



Published in final edited form as:

Neuroimage. 2013 November 15; 82: . doi:10.1016/j.neuroimage.2012.11.010.

Why share data? Lessons learned from the fMRIDC

John Darrell Van Horn and

Laboratory of Neuro Imaging (LONI), Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, 635 Charles E. Young Drive SW, Suite 225 Los Angeles, CA 90095-7334. Phone: (310) 267-5156, Fax: (310) 206-5518. jvanhorn@loni.ucla.edu

Michael S. Gazzaniga

Director, Sage Center for the Study of Mind, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. Phone: (805) 893-5448, Fax: (805) 893 4303. m.gazzaniga@psych.ucsb.edu

Abstract

Neuroimaging and the discipline of cognitive neuroscience have grown together in lock-step with each pushing the other toward an improved ability to explore and examine brain function and form. However successful neuroimaging and the examination of cognitive processes may seem today, the culture of data sharing in these fields remains underdeveloped. In this article, we discuss our own experience in the development of the fMRI Data Center (fMRIDC) – a large-scale effort to gather, curate, and openly share the complete data sets from published research articles of brain activation studies using fMRI. We outline the fMRIDC effort's beginnings, how it operated, note some of the sociological reactions we received, and provide several examples of prominent new studies performed using data drawn from the archive. Finally, we provide comment on what considerations are needed for successful neuroimaging databasing and data sharing as existing and emerging efforts take next steps in archiving and disseminating the field's valuable and irreplaceable data.

Keywords

neuroimaging; databasing; data sharing; neuroinformatics

INTRODUCTION

While their seeds were sown independently, it is no surprise that the field of cognitive neuroscience has grown in lock-step with the development of neuroimaging methodologies for examining *in vivo* brain function. The ability to peer into the brain at work, initially using positron emission tomography (PET) and later with functional magnetic resonance imaging (fMRI), helped to provide a rigorous empirical framework for mapping functions to specific combinations of brain regions. With neuroimaging as its primary tool set, cognitive neuroscience has matured as a brain science discipline, establishing its own scientific societies with well-attended annual meetings, and its own peer-reviewed journals. Increasingly, cognitive neuroimaging experiments are an expected part of large-scale brain imaging initiatives and programs worldwide (Van Horn 2004).

At the same time, research in neuroimaging methods has made enormous technological advances which permit the exploration of changes in blood oxygenation level dependent (BOLD) signals at finer and finer temporal time-scales. Much of this improvement has been

driven by the quest to understand ever more subtle time-dependent changes in cognitively-driven brain activity. Sampling rates for fMRI have gone from early studies collecting images once every 4 seconds to, more recently, once every few hundred milliseconds (Feinberg, Moeller *et al.* 2010). With each technological advance in imaging – be it with fMRI, diffusion weighted imaging, etc – scientists wishing to understand the details of cognitive systems have pushed the edge of the envelope on what that technology is capable of. As they approach that limit, MRI physicists, engineers, and manufacturers soon unveil still further improved technologies, permitting greater spatiotemporal data to be acquired.

Despite the wide availability of such imaging technologies, their empirical use to map brain function, and the incredible amounts of data being acquired, there continue to be many misunderstandings about the results and the processes that they represent. Further insights could be achieved if more of the data were made accessible for others to explore (Koslow 2000). Unfortunately, neuroimage data sharing as an expected element of cognitive neuroscience has not been fully accepted and data, once obtained, is often only subjected to the analytic treatments of those who collected it. These data sets are expensive to gather and, once presented in the literature, often languish, are archived to digital media, and, sadly, forgotten. Had such data been placed where others could examine it using new methods, algorithms, and techniques, greater examination of underlying, fundamental neural processes might have been possible. Although the efforts of the NIH, journals, and others have tried to encourage greater sharing (Poline, Breeze *et al.* 2012), the reality remains that very little of the neuroimaging data gathered each day in the field have been made available to those who could help provide much needed understanding.

However prevalent data sharing efforts may seem today, the culture of data sharing in many fields has not formed without growing-pains. Early on, biomedical researchers were reluctant to freely provide data to each other, let alone openly available archives (Goodman 1996; Rennolls 1997; Rockwell and Abeles 1998). Data sharing difficulty has also been true of agricultural, geological, atmospheric, and other sciences, too, where, in spite of the deployment of storage infrastructure, data archives have sat empty awaiting datasets to be uploaded (Nelson 2009). People have had concerns that the process of sharing data took time away from conducting their next set of experiments. They worry that they might be scooped by others on some effect present in the data that they failed to see (Marshall 2002). They want to know what direct benefit there was to them if they shared their data. Only with time, the refinement of data exchange mechanisms, standards, example useful informatics only possible with large collections of information, in addition to the encouragement of journals and scientific societies, have researchers come to understand the benefits of sharing their primary data.

