THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF SCIENCE

AN EXAMINATION ON THE ADVANTAGES OF OPEN DATA SHARING IN BEHAVIORAL NEUROSCIENCE RESEARCH

MATTHEW LEE
FALL 2014

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Science
with honors in Science

Reviewed and approved* by the following:

Rick Gilmore
Professor of Psychology
Thesis Supervisor

Henriette Evans
Advising Program Coordinator, Science and Premedicine Programs
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The importance of data sharing among scientific investigators is becoming more recognized in today's research society. Open data sharing has enabled the greatest possible use of data and has helped increase the rate of scientific discovery. Sharing data allows researchers to have their data validated, has let researchers support each other's investigations, and allows data to be used to its maximum potential. With that being said, there are obstacles standing in way of the widespread use of data sharing, especially in the behavioral neuroscience field. These obstacles include: a lack of standard tools for neuro-scientific data sharing, a lack of standard protocol to prepare data for sharing, and little incentive for researchers to share their data. This is troubling because the computational, neuroimaging, and behavioral methods used in behavioral neuroscience could be improved to enhance the rate and quality of research in the field.

This thesis will highlight today's most common open data sharing initiatives in the behavioral neuroscience field for functional neuroimaging and video data and note the benefits and drawbacks of each. Furthermore, this thesis will explain what is still needed to enhance data sharing in the behavioral neuroscience field and will incorporate recent open data sharing initiatives in other fields. Lastly, the direction in which open data sharing (in a broad sense) is heading will be observed. In whole, my thesis is an examination of the advantages of open data sharing in behavioral neuroscience research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank all of those involved in helping this thesis possible. Most notably, I would like to thank Dr. Rick Gilmore for his continued support during each phase of my thesis as well as during the past three years in the Brain Development Lab. I would also like to thank Henriette Evans, my honors advisor, for her guidance during this process and for being my second reader. Lastly, I would like to thank the Schreyer Honors College faculty and staff for their help during the planning, writing, formatting and publishing stages of this thesis.

Chapter 1

Introduction

Open Data Sharing Defined

In today's technologically growing society, research groups from a wide variety of fields are joining the open data sharing movement. According to the Open Knowledge Foundation (OKF), open data sharing can be defined as any form of knowledge that can be freely shared, used and further developed by any person at any moment. The term "open" as defined by OKF pertains to the freedom to use, reuse, and redistribute knowledge without limitations beyond a requirement for acknowledgment. Secondly, the term "sharing" emphasizes the importance of usability and availability of a data set. Typically, the work that is being shared is accessible through a free database using the Internet and should be available in a suitable and adjustable form. Lastly, the term "data" can be divided into three different areas including: content (forms of entertainment such as music, movies and books), evidence (from various scientific fields) and government information (Molloy, 2011).

Although it could be argued that open data sharing can be dated back to 1790 with the United States' first census after the American Revolution, open data sharing did not become widely visible in the mainstream until 2009 when the governments of the United States, United Kingdom, Canada and New Zealand announced new initiatives to share their information to the public (Hendler et al. 25). Open government data (OGD) has emerged as an essential communication channel between governments and their citizens. Government datasets are comprised of a wide range of information related to agriculture, climate, education, energy, finance, global development, healthcare, and public safety. Each government database releases hundreds of thousands of datasets from various government application programming interfaces and agencies (Ling 325).

This government-related open data sharing initiative sparked interest in other fields as well. The scientific community has recently begun the movement to share data among research groups in each respective field.

Science and Open Data Sharing

Data is a key foundation of science. The experimentation process of the scientific method is built around collecting, analyzing, publishing, adapting, and reusing data. Accurate data is vital for the formation of scientific decisions and development of scientific principles, and is essential to make informed decisions and manage scientific resources. According to the National Science Foundation (NSF), science in the 21st century has become substantially more data-intensive and collaborative due to the recent developments in digital technology, computational modeling, automated data acquisition, and virtual communication (National Science Foundation, 2010). In fact, this new research era has been called "the fourth paradigm" which focuses on a data-intensive scientific discovery where data sets have moved from literature to online (Hey et al., 2009). As this new scientific research paradigm draws closer to more data-driven information, collaboration through open data sharing becomes important.

