus, T., Pausova, Z., Penninx, B. W., Potkin, S. G., Samann, P. G., Saykin, A. J., Schumann, G., Smoller, J. W., Wardlaw, J. M., Weale, M. E., Martin, N. G., Franke, B., Wright, M. J., & Thompson, P. M. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genetics*.
- Toga, A. W. (2002a). Imaging databases and neuroscience. *The Neuroscientist*, 8(5), 423–436.
- Toga, A. W. (2002b). The laboratory of neuro imaging: what it is, why it is, and how it came to be. *IEEE Transactions on Medical Imaging*, 21(11), 1333–1343.
- Toga, A. W., & Crawford, K. L. (2010). The informatics core of the Alzheimer's disease neuroimaging initiative. *Alzheimer's and Dementia*, 6(3), 247–256.
- Toga, A. W., Clark, K. A., Thompson, P. M., Shattuck, D. W., & Van Horn, J. D. (2012). Mapping the human connectome. *Neurosurgery*, 71(1), 1–5.
- Van Essen, D. C. (2005). A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *NeuroImage*, 28(3), 635–662.
- Van Horn, J. D., & Ball, C. A. (2008). Domain-specific data sharing in neuroscience: what do we have to learn from each other? *Neuroinformatics*, 6(2), 117–121.
- Van Horn, J. D., & Gazzaniga, M. S. (2005). In S. H. Koslow & A. Subramanian (Eds.), *Maximizing information content in shared neuroimaging studies of cognitive function. Databasing the brain: From data to knowledge*. New York: John Wiley and Sons.
- Van Horn, J. D., & Gazzaniga, M. S. (2012). Why share data? Lessons learned from the fMRIDC. *Neuroimage*.
- Van Horn, J. D., & Ishai, A. (2007). Mapping the human brain: new insights from FMRI data sharing. *Neuroinformatics*, 5(3), 146–153.
- Van Horn, J. D., & Toga, A. W. (2009a). Is it time to re-prioritize neuroimaging databases and digital repositories? *NeuroImage*, 47(4), 1720–1734.
- Van Horn, J. D., & Toga, A. W. (2009b). Multisite neuroimaging trials. *Current Opinion in Neurology*, 22(4), 370–378. PMID: 2777976.
- Van Horn, J. D., Grethe, J. S., Kostelec, P., Woodward, J. B., Aslam, J. A., Rus, D., Rockmore, D., & Gazzaniga, M. S. (2001). The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 356(1412), 1323–1339. PMID: 1088517.
- Van Horn, J. D., Wolfe, J., Agnoli, A., Woodward, J., Schmitt, M., Dobson, J., Schumacher, S., & Vance, B. (2005). Neuroimaging databases as a resource for scientific discovery. *International Review of Neurobiology*, 66, 55–87.
- Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., & de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *NeuroImage*, 41(4), 1267–1277.
- Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Siuciak, J. A., Soares, H., Toga, A. W., & Trojanowski, J. Q. (2012). The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, 8(1 Suppl), S1–S68. PMID: 3329969.
- Zhan, L., Leow, A. D., Jahanshad, N., Chiang, M. C., Barysheva, M., Lee, A. D., Toga, A. W., McMahon, K. L., de Zubicaray, G. I., Wright, M. J., & Thompson, P. M. (2010). How does angular resolution affect diffusion imaging measures? *NeuroImage*, 49(2), 1357–1371.