Yet in those fields where data sharing has caught on there appear to be incredible success stories where in some instances data are made available within days of their collection (Goel, Muthusamy *et al.* 2011; Dreszer, Karolchik *et al.* 2012; Milia, Congiu *et al.* 2012). Researchers are expected to provide genetic and gene expression data to NIH-based archives as soon as possible following collection (e.g. the NIH Database of Genotypes and Phenotypes (dbGAP); <http://www.ncbi.nlm.nih.gov/gap>). This has given rise to new forms of data-driven biomedical science.

Many now view biological databases like GenBank (Benson, Karsch-Mizrachi *et al.* 2012) and its associated National Center for Biotechnology Information (NCBI) assets as incredible success stories (<http://www.genomeweb.com/quarter-century-genbank>). Indeed, the sub-discipline of bioinformatics did not exist prior to the establishment of databases such as GenBank and accompanying data repositories. With these informatics tools in hand, one does not strictly need to be a microbiologist to conduct discovery-oriented science in the

field of genetics (Kolker, Stewart *et al.* 2012). So equipped, researchers have developed novel insights into the roles of genes in many human diseases (Roy-Engel, Carroll *et al.* 2001).

With the emergence and co-evolution of cognitive neuroscience and neuroimaging, we have long believed that the fMRI studies conducted and published to map the brain at work can form rich resources which can be mined, analyzed, and provide fundamental understanding of neural processes involved in mental operations. As is the case in many neuroimaging experiments, there are often dimensions of the data that are not fully explored or even recognized by the researchers obtaining it. If such data can be archived, indexed with accompanying meta-data, and combined, there is an enormous opportunity to obtain deep insights into the workings of the brain and mind.

In this article, we share our own experiences with the databasing of brain imaging data from published fMRI activation experiments hoping that it will serve as a useful example of how 1) sociological limitations to sharing can be overcome, 2) new and interesting science can emerge from shared data, and 3) the whims of funding support can make or break data archives. In what follows, we discuss our experience in the formation of the fMRI Data Center (fMRIDC) project - illustrating how the project got its start, the initial sociological concerns it experienced, its growth, several data re-use success stories, and provide comment on its current status.

The fMRI Data Center

In many ways, the fMRIDC was a novel experiment - one geared toward testing the notion that raw, complete fMRI studies (BOLD imaging time series, structural MRI, stimulus time courses, and other accompanying data) could be gathered from neuroimaging researchers and made openly available to anyone else. It had been shown that it was possible to build a useful database framework around collections of brain activation foci in Talairach/MNI atlas space for the purposes of meta-analytic inquiry (Fox, Mikiten *et al.* 1994). Yet, the contents of such a database are not entirely reflective of all the information available for analysis in those studies (Van Horn and Gazzaniga 2005) and which, as it happens, may collectively contain publication biases which need to be acknowledged and controlled for in order to maximize utility (Jennings and Van Horn 2012). As such, researchers wishing to assess new methods of comprehensive analysis, to scrutinize novel hypotheses, and prompt new discourse on the nature of fundamental cognitive operations as viewed using functional imaging would need access to the complete fMRI data sets from published research articles. This was the challenge we accepted.

The fMRIDC got its start in 1999 when we along with Daniel Rockmore, Peter Kostelec, Jeffrey Grethe, and Javed Aslam at Dartmouth College felt that the time was right to build a database for cognitive neuroimaging studies. The project was awarded funding from the National Science Foundation (NSF) and from the W.M. Keck Foundation to commence archive construction. An advisory committee of leading cognitive neuroscientists and brain imagers was formed to provide their recommendations and feedback. To seed the fMRIDC experiment, we worked closely with the *Journal of Cognitive Neuroscience* (JOCN) to ensure a representative corpus of studies would be obtained. To launch the database and obtain an initial set of fMRI data, a special issue of the journal appeared in 2000 featured thirteen neuroimaging activation studies from leading research teams from around the world (D'Esposito, National Science Foundation *et al.* 2000). Our interactions with other journals were highly supportive and, since several had already made database deposition of data from genetics, proteomics, and x-ray crystallographic experiments a condition of publication,

arrangements were made to do the same for studies using fMRI published in those periodicals.