Many different scientific fields, ranging from biology to chemistry to physical sciences heavily rely on collaboration and social interaction to aid in research efforts. Traditionally, this group effort has been seen within a lab group or research team. Recently, the digital age has allowed researchers to expand their network to other teams, specialists, and colleagues in their field (Hey et al., 2009). International foundations such as the World Data System (Minster, 2014) and OKF have pushed for open data sharing among the scientific community to not only unite researchers, but to improve the research process as a whole (Molloy, 2011).

Advantages of Open Data Sharing in Science

There are many reasons for this recent upsurge of open data sharing within the scientific community. First, data sharing allows for verification of a research study, which is a vital component of the scientific research process (Tenopir et al., 2011). Furthermore, sharing data can lead to different variations and adaptations of experiments. This can contribute to the progress and advancement of scientific research (Molloy, 2011). Increased access to scientific data has often led to significant scientific discovery and opportunity for educational growth in both academic and research settings. In addition, when data sets are easily accessible, recollection of data becomes more efficient, and this allows researchers to avoid wasting resources (Tenopir et al., 2011). Making data widely available also provides safeguards against data fabrication, which presents a large problem even in today's research community (Eisenach 955). Lastly, data sharing has proven to be an important aspect of the research process. According to a survey performed by the Publishing Research Consortium (PRC), all 3823 respondents (a group comprised of scientists and researchers of a multitude of backgrounds, locations, and fields of study), ranked access to data sets as "highly important" or "important" to their research ("Access vs. Importance", 2010). As more data becomes more easily accessible, the scientific process becomes more efficient. This yields benefits to society (Tenopir et al., 2011).

From a business perspective, sharing data can reduce the cost of research and increase the return on current investments. Collecting data using neuroimaging methods can be highly expensive and time consuming. Due to the lack of data sharing in this field, there is the risk of wasting a significant amount of money by duplicating already performed data (Poline et al. 47). Significant money, time and resources could be saved if there was a standard method of sharing neuroscience data. These savings could then be used for further data collection, analysis, and presentation of data.

Overview

Open data sharing is growing significantly among biological, chemical and other similarly-related fields, but has yet to become prominent in the behavioral neuroscience field. For instance, biological and chemical laboratories frequently share genetic and protein sequences respectively. Behavioral neuroscience is the study of the nervous system and how it affects behavior, focusing on perception, learning and memory, and motor performance (Teeters et al. 47). The research methods regarding the relationship between the human brain and behavior is fairly distinct as compared to other scientific research methods. A large aspect of behavioral neuroscience research lies in functional neuroimaging, computer modeling, and videoanalyses. Despite the vast amounts of digital data within this field, researchers rarely share the aforementioned forms of data with other lab groups (Poline et al. 47). This presents a large problem for behavioral neuroscience research because impedes rapid advancement in the field.

This paper highlights the rising trend of open data sharing in today's scientific community, with specific focus on the behavioral neuroscience field. It explores the advantages of as the barriers to open data sharing and its importance to behavioral neuroscience. The paper emphasizes many of the ongoing efforts among various research labs throughout the world and provides insight into the future of open data sharing in behavioral neuroscience.

Chapter 2

Open Data Sharing in Behavioral Neuroscience

The Need

The main reason that data should be shared within the behavioral neuroscience field is to accelerate the progress in the understanding of one of the most important organs of the human body, the brain. The Functional Biomedical Informatics Research Network (FBIRN) is just one testament to the success and advancement sharing behavioral data sets can bring to the field (Glover et al. 39). The network is essentially a data repository for neuroimaging data, specifically related toschizophrenia research. A number of research groups including the Mind Research Network (Kim et al. 3795) and the University of California, Irvine Department of Psychiatry (Potkin et al. 15) reported using shared data from FBIRN—collected from a multitude of institutions—to accelerate their understanding of schizophrenia.

Another benefit of sharing data is related to the clinical outcomes of various experiments. Data sharing could lead to higher successes in clinical diagnosis and treatment of disorders. The idea is that neuroimaging research in the form of electroencephalography (EEG), functional magnetic resonance imaging (fMRI) and diffuse optical imaging data can be shared to help research groups make accurate treatments and clinical decisions. For example, the results from a general study on motor development could be applicable to a different study trying to determine hormone effects on coordination. The results from these studies could then be used for a different study and continue to make advancements in other behavioral neuroscience clinical experiments.

