Brain Research Through Advancing Innovative Neurotechnologies® (BRAIN) Neuroethics Working Group (NEWG) Workshop on the Ethics of Sharing Human Brain Data Collected in Biomedical Research

National Institute of Neurological Disorders and Stroke

July 17-18, 2023

Virtual Meeting

Draft September 19, 2023

This meeting summary was prepared by Rose Li and Associates, Inc., under contract to the National Institute for Neurological Disorders and Stroke (NINDS). The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of NINDS. Contributions to this meeting summary by the following individuals are gratefully acknowledged: Bethany Stokes, Gina Castelvecchi, Dana Carluccio, Nancy Tuvesson, Nina Hsu.

Table of Contents

Acronym Definitions ..... **iii**
Executive Summary ..... **iv**
What are brain data? iv
Are there ethically salient differences between brain data and other types of biomedical data? v
How should the research community balance the benefits of data sharing with potential risks? v
When should strategies be applied that reduce risks, knowing that, in some cases, these strategies may reduce the value of shared data? vi
Meeting Summary **1**
Welcome and Meeting Overview **1**
Introduction and Background: Ethics of Sharing Human Brain Data Collected in Biomedical Research **1**
Panel 1: The data that is collected and stored by neurotechnologies and the inferences that can be made from these data **3**
Benefits and risks of sharing functional MRI data 3
Human EEG Data 4
Benefits and Risks of Sharing Structural Brain Imaging 4
BCIs and Data Sharing 5
Ethics of Sharing Individual-Level Human Brain Data Collected in Biomedical Research 6
Discussion 6
Panel 2: Inferences to be drawn from data and their implications for data sharing **7**
OpenBCI Technology and Survey 7
Decoding Language Using Noninvasive Brain Imaging 8
Implantable Neurotechnologies: Ethical Considerations in Neural Data Recording, Use & Sharing 8
Discussion 9
Panel 3: The potential risks to communities **10**
Confronting exclusionary practices within neuroimaging 10
Indigenous Data Sovereignty and Brain Data 11
Discussion 12
Panel 4: The potential risks to individuals of sharing different types of human brain data **13**
Risk of Reidentification in Neuroimaging 13
Complicating Notions of Individuals, Neuro Data, and Data Sharing 14
Discussion 16
Panel 5: Research participants' perspectives on sharing human brain data **17**
My Story and My Experience with BCI 17
Advancing Collective Rights, Interests, and Voices of Patient Communities in Healthcare Technology 18
Perspectives on Sharing Data in BRAIN Initiative Studies 19
Discussion 21
Breakout Room Reports **22**
Technological Developments to Support Ethical Data Sharing 22
Practices to Enable Participant Involvement in Data Sharing Decisions 23
Data Storage and Cybersecurity Practices to Prevent Unauthorized Data Use 23
Conclusion **24**
Appendix: Agenda **25**

Acronym Definitions

3D	three-dimensional
ABCD	Adolescent Brain Cognitive Development
AD	Alzheimer's disease
ALS	amyotrophic lateral sclerosis
BCI	brain-computer interface
BOLD	blood-oxygen-level-dependent
BRAIN	Brain Research Through Advancing Innovative Neurotechnologies
CARE	Collective Benefit, Authority to Control, Responsibility, and Ethics
CITI	Collaborative Institutional Training Initiative
COVID-19	coronavirus disease 2019
DBS	Deep Brain Stimulation
DMS	Data Management and Sharing
EEG	electroencephalogram
ENIGMA	Enhanced Neuroimaging Genetics Meta Analysis
FAIR	Findable, Accessible, Interoperable, and Reusable
FDA	Food and Drug Administration
fMRI	functional magnetic resonance imaging
GINA	Genetic Information Nondiscrimination Act
GPi	globus pallidus internus
GPT	Generative Pre-Trained Transformer
HCP	Human Connectome Project
HGDP	Human Genome Diversity Project
HHS	Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
HIV	human immunodeficiency virus
ML	machine learning
MRI	magnetic resonance imaging
NBDC	Native BioData Consortium
NEWG	Neuroethics Working Group
NFT	non-fungible token
NIH	National Institutes of Health
NIMH	National Institute of Mental Health
NINDS	National Institute for Neurological Disorders and Stroke
PHI	protected health information
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SENICa	Standards Ecosystem for Neutral Interface Consent and Neuroethics