However, upon becoming aware of our efforts and goals, fMRI researchers angered by journal requirements to provide copies of the fMRI data from their published articles began a letter writing campaign seeking to muster opposition against providing data to the fMRIDC¹ – an effort which was featured in the news and editorial sections of several influential journals (Aldhous 2000; Bookheimer 2000). Editorials and commentaries over fMRI data sharing were aired in the pages of *Science* (Marshall 2000), *Nature* (Editorial 2000b), *Nature Neuroscience* (Editorial 2000a), as well as in the journal *NeuroImage* (Toga 2002) expressing concern over the data sharing requirement, over what possession of the data implied, human subjects concerns, and, if databasing was to be conducted at all, how it should be conducted “properly”. Leadership groups in the field argued that fMRI was not mature enough to begin archiving its data (Governing Council of the Organization for Human Brain Mapping 2001), conjecturing that until the BOLD response was better understood, it was too early to consider databasing the images of published studies. People privately complained that those who collected the data owned it, that they would be remiss in simply giving it away, and that a small team based at a small Ivy League institution was not the best group to take on the task of archiving it. As a result of apparent magnitude of these concerns, many of the journals, who had initially been so supportive, decided not to require submission of data from the fMRI studies they published. Instead they hoped to wait out the controversy and let the community itself resolve the issue.

Having believed that we had embarked on a journey to develop a useful resource to support the work of cognitive neuroscience, promote data re-use and new discovery, the reactions of our colleagues caught us somewhat off guard. We were honestly surprised by the negative and hostile response when we had believed that creation of a data archive would be of benefit for the neuroimaging community. Perhaps, they had a point. Maybe the field wasn't ready for fMRI data sharing? Perhaps, it was too early to start such a project? We struggled with how best to move forward or whether to move forward at all.

Yet, rather than surrender to the view that we were premature to begin archiving fMRI data, we decided to continue - full steam ahead (Van Horn, Grethe *et al.* 2001). We chose to believe that if we waited for the field to mature that the best parts of this rapidly developing field would be over. Indeed, we felt that we would miss a valuable opportunity to take an ongoing set of snapshots of representative fMRI studies as they became more rigorous and exacting. While challenges would exist in gathering, organizing, and storing complete data sets, we were convinced, as were others (Johnson 2002), that these limitations could be overcome. With these driving principles in mind, despite receiving a disappointing level of community support, we began our fMRI databasing efforts in earnest.

Databasing Hardware and Man-Power

Our first step was to put into place the necessary computational and storage infrastructure needed to house complete fMRI studies. Even in the early 2000's, fMRI studies were considered “big data” requiring appropriate computational infrastructure for archival storage and processing (Van Horn, Dobson *et al.* 2006). To accommodate submissions of study data, the fMRIDC deployed several large-scale computer servers along with several terabytes of

¹The link to the original letter of opposition, authored by Isabel Gauthier and colleagues has been lost, though its text can be found courtesy of the SPM JISMAIL archives: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind00&L=spm&P=R113930&I=spm&9=A&I=-3&J=on&d=No+Match%3BMatch%3BMatches&z=4>.

disk storage. These systems were necessary to provide internet connectivity, file management, processing capability for data anonymization, etc.

One aspect of data sharing became abundantly clear to us very soon after getting underway – it turns out that people are not very good about managing their own data. We were amazed by the variety of ways in which neuroimaging data sets were contributed to the archive and how they were stored. We received data on tape, DVD, CD, and, of course, via FTP/SCP. However, it was not uncommon for researchers to give up entirely attempting to organize their data for submission and they literally pulled the hard drives from their computers and mailed them to us! File formats varied widely and file structures even more.

Once data arrived at the fMRIDC, two full-time study curators organized all the information, ensured a consistent file structure, image file types, and verified the validity of all meta-data and experimental protocol descriptor files. In some cases, study authors failed to completely anonymize their data (or didn't at all), leaving identifying information in various fields of header files, in experimental time course files, and other accompanying file locations. These had to be systematically located and be removed in order for the study to comply with HIPAA requirements for subject privacy. At each step, the curatorial team worked to stay in touch with the submitting authors to get clarification when needed, obtain additional files when needed, and to ensure accuracy. Depending on the size of the study, these processes could take a couple days or over a month to perform.

Copies of all completely curated and packaged datasets were maintained on spinning disk. While perhaps small by today's standards, the fMRIDC dedicated over 12 terabytes of storage to the database. A pair of system administrators ensured that these systems maintained 98% up-time and supervised database backups which involved having redundant copies of the entire database, nearline versions of each study, and copies of all data stored in an off-site location. Computer software engineers wrote customized code to automate various data processing tasks (e.g. face anonymization on high-resolution structural images), perform web design, and to create image viewing and meta-data ontology inspection interfaces (Van Horn, Woodward *et al.* 2002). Having a dedicated curatorial, system administration, and software engineering support staff was a critical element of fMRIDC activities and their roles cannot be overstated.