The last reason to openly share neuroimaging data is because other scientific fields have had success with data sharing. One example lies in the field of genetics. HapMap and GenBank are two of the

largest open data repositories for human genome patterns. These types of open databases have led to significant discoveries in genetically defined risks and diseases (Manolio et al. 1590). According to Kaye et al. major funding applications for genomics research now include an open data sharing requirement because of the advantages this type of system brings to the field. These types of requirements from various funding agencies are becoming more prevalent in other life science research fields as well (Kaye et al. 331). This concept can be translated to the field of behavioral neuroscience research, as this field would also flourish with more standardized practices of open data sharing.

Current Barriers

There are three main reasons why more data is not being shared in the behavioral neuroscience field: motivation, ethical and legal concerns, and technical issues. Motivation has to do with why researchers should share their data and why they would use other researchers' data. The sharing of EEG or fMRI data can be both costly and time consuming (Wilhelm et al. 1201). With regards to money, the process of sharing data may incur costs related to the type of data being shared, the number of submissions, and the quantity of data depending on the number of subjects involved. Costs for maintaining the infrastructure or the physical repository for the shared data may fall to the researcher as there is a consistent need to administer, maintain, and manage the data (Poline et al. 47). Thus, in a competitive research field where there is little or tight competition for funding, motivation is a key factor in sharing data. Furthermore, sharing data leads to a higher chance that someone may dispute a particular result or finding. Lastly, researchers may not be fully confident about their research methods and may not feel comfortable sharing their data with the entire research community for fear of being debunked (Poline et al. 47).

Aside from lack of motivation, ethical and legal issues may act as impediments to open data sharing. There are various regulations and ethical standards when related to the collection and use of

human subject data. Most research studies dealing with human participants must secure informed consent from research participants, often by means of a signed consent form that describes how the subject data will be used. Research protocols and consent forms must gain approval from the Institutional/Ethical Review Boards (IRB or ERB). Since many of these informed consent forms do not include the permission to share data openly, researchers who would like to share data may have to revise their research protocols and secure new or additional written consent from participants tested in the past. Securing these approvals is often time consuming and may be deemed superfluous by some researchers. In fact, Kehagia et al. found that many researchers believe they are already spending too much time completing the ethics review processes and other administrative tasks (Kehagia et al. 775).

Technical problems may arise when sharing neuroimaging data. There are some data repositories that collect masses of data from a wide range of disparate studies, but most repositories are focused on a specific project and the data may not be easy to convert from tool to tool (Poline et al. 47). For instance, a database including fMRI data may be difficult to include in a database holding behavioral video data. Furthermore, the level of detail in the data set along with the organization and format of the data being shared becomes an issue. Since data from each study will have a different level of detail and a unique organization and format, a standard form of technology must be used to aggregate data in a uniform manner (Daswani et al. 1). This poses some additional obstacles. The process of sharing data requires some technical expertise, which not all research labs may possess or have the means to acquire. If a standard form of repository technology was chosen, research groups would have to be made aware of the technical infrastructure of the database. For instance, an open database may pose security challenges when attempting to keep data available, reliable, and anonymous. Since most behavioral neuroscience-related data deals with human subjects, data security is important.

All in all, these barriers are keeping researchers from openly sharing their data. Open data sharing initiatives in the behavioral neuroscience field are starting to overcome some of these obstacles, creating more support for the movement.

Chapter 3

Current Open Data Initiatives in Behavioral Neuroscience

Introduction

There are some open data initiatives that have recently begun in the behavioral neuroscience field. These initiatives can be divided into two main categories, based on the type of data being shared and the methods used to collect the data. These categories include: neuroimaging data and video data. Neuroimaging models can be divided into three distinct subtypes including: functional, neurochemical and structural. Functional neuroimaging models, which will make up the bulk of many behavioral neuroscience techniques include: functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) (Bandettini 260). Neurochemical methodology focuses on the structure and function of the nervous system through techniques including magnetic resonance spectroscopy (MRS), single-photon emission computed tomography (SPECT) and positron emission tomography (PET). Structural models include computed tomography (CT), magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI).

Many behavioral neuroscience labs collect and analyze data through video recordings. Video technology in behavioral research has proven to be beneficial in measuring complex behaviors, particularly those behaviors that occur over short periods of time. It provides many advantages including reproducibility, or the power to rewind and replay data, the ability to control the data collection session with a smaller chance for human error, and the capability of observing behavior from different perspectives (Haidet et al. 465)

Functional Neuroimaging Data Sharing

With regards to neuroimaging models, there are many different data sharing efforts for each of the distinct subtypes. Functional neuroimaging models are a major component of behavioral neuroscience research. Open sharing of fMRI data in particular brings many advantages. One advantage is that researchers are able to more easily examine trends and patterns of brain function across a number of datasets. This would also allow researchers to clearly determine connections between cognitive tasks and the patterns they observe using fMRI methods. By finding these connections and relationships between tasks and brain activity, researchers are able to make more robust interpretations and devise more comprehensive explanations of various mental processes. Furthermore, sharing data supports reanalysis of existing data with new and improved analytical methods. This would not only improve the quality of existing data, but will add to the value of behavioral neuroscience knowledge as a whole (Van Horn and Gazzaniga 314).