Executive Summary

The Neuroethics Working Group (NEWG) of the Brain Research Through Advancing Innovative Neurotechnologies® (BRAIN) Initiative held a workshop on human brain data sharing on July 17 and 18, 2023. This workshop convened stakeholders from academia, government, industry, and nonprofit organizations, as well as patient advocates. The goals of this meeting were to (1) explore meaningful ways to categorize human brain data by potential risks of data sharing and (2) consider any resulting differences in how researchers should treat and use those data. Meeting participants heard presentations throughout five panel sessions, each followed by a moderated discussion: The data that is collected and stored by neurotechnologies and the inferences that can be made from these data (Panel 1), Inferences to be drawn from data and their implications for data sharing (Panel 2), The potential risks to communities (Panel 3), The potential risks to individuals of sharing different types of human brain data (Panel 4), and Research participants' perspectives on sharing human brain data (Panel 5). Each discussion, as well as the Breakout Group sessions, focused on specific questions posed by meeting facilitators and are summarized in the sections below:

What are brain data?

Brain data types can include (but are not limited to) neuroimaging and brain electrophysiological data, and both of these can be combined with other data types not directly obtained from the brain to enable researchers to make additional brain-related inferences. These various kinds of data have distinct sets of implications for patient privacy, health disparities, and other key issues.

Different neuroimaging modalities (e.g., structural magnetic resonance imaging [MRI], functional MRI [fMRI]) produce different data outputs whose sharing poses varied risks (e.g., reidentification or misuse). Structural MRI data enables anatomical segmentation of brain features and are often used for registration of other imaging data. fMRI data are used to study brain activity over time, either at resting state or during specific tasks.

Electroencephalography (EEG) is used to record electrical activity of the brain to infer diagnoses of neurological conditions or to characterize neural oscillations at the resting state or during specific tasks. More specific measures of individual neurons or small groups of neurons are achievable with intracranial sensors, such as implantable brain-computer interfaces (BCIs) and deep brain stimulation (DBS) devices with built-in sensors. Various inferences can be made from neuronal electrophysiology readings, including movement intention, language use, expected sensory perceptions, correlates of behavior, health biomarkers, sensory perception, cognitive states, affective states, memories, sleep habits, psychoactive medication use, and potential to develop future health conditions. Some of these inferences are more feasible than others, but continued technological developments and integration with other data types will enable further insights into specific conditions and more personalized medicine approaches.

Various other data types are collected during brain studies, such as behaviors, movement kinematics, gait kinematics, and computer use. When integrated with brain data, these data

types can result in identification of neural correlates of these data types, and thus the potential for prediction of future actions.

Are there ethically salient differences between brain data and other types of biomedical data?

Because brain data can be used to draw inferences about an individual's private thoughts and feelings, these data are particularly sensitive and can expose patients and populations to individual and community risks and harms if used inappropriately. In addition, the potential for predicting future conditions could result in discrimination and health insurance denials. Research participants have indicated general comfort with sharing their brain data with academic researchers and less comfort with sharing the same data with large corporations or the government. However, when a study integrates brain data with other particularly sensitive data types—such as HIV status, drug use, gene mutations with clinical implications, abortion history, and online activity—sharing these data could pose risks of both reidentification and the misuse of highly private information.

How should the research community balance the benefits of data sharing with potential risks?