Data Sharing Sociology

Data ownership issues were among the many initial concerns people had about the fMRIDC. We handled this particular issue through an embargo process which allowed for up to six months post-article publication for authors to conduct any other analyses they had planned. This would allay the fears of researchers who imagined working tirelessly on a follow-up project, only to have someone else beat them to the punch by publishing on the results of their data first. Interestingly,

The integration of data sharing into the publication timeline was a particularly important step. Once an article had been accepted for publication at JOCN, the fMRIDC was notified and the authors were given detailed instructions on how to begin preparing data for deposition to the fMRIDC archive. Communication with study authors was essential to ensure accuracy and completeness. Reminders were sent to lead study authors as needed and, if necessary in the case of non-responsiveness, to senior authors indicating the possibility of delays in the publication process if data were not provided. The goal was to have received all the neuroimaging and meta-data, have it entered into the database, and packaged for dissemination simultaneous with the issue of the journal containing the published article. The archive also served as a useful resource also for investigators whose funding agencies required the sharing of data as a condition of funding. Numerous other

researchers kindly volunteered their data for inclusion. Using this approach, through the course of its full-scale operation (2000–2006), the fMRIDC was able to obtain and archive over 100 complete fMRI studies from the published literature and focus on numerous cognitive paradigms. This represents many terabytes worth of data from many thousands of individual subjects and many dozens of individual scanners from around the world².

Examples of Data Re-Use

The importance of the fMRIDC as an entity for us was never in how many bits of data were stored but in the use of the resource toward new applications and new insights into cognitive function. Through the use of data obtained from the fMRIDC several papers were published by top groups from around the world in leading journals that (arguably) might not have been attempted otherwise or would have been costly to perform *de novo*.

One particularly interesting example of data re-use involves the use of data from older patients with pre-Alzheimer's dementia, otherwise healthy older subjects, and young healthy subjects originally collected by Buckner et al. to make note of the challenges in estimating the hemodynamic response function in older subjects and those with diseases of brain aging (Buckner, Snyder *et al.* 2000). Greicius and coworkers (Greicius, Srivastava *et al.* 2004) obtained these data from the fMRIDC archive and performed an independent components analysis to identify and measure the degree of default mode activity present in each group. They noted that older healthy adults had diminished magnitude of default mode activity as compared to younger subjects, while those subjects with pre-Alzheimer's dementia showed still further reductions. These results were interesting in that they provided the Buckner data set with a clinical importance by suggesting that alterations in default mode activity may be a useful biomarker for degenerative brain disease.

Several other examples exist where data from the fMRIDC archive were re-used and intriguing new results and/or conclusions were obtained (Lloyd 2002; Carlson, Schrater *et al.* 2003; Mechelli, Price *et al.* 2003; Aizenstein, Butters *et al.* 2009). With such activity to take old data and form new results from it, even those journals which featured the brain imaging community's misgivings about fMRI data sharing had to admit that the fMRIDC might, in fact, be on to something (Barinaga 2003; Leslie 2003). What is particularly satisfying is that published re-examinations of data obtained from the fMRIDC have continued even until recently (Liou, Savostyanov *et al.* 2012).

New Times, New Data Sharing Expectations

But just as neuroimaging community was coming around to the idea that sharing their data could help to promote and propel cognitive neuroscience, the storm clouds on the funding horizon meant that change was coming for the fMRIDC. The NSF decided to establish new programs and discontinue the one that had funded our efforts. The Keck Foundation preferred to get new efforts started and had no interest in longer term support for neuroimaging databases. The National Institutes of Mental Health, who had once been so supportive of neuroinformatics activities, decided to redirect its efforts on informatics elsewhere, cancelling the well established Human Brain Project (Huerta, Koslow *et al.* 1993; Shepherd, Mirsky *et al.* 1998; Brinkley and Rosse 2002), leaving no funding mechanisms in place to support data sharing, databasing, mining, and meta-analytic activities (<http://grants.nih.gov/grants/guide/notice-files/NOT-MH-05-014.html>). Despite protests from us and other neuroscience leaders that neuroinformatics deserved continued NIH support

²We note the integration of fMRIDC data submission into the publication was greatly facilitated through the support of Dr. Mark D'Esposito, editor-in-chief of the *Journal of Cognitive Neuroscience*. His editorial involvement and enthusiasm for the project during its heyday is greatly appreciated.

(Gazzaniga, Van Horn *et al.* 2006) it was clear that the NIH was moving in other directions. With changes in the focus of major funding agencies and a move of project senior personnel from Dartmouth to the University of California Santa Barbara, necessitating a move of the whole fMRIDC operation, it was difficult to rationalize the case for ongoing full-scale effort. With uncertainty in funding and no dedicated staff working to curate studies, JOCN determined that it would no longer expect data submissions to the archive as a condition of publication. So like all experiments, the fMRIDC's efforts to curate new fMRI cognitive neuroimaging studies had reached its conclusion.