Open data sharing of fMRI data first began in 1999 with the fMRI Data Center (FMRIDC) which compiled 107 fMRI datasets and provided data for at least ten papers. Controversy over fMRIDC arose when two journals, the *Journal of Cognitive Neuroscience* and the *Proceedings of the National Academy of Sciences* required that researchers submit their data to the repository (Van Horn and Gazzaniga 314). Due to the number of outspoken concerns from various researchers, many of the journals got rid of the requirement. fMRIDC continued its existence, despite the lack of interest from the neuroimaging community. Eventually, however, the center discontinued its collection of new fMRI data sets. fMRIDC paved the way for other fMRI open data sharing projects including the International Neuroimaging Data-Sharing Initiative (INDI), previously known as the 1000 Functional Connectomes Project (Mennes et al. 683). INDI has compiled more than 5,000 fMRI data sets from research groups around the world. The unique aspect of the initiative is that it makes specific phenotypic information such as sex and age available for sharing. Beyond phenotypic data sets, INDI collected pre-publication data, as well, which allowed researchers to share data in parts or in full whether or not the data had appeared in print. These

new aspects demonstrated the value of fMRI data sharing in both replication and testing of emerging experiments (Mennes et al. 683).

The most recent sharing initiative related to fMRI data is the OpenfMRI project. This project promotes the dissemination of raw and processed neuroimaging datasets, focusing primarily on task-based fMRI studies (Poldrack et al., 2013). OpenfMRI was created to improve aspects of fMRIDC and INDI. For instance, the project organizes its metadata by specific research goals instead of displaying data sets for all possible variables of each experiment. This focus not only makes it more efficient for researchers when they upload and use data, but allows OpenfMRI to keep track of the type of data sets uploaded in an organized fashion (Poldrack et al., 2013). Furthermore, the database has made data easily accessible and downloadable over the Internet. The main goal of OpenfMRI is to facilitate whole-brain meta-analysis and thus any data set submitted to the project must include task-based fMRI data with emphasis on the whole brain. With regards to behavioral data, OpenfMRI organizes the data by trial variables where each file specifies the independent and dependent variables so that the data can be grouped with similar studies (Poldrack 59). Additional variables that are not trial-based can be inputted into each study and OpenfMRI are working on displaying the data using extensible markup language (XML). The database currently contains 28 data sets from over 10 different laboratories with 660 subjects across all datasets (Poldrack et al. 2013). Although many of the datasets are provided from the Poldrack Lab at Stanford University, OpenfMRI contains data from a wide variety of studies. As of March 2013, there have been 914 downloads from the database and four reported publications as a result of using datasets from the OpenfMRI database (Carp, 2012). OpenfMRI continues to accept uploads from researchers and is optimistic in its potential to provide insight to the field of behavioral neuroscience.

There are many projects dealing with EEG data sharing which include: the Bern-Barcelona EEG Database, Freiburg Epilepsy Data, and the Australian EEG Database (AED). AED is the leading the EEG sharing initiative. AED is a searchable online database which houses more than 18,500 EEG records from the John Hunter Hospital (JHH) (Hunter et al. 76). It was established in 2001 as a partner project between

the University of Newcastle and JHH. The database contains datasets from patients ranging from infants at 24 weeks through people who are over 90 years of age. The database has many advantages: AED allows researchers to more efficiently refer patients for specific studies, which shortens the time for recruitment significantly. Another advantage is the abundant number of datasets, which allows researchers to target specific criteria when selecting records for testing. AED is primarily used for projects relating to neonatal seizures and infantile spasms, but the diverseness and number of EEG data sets have helped researchers identify a variety of other rare conditions (Hunter et al. 76).