We can look upon it now and recognize that the fMRI Data Center was, in fact, a very successful endeavor. Even today, with over 100 complete, published studies, containing thousands of subjects, and countless functional imaging time points, the fMRIDC arguably contains more fMRI data from published cognitive activation studies than any other archive of its kind (<http://www.fmridc.org>). Useful new research has appeared in the literature on the basis of the available studies in the archive and new collaborations formed which may not have happened in the absence of openly shared data. That the fMRIDC has been able to contribute in these ways is, from our perspective, the principle hallmark of its success. Its early detractors, at the time so opposed to it, now frequently confess to us that it was, in fact, a pioneering idea and are impressed how far we were able to take it. Investigators worldwide continue to request studies from the archive for re-analysis or use in education. That even in its reduced form the fMRIDC is still enabling researchers to generate new research is a remarkable statement about the potential for emerging neuroimaging databases to find a market for shared data.

Advice for the Next Era of Neuroimaging Databasing

And emerge they have. New functional imaging databases have begun in only the past few years and appear to have found researchers now ready to contribute their data and use it, too. The 1000 Functional Connectomes Project (FCP) and its International Neuroimaging Data Initiative (INDI) have grown enormously, through grass-roots effort, into the primary resources for resting-state fMRI data (Biswal, Mennes *et al.* 2010; Milham 2012). The OpenfMRI project (<http://openfmri.org>) is currently gaining speed. The NIH itself now sees a bright future for neuroimaging data sharing and has decided to develop a series of disease focused neuroimaging database resources. The first of these to emerge has been the National Database for Autism Research (NDAR) which seeks to gather all brain imaging data from NIH-funded studies of autism spectrum disorder (Hall, Huerta *et al.* 2012). Other NIH-based neuroimaging repositories are being deployed (e.g. for brain trauma; <https://fitbir.nih.gov/tbi-portal>) or are under development.

From our perspective of having been through the complete life-cycle of neuroimaging databasing, we are in a unique position to offer recommendations for existing and emerging databasing efforts as we enter a new era for data exchange and sharing. This is not to suggest that we have cornered the market on the wisdom of data sharing or have some deep philosophical agenda for databasing best practices. Rather, we share our views based on the practicalities of what it takes to motivate, manage, and deliver data efficiently and effectively.

There are many ways to motivate the sharing of data

Voluntary sharing of data may be the ideal, but many researchers find themselves very busy and, given the choice of how to spend their time, may not find the benefits to the community to be compelling enough to make the effort. Additionally, only having data from a “coalition of the willing” may not fully capture the breadth and depth of imaging experimental methods being applied across the field. On the other hand, sharing required as a condition of

funding or journal publication can ensure a steady stream of data. However, the mere fact that sharing is *required* may be met with resistance and prove to be unpopular. Working closely with funding agencies and/or journal editors will be an important element in such instances.

Good curation is essential

Depending on busy investigators to re-format their data specifically for your database is unlikely to be a sound model. For every investigator, there is every type of organizational scheme and to create a universal data entry method or ontological structure, despite some progress in this regard (Bug, Astahkov *et al.* 2008; Turner, Mejino *et al.* 2010), remains massively complex and has the potential to be an obstacle to innovation and progression. While electronic data capture methods have a very important role to play (Turner and Van Horn 2012), it is essential to have dedicated human curators ensuring the consistency of the data, troubleshooting format issues, interacting with the contributing authors, and extracting useful meta-data to summarize the study. If we are to move beyond anonymous FTP-based file repositories, such guided organization and standards for curation will be critical (Van Horn and Ball 2008; Van Horn and van Pelt 2008).

Multiple data sharing models exist and can co-exist

Some may argue that only sharing a portion of the data processing workflow or summary images and activation results are sufficient for the purposes of data sharing. Others will argue that original, raw data are the only way to go. In our view, there are multiple points of entry into the data processing stream from the original data, preprocessed images, to summary results. Each point has its own interested audience. However, no database is likely to have the resources to accommodate them all and some resources will do a better job with one than another. Interoperability between efforts will be an important step. In short, there is plenty of work to go around.

Deliver study information in whatever forms necessary

While providing data online seems obvious, some potential users may not have sufficient hard disk space to accommodate a large study. If they download data and it crashes their hard drive, they will blame you. In its early days, the fMRIDC chose to provide data on CD and DVD to avoid such instances and to allow users to place data where they wanted it at their sites. It now provides data as multiple downloadable files which helps people stage the amount of data they receive at once. However, providing certain study or summary information as PDFs, XML tables, even ASCII text files makes the data as widely useable as possible.

Listen to the community but have a long-range vision

Many will have opinions about how you should go about your databasing and what it should deliver. Take your peers' suggestions, recommendations, and complaints to heart. They are your user-base, audience, advocates, and critics. Keep abreast of what types of data they are collecting, how they are collecting it, and adjust your methods accordingly as you do so. Form an external advisory board to help you stay in touch with what is happening in the field that can assist in make your database stronger, more representative, and to better serve your target audience. Get input from the community but ensure that your effort has a long-term goal and don't waiver from getting there.

Have a plan for what happens AFTER your grant funding runs out

The NIH or NSF may have funded your effort now, but that does not mean they will continue to do so in the future – regardless of your database's real or perceived degree of

success. Federal and private funding themes change and can change quickly. In lean times, you may find that your staff has fled for more secure jobs elsewhere. It may also be hard to replace aging computers and storage space, which threatens data preservation. For your data sharing effort to ride out a funding shortfall, you should work to become self-sufficient, if possible, with an effective business model that preserves ongoing data deposition and ensures open and unfettered access to database contents.

While, most importantly, a database should have its contents used toward new and novel applications. If data is entered into the database but doesn't come out again, then the value of the archive is likely going to be minimal. If data are being re-used and are generating new, independently produced peer-reviewed publications, then it could be argued that this is the best hallmark of a database's success. Database creators should work to encourage workers both within and outside of their own fields to explore the archive, mine it, combine studies, and extract useful new information. Moreover, re-used data should not simply form another collection of derivative results, but constitute a true synthesis of effects which expose new relationships not observable previously by the original studies. These should, in turn, generate new, testable hypotheses which can spur on further empirical data collection. Data sharing in the fMRI world will truly catch on when such studies begin to yield important results not simply just more results.

Discussion

Cognitive neuroscience and *in vivo* neuroimaging using fMRI have each other to thank for the levels of maturity they have both attained. Research in one has greatly depended on work in the other and both are the better for it. Shared data obtained in cognitive neuroimaging studies has also helped to push the advancement of analytic methods, data mining, modeling, and visualization techniques (Van Horn and Ishai 2007). Examples of such symbiotic science are often difficult to come by and illustrate that more can be gained through dynamic interactions than through silo-based research.

To fully re-invigorate an archive such as the fMRIDC from scratch with NIH or NSF support would be costly - likely more than the cost of several individual fMRI-related RO1 grants. However, the value to the neuroscience community of having these unique fMRI datasets available and adding to this collection with newer, larger, more precise, more interesting data sets would be incalculable. We only need look to the field of bioinformatics, which didn't exist before people began sharing their genomic data, and which has been essential in identifying the functional significance of numerous genes and whose methods are now utilized daily by neuroimaging scientists via genome-wide association studies (GWAS) and the effects of genes on cognition/activation. Having an archive that could incorporate all of this will be essential as we tackle new and increasingly complex puzzles in functional neuroimaging. New NIH-based resources and independent efforts like FCP/INDI may mean that data sharing may soon become the expectation rather than the exception. With lessons learned from our fMRIDC experiment, we look forward to further progress and new developments as the field of neuroimaging and cognitive neuroscience continue to grow and share their data.

References

- Aizenstein HJ, Butters MA, Wu M, Mazurkewicz LM, Stenger VA, Gianaros PJ, Becker JT, Reynolds CF 3rd, Carter CS. Altered functioning of the executive control circuit in late-life depression: episodic and persistent phenomena. *Am J Geriatr Psychiatry*. 2009; 17(1):30–42. '2626170.' 2626170. [PubMed: 19001356]
- Aldhous P. Prospect of data sharing gives brain mappers a headache. *Nature*. 2000; 406(6795):445. [PubMed: 10952275]