Neurochemical and Structural Neuroimaging Data Sharing

Neurochemical (MRS, SPECT, PET) and structural (CT, MRI, DTI) neuroimaging data are not as widely shared in the behavioral neuroscience field as compared to functional neuroimaging data. Neurochemical neuroimaging data in particular does not seem to have a prominent data sharing project. This is one area which may be difficult to promote due to the aforementioned barriers, particularly with regards to formatting issues. However, the structural neuroimaging data does have a large initiative for MRI data sharing, called Open Access Series of Imaging Studies (OASIS).

OASIS is sponsored by the Washington University Alzheimer's Disease Research Center headed by Dr. Randy Buckner. The project's primary focus is on the collection of MRI samples from demented and non-demented adults between the ages 60 and 90. OASIS distributes MRI data sets of the brain to promote new developments in behavioral neuroscience. The project has had a number of positive impacts on the behavioral neuroscience field. Firstly, OASIS encourages the exploration of detailed research studies to supply data that would be challenging to obtain from personal laboratories. Also, due to the wide range of ages and health conditions that the project houses, researchers use the MRI images to reassess current analytic techniques related to the human brain. Furthermore, data sets are commonly used as benchmark targets or point of references for various studies. This allows researchers to compare and

contrast their methods with methods from other laboratories. Lastly, the project can serve as an educational opportunity for various courses and workshops focusing on MRI images or even data sharing initiatives as a whole (Marcus et al. 2677).

There is much room for expansion in the field of behavioral neuroscience in terms of neurochemical and structural neuroimaging data sharing. Although these types of data sharing are in their beginning stages, projects such as OASIS may lead the way for future projects related to structural and more neurochemical-related data sharing.

Video Data Sharing in Behavioral Neuroscience

The study of people's behavior, as it relates to the structure or activity of the nervous system, requires the close examination of a number of factors including environment, time scale, etc. Due to the complexity involved in observing behavior, many researchers rely on video recordings of various behaviors in both experimental and natural settings to gather data. In fact, the average developmental lab records 12 hours of video a week (Tarr and Warren 1089). Video is often used as a common data collection method because it is simple to use and incorporate in a research lab setting and it allows researchers to effectively capture the complexity of behavior (Tarr and Warren 1089). Until recently, researchers in this field did not share video data often, which hindered advancements in behavioral neuroscience.

Databrary is an open sharing initiative which focuses on sharing video data (Adolph et al. 244). The project is heavily guided by peer input and it places a large emphasis on community building, among those in the behavioral neuroscience field. As of November 2014, Databrary has 53 authorized researchers, and 39 researchers who have registered and have applications pending. More than 100 development scientists who support the project's goal of open data sharing having written letters of support indicating their willingness to share video data (Adolph et al. 244). The repository strives to

overcome the challenges of video data sharing efforts. For instance, open video data sharing presents technical, financial, and privacy-related obstacles. Technical barriers are seen when sharing video files, as many files are not under the same format. Databrary helps reduce these technical difficulties by providing a free open source video coding program called Datavyu (datavyu.org) which allows the integration, analysis and visualization of video data in a common format. Furthermore, technical issues arise as few repositories have the capability of storing large video data sets. Databrary serves as a solution to this problem, as users are able to upload and download objects that are quite large (Adolph et al. 244). With regards to financial challenges, to keep an open data sharing project alive, there must be a funding initiative to provide for repository space, security, reliability of data sets, and other expenses related to housing and collecting data. Finally, the privacy barrier stems from the fact that participants on video must be anonymous and it is a much different process than when using images or simple text files. To deal with this issue, Databrary has created standard templates for seeking permission from people depicted in videos regarding their willingness to permit video data to be shared with other researchers. This framework builds on the existing foundation of participant permissions (Adolph et al. 244).

Having access to video data sets from various labs could disclose a lot about the experimental procedure of a study. Video not only allows researchers to replay a participant's actions, but enables researchers to observe a participant's behavior with much more detail. The physical and social environment, stimuli, facial expressions, voice tone and inflection, posture, and gestures can all be observed more easily and thoroughly through video data (Tarr and Warren 1089). This detail can provide a more profound understanding of a certain behavior as compared to photos, diagrams or verbal descriptions of experiments.

Databrary offers a method for researchers to explore, share, and discover new findings through video data. The repository enables behavioral neuroscientists to not only compare their findings with other lab groups, but encourages advancement in their research. Although it is still in its working stages—

the 1.0 version of the software was released in October 2014—Databrary is well on its way to having a positive impact on the behavioral neuroscience field.