- Barinaga M. Neuroimaging. Still debated, brain image archives are catching on. *Science*. 2003; 300(5616):43–45. [PubMed: 12677038]
- Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2012; 40:D48–D53. (Database issue) '3245039:' 3245039. [PubMed: 22144687]
- Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski AM, Ernst M, Fair D, Hampson M, Hoptman MJ, Hyde JS, Kiviniemi VJ, Kotter R, Li SJ, Lin CP, Lowe MJ, Mackay C, Madden DJ, Madsen KH, Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SA, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, G JS, Sorg C, Teng GJ, Veijola J, Villringer A, Walter M, Wang L, Weng XC, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang YF, Zhang HY, Castellanos FX, Milham MP. Toward discovery science of human brain function. *Proc Natl Acad Sci U S A*. 2010
- Bookheimer S. Brain Mapping Researchers Voice Concern Over Compulsory Data Sharing. *Neuroreport*. 2000; 11(17):A13.
- Brinkley JF, Rosse C. Imaging and the Human Brain Project: a review. *Methods Inf Med*. 2002; 41(4): 245–260. [PubMed: 12425235]
- Buckner RL, Snyder AZ, Sanders AL, Raichle ME, Morris JC. Functional brain imaging of young, nondemented, and demented older adults. *J Cogn Neurosci*. 2000; 12(Suppl 2):24–34. [PubMed: 11506645]
- Bug W, Astahkov V, Boline J, Fennema-Notestine C, Grethe JS, Gupta A, Kennedy DN, Rubin DL, Sanders B, Turner JA, Martone ME. Data federation in the Biomedical Informatics Research Network: tools for semantic annotation and query of distributed multiscale brain data. *AMIA Annu Symp Proc*. 2008:1220. [PubMed: 18999211]
- Carlson TA, Schrater P, He S. Patterns of Activity in the Categorical Representations of Objects. *J. Cogn. Neurosci*. 2003; 15(5):704–717. [PubMed: 12965044]
- D'Esposito M. National Science Foundation and W.M. Keck Foundation. Special Issue Celebrating the Launching of the NSF/Keck Foundation National fMRI Data Center. 2000
- Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, Meyer LR, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Pohl A, Malladi VS, Li CH, Learned K, Kirkup V, Hsu F, Harte RA, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, James Kent W. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res*. 2012; 40:D918–D923. (Database issue) '3245018:' 3245018. [PubMed: 22086951]
- Editorial. A debate over fMRI data sharing. *Nat Neurosci*. 2000a; 3(9):845–846. [PubMed: 10966604]
- Editorial. Whose scans are they anyway? *Nature*. 2000b; 406:443.
- Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, Miller KL, Ugurbil K, Yacoub E. Multiplexed Echo Planar Imaging for Sub-Second Whole Brain fMRI and Fast Diffusion Imaging. *PLoS ONE*. 2010; 5(12):e15710. [PubMed: 21187930]
- Fox, PT.; Mikiten, S.; Davis, G.; Lancaster, J. BrainMap: A database of human function brain mapping:. In: Thatcher, RW.; Hallett, M.; Zeffiro, T.; John, ER.; Heurta, M., editors. *Functional Neuroimaging Technical Foundations*. San Diego: Academic Press; 1994. p. 95-105.
- Gazzaniga MS, Van Horn JD, Bloom F, Shepherd GM, Raichle ME, Jones EG. Continuing Progress in Neuroinformatics. *Science*. 2006; 311(5758):176. [PubMed: 16410506]
- Goel R, Muthusamy B, Pandey A, Prasad TS. Human protein reference database and human proteinpedia as discovery resources for molecular biotechnology. *Mol Biotechnol*. 2011; 48(1):87–95. [PubMed: 20927658]
- Goodman KW. Ethics, genomics, and information retrieval. *Comput Biol Med*. 1996; 26(3):223–229. [PubMed: 8725773]
- Governing Council of the Organization for Human Brain Mapping. *Neuroimaging Databases*. *Science*. 2001; 292:1673–1676. [PubMed: 11387466]
- Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proc Natl Acad Sci U S A*. 2004; 101(13):4637–4642. [PubMed: 15070770]