Chapter 4

Conclusion

The Future of Open Data Sharing in Behavioral Neuroscience

The concept of open data sharing is becoming increasingly prevalent in today's society. Researchers in the social and natural sciences, business, and even the government are finding themselves under pressure to devise ways to share data openly. With regards to natural science, raw data from experiments is being shared among different laboratories. In business, financial statements and business practices are being shared among different companies. In the government (United States government in particular), various sectors have made their data openly available to the general public (Poline et al. 47). As mentioned previously, numerous advantages have been discovered from open data sharing in these fields. Due to these advantages, open data sharing is becoming a prominent initiative within many different areas. Although this growth is beginning to develop in the field of behavioral neuroscience, there are still many barriers that need to be overcome for open data sharing to have a significant impact on the field. One barrier is the publication process. Some journals in the field have stopped agreeing to store or publish supplemental materials such as data. One example of this is the *Journal of* Neuroscience, the flagship journal of the Society for Neuroscience (Poline et al. 47). As mentioned previously, the biggest obstacle is "buy in" from the behavioral neuroscience community. Without the support of those in the field, open data cannot become a success. The success of open data sharing requires the commitment and support of researchers to not only provide data sets, but to use those data sets to improve their own research practices and the field as a whole.

As society becomes more data-driven, open data sharing will need to play a larger role to fuel discoveries and advancements in research. The main method to increase buy in from researchers is to provide a reason for them to contribute their data.

Open Data Sharing Benefits the Bottom Line

The biggest question researchers ask about data sharing is, "What's in it for me?" This is a valid question as the opportunity for advancement in a researcher's field may not be the most convincing reason for contributing their data. There must be an incentive to ensure buy in. Thus, the business case for open data revolves around the idea of a return on investment. In other words, are the costs for open data sharing going to lead to higher benefits and profits? At the end of the day, behavioral neuroscience research is significant because it helps humans find new insight into the connection between the nervous system and behavior. However, there must be funding and financial capitalization reasons for researchers to develop these new discoveries. Unfortunately, it is difficult to monetize or put a dollar amount on research. Research cannot be capitalized as it is difficult to quantify its future benefits. This is primarily because research is a not fully considered an asset, as an asset is something that has future economic benefits. Thus, one of the best methods to acquire buy in from other researchers on open data sharing in behavioral neuroscience is to show how other researchers have benefitted from open data sharing in other fields.

In January 2014, the Yale University Open Data Access (YODA) project and Johnson & Johnson (J&J) made an agreement to allow scientists to gain access to the company's clinical trial data sets (Peart, 2014). Currently, J&J has shared pharmaceutical trial data and are planning to release clinical data on their medical devices and consumer packaged goods business. This event has sparked the acceleration of corporations and enterprises being more open and transparent (Peart, 2014). When other companies and organizations see a world renowned leader such as J&J begin to share data, and observe the advantages of

doing so, it encourages them to share data as well. Since the agreement with J&J, YODA has collected a significant amount of clinical trial data from various organizations, much of which is never published ("The YODA Project"). The project is dedicated to data transparency in hopes benefitting not only patients and the medical community, but society as a whole (Krumholz et al. 910).

It is evident that data sharing has benefitted other fields of research. Seeing these types of successes should encourage researchers to begin participating in open data sharing in the behavioral neuroscience field. The initiatives are beginning to grow in the field, but must gain more buy-in for them to become successful in providing the opportunity for new discoveries. Ultimately, the mission of researchers in behavioral neuroscience is to help society understand the biological basis of behavior. Open data sharing can help bring this goal closer to reality.

Bibliography

"Access vs. Importance." *Publishing Research Consortium* (2010): n. pag.

Adolph, Karen E., Rick O. Gilmore, Clinton Freeman, Penelope Sanderson, and David Millman. "Toward Open Behavioral Science." *Psychological Inquiry* 23.3 (2012): 244-47

Bandettini, Peter A. "What's New in Neuroimaging Methods?" *Annals of the New York Academy of Sciences* 1156.1 (2009): 260-93.

Carp, Joshua. "On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of FMRI Experiments." *Frontiers in Neuroscience* 6 (2012): n. pag.

Daswani, Neil, Hector Garcia-Molina, and Beverly Yang. "Open Problems in Data-Sharing Peer-to-Peer Systems." *Database Theory* 2572 (2003): 1-15.

Eisenach, James C. "Data Fabrication and Article Retraction." *Anesthesiology* 110.5 (2009): 955-56.