- 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]
- Huerta MF, Koslow SH, Leshner AI. The Human Brain Project: an international resource. *Trends Neurosci*. 1993; 16(11):436–438. [PubMed: 7507612]
- Jennings RG, Van Horn JD. Publication bias in neuroimaging research: implications for meta-analyses. *Neuroinformatics*. 2012; 10(1):67–80. [PubMed: 21643733]
- Johnson DH. The Power of Psychology's Databases. *The American Psychological Society Observer*. 2002; 15
- Kolker E, Stewart E, Ozdemir V. Opportunities and challenges for the life sciences community. *Omics*. 2012; 16(3):138–147. '3300061:' 3300061. [PubMed: 22401659]
- Koslow SH. Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*. 2000; 3(4):863–865.
- Leslie M. DATABASE: Swap Meet for Brain Mappers. *Science*. 2003; 299(5613):1637.
- Liou M, Savostyanov AN, Simak AA, Wu W-C, Huang C-T, Cheng PE. An Information System in the Brain: Evidence from fMRI BOLD Responses. *Chinese Journal of Psychology*. 2012; 54(1):1–26.
- Lloyd D. Functional MRI and the study of human consciousness. *J Cogn Neurosci*. 2002; 14(6):818–831. [PubMed: 12191448]
- Marshall E. A Ruckus Over Releasing Images of the Human Brain. *Science*. 2000; 289(5484):1458–1459. [PubMed: 10991728]
- Marshall E. Data sharing. DNA sequencer protests being scooped with his own data. *Science*. 2002; 295(5558):1206–1207. [PubMed: 11847311]
- Mechelli A, Price CJ, Noppeney U, Friston KJ. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *J Cogn Neurosci*. 2003; 15(7):925–934. [PubMed: 14628754]
- Milham MP. Open neuroscience solutions for the connectome-wide association era. *Neuron*. 2012; 73(2):214–218. [PubMed: 22284177]
- Milia N, Congiu A, Anagnostou P, Montinaro F, Capocasa M, Sanna E, Destro Bisol G. Mine, yours, ours? Sharing data on human genetic variation. *PLoS ONE*. 2012; 7(6):e37552. '3367958:' 3367958. [PubMed: 22679483]
- Nelson B. Data sharing: Empty archives. *Nature*. 2009; 461(7261):160–163. [PubMed: 19741679]
- Poline JB, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, Haselgrove C, Helmer KG, Keator DB, Marcus DS, Poldrack RA, Schwartz Y, Ashburner J, Kennedy DN. Data sharing in neuroimaging research. *Front Neuroinform*. 2012; 6:9. '3319918:' 3319918. [PubMed: 22493576]
- Rennolls K. Intersalt data. Science demands data sharing. *Bmj*. 1997; 315(7106):486–487. [PubMed: 9284679]
- Rockwell RC, Abeles RP. Sharing and archiving data is fundamental to scientific progress. *J Gerontol B Psychol Sci Soc Sci*. 1998; 53(1):S5–S8. [PubMed: 9469174]
- Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL. Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*. 2001; 159(1):279–290. '1461783:' 1461783. [PubMed: 11560904]
- Shepherd GM, Mirsky JS, Healy MD, Singer MS, Skoufos E, Hines MS, Nadkarni PM, Miller PL. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends Neurosci*. 1998; 21(11):460–468. [PubMed: 9829685]
- Toga A. Neuroimaging Databases: The Good, The Bad, and The Ugly. *Nature Reviews Neuroscience*. 2002 Apr.3:302–309.
- Turner JA, Mejino JL, Brinkley JF, Detwiler LT, Lee HJ, Martone ME, Rubin DL. Application of neuroanatomical ontologies for neuroimaging data annotation. *Front Neuroinformatics*. 2010; 4 '2912099:' 2912099.
- Turner JA, Van Horn JD. Electronic data capture, representation, and applications for neuroimaging. *Front Neuroinform*. 2012; 6:16. '3345526:' 3345526. [PubMed: 22586393]
- Van Horn, JD. Functional Neuroimaging and Cognitive Neuroscience: Past Successes, Future Directions. In: Gazzaniga, MS., editor. *The New Cognitive Neurosciences*. Cambridge, MA: The MIT Press; 2004.

- Van Horn JD, Ball CA. Domain-specific data sharing in neuroscience: what do we have to learn from each other? *Neuroinformatics*. 2008; 6(2):117–121. [PubMed: 18473189]
- Van Horn, JD.; Dobson, J.; Woodward, J.; Wilde, M.; Zhao, Y.; Voeckler, J.; Foster, I. Grid-Based Computing and the Future of Neuroscience Computation. In: Senior, C.; Russell, T.; Gazzaniga, MS., editors. *Methods in Mind*. Cambridge: MIT Press; 2006. p. 141-170.
- Van Horn, JD.; Gazzaniga, MS. Maximizing Information Content in shared Neuroimaging Studies of Cognitive Function. In: Koslow, SH.; Subramanian, A., editors. *Databasing the Brain: From Data to Knowledge*. New York: John Wiley and Sons; 2005.
- Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, Rockmore D, Gazzaniga MS. The Functional Magnetic Resonance Imaging Data Center (fMRIDC): The Challenges and Rewards of Large-Scale Databasing of Neuroimaging Studies. *Philos Trans R Soc Lond B Biol Sci*. 2001; 356:1323–1339. [PubMed: 11545705]
- Van Horn JD, Ishai A. Mapping the human brain: new insights from FMRI data sharing. *Neuroinformatics*. 2007; 5(3):146–153. [PubMed: 17917125]
- Van Horn JD, van Pelt J. 1st INCF Workshop on Sustainability of Neuroscience Databases. *Nature Precedings*. 2008 <<http://dx.doi.org/10.1038/npre.2008.1983.1>>.
- Van Horn, JD.; Woodward, JB.; Simonds, G.; Vance, B.; Grethe, JS.; Montague, M.; Aslam, JA.; Rus, D.; Rockmore, D.; Gazzaniga, MS. The fMRI Data Center: Software Tools for Neuroimaging Data Management, Inspection, and Sharing. In: Kotter, R., editor. *A Practical Guide to Neuroscience Databases and Associated Tools*. Amsterdam: Kluwer; 2002. p. 221-235.

Highlights

1. The fMRI Data Center was established to gather data from published research articles
2. Curators worked closely with publishers and authors to ensure full submissions
3. Studies provided to users worldwide forming the basis for new research articles
4. Provides a productive and replicable model for neuroimaging databasing
5. Lessons learned can inform and guide existing and new fMRI data sharing