Glover, Gary H., Bryon A. Mueller, Jessica A. Turner, Theo G.m. Van Erp, Thomas T. Liu, Douglas N. Greve, James T. Voyvodic, Jerod Rasmussen, Gregory G. Brown, David B. Keator, Vince D. Calhoun, Hyo Jong Lee, Judith M. Ford, Daniel H. Mathalon, Michele Diaz, Daniel S. O'leary, Syam Gadde, Adrian Preda, Kelvin O. Lim, Cynthia G. Wible, Hal S. Stern, Aysenil Belger, Gregory Mccarthy, Burak Ozyurt, and Steven G. Potkin. "Function Biomedical Informatics Research Network Recommendations for Prospective Multicenter Functional MRI Studies." *Journal of Magnetic Resonance Imaging* 36.1 (2012): 39-54.

Haidet, Kim Kopenhaver, Judith Tate, Dana Divirgilio-Thomas, Ann Kolanowski, and Mary Beth Happ. "Methods to Improve Reliability of Video-recorded Behavioral Data." *Research in Nursing & Health* 32.4 (2009): 465-74.

Hendler, James, Jeanne Holm, Chris Musialek, and George Thomas. "US Government Linked Open Data:

Semantic.data.gov." *Intelligent Systems* 27.3 (2012): 25-31.

Hey, Anthony J. G., Stewart Tansley, and Kristin Michele. Tolle. *The Fourth Paradigm: Data-intensive Scientific Discovery*. Redmond, WA: Microsoft Research, 2009.

Hunter, M., R. L. L. Smith, W. Hyslop, O.A. Rosso, R. Gerlach, J.A.P. Rostas, D.B. Williams, and F. Henskens. "The Australian EEG Database." *Clinical EEG and Neuroscience* 36.2 (2005): 76-81.

Kaye, Jane, Catherine Heeney, Naomi Hawkins, Jantina De Vries, and Paula Boddington. "Data Sharing in Genomics — Re-shaping Scientific Practice." *Nature Reviews Genetics* 10.5 (2009): 331-35.

Kehagia, A. A., K. Tairyan, C. Federico, G. H. Glover, and J. Illes. "More Education, Less Administration: Reflections of Neuroimagers' Attitudes to Ethics Through the Qualitative Looking Glass." *Science and Engineering Ethics* 18.4 (2012): 775-88.

Kim, Dae Il, Dara S. Manoach, Daniel H. Mathalon, Jessica A. Turner, Maggie Mannell, Greg G. Brown, Judith M. Ford, Randy L. Gollub, Tonya White, Cynthia Wible, Aysenil Belger, H.jeremy Bockholt, Vince P. Clark, John Lauriello, Daniel O'leary, Bryon A. Mueller, Kelvin O. Lim, Nancy Andreasen, Steve G. Potkin, and Vince D. Calhoun. "Dysregulation of Working Memory and Default-mode Networks in Schizophrenia Using Independent Component Analysis, an FBIRN and MCIC Study." *Human Brain Mapping* 30.11 (2009): 3795-811.

Krumholz, Harlan M., Joseph S. Ross, Cary P. Gross, Ezekiel J. Emanuel, Beth Hodshon, Jessica D. Ritchie, Jeffrey B. Low, and Richard Lehman. "A Historic Moment for Open Science: The Yale University Open Data Access Project and Medtronic." Annals of Internal Medicine 158.12 (2013): 910.

Ling, Ding. "TWC LOGD: A Portal for Linked Open Government Data Ecosystems," *J. Web Semantics*, 9.3 (2011): 325–333.

Manolio, Teri A., Lisa D. Brooks, and Francis S. Collins. "A HapMap Harvest of Insights into the Genetics of Common Disease." *Journal of Clinical Investigation* 118.5 (2008): 1590-605.

Marcus, Daniel S., Anthony F. Fotenos, John G. Csernansky, John C. Morris, and Randy L. Buckner.

"Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults." *Journal of Cognitive Neuroscience* 22.12 (2010): 2677-684.

Mennes, Maarten, Bharat B. Biswal, F. Xavier Castellanos, and Michael P. Milham. "Making Data Sharing Work: The FCP/INDI Experience." *NeuroImage* 82 (2013): 683-91.

Minster, Bernard. "Data Services for Global Science." World Data System, n.d. Web. 17 Aug. 2014.

Molloy, Jennifer C. "The Open Knowledge Foundation: Open Data Means Better Science." *PLoS Biology* 9.12 (2011): *PLOS Biology*.

National Science Foundation. *Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans*. Arlington: National Science Foundation, 2010. Print.

Peart, Karen N. "Yale Program's Agreement with Johnson & Johnson Allows Broad Access to Clinical Trial Data." Yale News. Yale University, 30 Jan. 2014. Web. 17 Oct. 2014.

Poldrack, R. "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences* 10.2 (2006): 59-63.

Poldrack, Russell A., Deanna M. Barch, Jason P. Mitchell, Tor D. Wager, Anthony D. Wagner, Joseph T. Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael P. Milham. "Toward Open Sharing of Task-based FMRI Data: The OpenfMRI Project." *Frontiers in Neuroinformatics* 7 (2013): n. pag.

Poline, Jean-Baptiste, Janis L. Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O. Halchenko, Michael Hanke, Christian Haselgrove, Karl G. Helmer, David B. Keator, Daniel S. Marcus, Russell A. Poldrack, Yannick Schwartz, John Ashburner, and David N. Kennedy. "Data Sharing in Neuroimaging Research." *Frontiers in Neuroinformatics* 6 (2012): 47-55.

Potkin, S. G., and J. M. Ford. "Widespread Cortical Dysfunction in Schizophrenia: The FBIRN Imaging Consortium."*Schizophrenia Bulletin* 35.1 (2009): 15-18.

Tarr, Michael J., and William H. Warren. "Virtual Reality in Behavioral Neuroscience and beyond." Nature Neuroscience 5.Supp (2002): 1089-092.

Teeters, Jeffrey L., Kenneth D. Harris, K. Jarrod Millman, Bruno A. Olshausen, and Friedrich T.

Sommer. “Data Sharing for Computational Neuroscience.” *Neuroinformatics* 6.1 (2008): 47-55.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. “Data Sharing by Scientists: Practices and Perceptions.” Ed. Cameron Neylon. *PLoS ONE* 6.6 (2011): E21101.

"The YODA Project." The YODA Project. Yale University, n.d. Web. 17 Oct. 2014.

Van Horn, John D., and Michael S. Gazzaniga. “Databasing FMRI Studies.” *Nature Reviews Neuroscience* 3.4 (2002): 314-18.

Van Horn, John D., and Michael S. Gazzaniga. “Why Share Data? Lessons Learned from the FMRIDC.” *NeuroImage* 82 (2013): 677-82.

Wilhelm, Erin E., Emily Oster, and Ira Shoulson. "Approaches and Costs for Sharing Clinical Research Data." *Jama* 311.12 (2014): 1201.

Academic Vita

Matthew Lee

mzl147@psu.edu | 5115 Sunset Ridge Dr., Mason, OH 45040 | (814) 321-4026

Education:

The Pennsylvania State University, University Park, PA — August 2014-Present
Master in Business Administration Candidate, May 2016

The Pennsylvania State University, Schreyer Honors College, University Park, PA — August 2011-Present
5-Year Accelerated Science BS/MBA Program
Bachelor of Science in General Science, Honors in General Science

Experience:

The Clorox Company, Alpharetta, GA — June-December 2013
Global Strategic Sourcing Department Intern

- Developed program that efficiently assesses quality of supplier's products using IASTA software; currently being utilized by over 1000 suppliers
- Analyzed and managed changes in payment terms for over 700 suppliers as part of working capital improvement initiative, which resulted in over $18 million in savings
- Co-coordinated recruitment program at Fall 2013 Career Fair at Penn State to evaluate and interview potential interns/employees

Sportsmetrics, Cincinnati, OH — Summer 2012
Athletic Trainer Intern

- Under Dr. Frank Noyes, assisted in female ACL injury prevention research
- Assisted in organization and implementation of thirty 4-hour work-out sessions for forty female high school students

Vatsalya Children's Village Service Project, Jaipur, India — Summer 2012
Children's Orphanage Volunteer

- Developed and implemented lesson plans to teach English and basic science, mentored children, aided with maintenance and construction of local school

Activities:

Penn State Men's Club Soccer — 2011-Present

- NIRSA Region 1 Champions 2011, 2012, 2013
- NIRSA National Championship Tournament Elite 8 Finalists 2011, 2012

The Pennsylvania State University, University Park, PA — 2012-Present
Research Assistant, Department of Neuropsychology

- Under mentorship of Dr. Rick Gilmore study neuroscience of perception and memory among adolescents using EEG and MRI methods
- Writing thesis on advantages of open data sharing in behavioral neuroscience field