Consider Inherent Risks of Different Data Types

Certain data types pose more severe risks to individuals and communities than other data types. To assess these risks, researchers should consider the ways in which data are *actionable*—in other words, if data were shared with bad actors, what could happen as a result? These possibilities are not necessarily limited by the characteristics of the data themselves: in some cases, the abundant promotion of brain data and resulting insights could result in bad actors attempting to nefariously obtain and use such data in ways that may cause harm to study participants themselves (e.g., hackers may use data to claim that have unsubstantiated data insights that impact patient wellbeing). However, mitigating these risks should also be balanced against the obligation of researchers to maximize the utility of study participant data to make study participation worthwhile.

Structural MRI data poses a risk of facial reconstruction and subsequent reidentification of a study participant, which would also associate the participant's identity with any other types of data collected during the same study. Defacing algorithms can reduce the risk of re-identification but also impact registration of other neuroimaging data and downstream analyses. Current research suggests that the downstream effects of defacing are minimal. Additional research is required to better understand the real risks of reidentification with structural MRI data, particularly when using very large datasets.

Develop Participant-Centered Data Sharing Models

Meeting participants discussed various data sharing models centered around empowering study participants to control the uses of their data. First, they emphasized the importance of promoting *data sovereignty* for study participants. The concept of data sovereignty restores data sharing autonomy to minority populations, particularly those from Indigenous

populations. The research community should provide Indigenous communities with the tools and expertise required to control secondary uses of community members' data.

In addition, meeting participants suggested the use of trusted data brokers, who can be enlisted to make decisions about secondary data uses on behalf of study participants. This model lessens the burden placed on individual study participants, who might otherwise need to consider reconsenting to each new use of their data. Study participants can select brokers based on shared values and practices, and patient advocacy groups could fill a void in trusted data brokers.

Improve the Informed Consent Process

The current research consent process is relatively static: once study participants consent to participate in research and to subsequent data sharing, their data can remain available indefinitely. This process typically uses boilerplate consent forms and lacks meaningful, informative, bidirectional discussions with potential study participants. In addition, the risks of data sharing may not be fully known at the time of initial consent. Education of the general public about current and future risks of research participation and data sharing can help bridge this gap in understanding while the informed consent process is more fundamentally revamped. Community organizations are well positioned to assist with this educational effort. To improve the consent process itself, consent forms require more granularity and specificity, including multiple separate consents for study participation and data sharing when possible. Researchers involved in the informed consent process should engage future study participants in meaningful discussion about the realities of data sharing risks as well as methods used to mitigate these risks.

In addition, meeting participants suggested a more dynamic consent process, in which study participants can revoke data sharing consent at any time. Dynamic consent could be modeled after the European Union's "right to be forgotten," which involves data destruction after a certain time period, mitigating future risks of new analytical technologies to study participants. In addition, rather than reconsenting for every secondary data use, study participants could establish preset preferences for data sharing, including allowable data uses, institution types (e.g., granting access to academic institutions but not to the federal government), and time limits for access. This approach reduces the overall burden on study participants while still honoring their wishes regarding future uses of their data. However, even with improvement of the consent process, researchers still need to implement additional safeguards to protect patients from risks of data sharing.

When should strategies be applied that reduce risks, knowing that, in some cases, these strategies may reduce the value of shared data?

Meeting participants suggested ways to prioritize strategies to reduce risks to study participants, even if those strategies may reduce the value of shared data. The risk of reidentification was not one of participants' main concerns: although they acknowledged that it is a primary concern for many researchers, most neuroimaging data would be relatively difficult to use for identification unless integrated with several other identifying data types. Moreover,

many patients have reported being comfortable with the use of their data for the advancement of science, noting some concern about reidentification but a willingness to participate for the greater goals of the studies.

Meeting participants emphasized that risk reduction strategies should focus on removing or lessening those risks and harms that are major concerns for underrepresented populations. Engaging with these populations is a priority because they have long been excluded from or missed during study recruitment, leading to a lack of understanding of how specific technologies or therapies work in specific populations. In addition, many of these populations have a deep mistrust of the research community due to previous conduct, such as using data for projects that violate the population's preferences. Building trust with populations is paramount and requires that researchers engage authentically, involve individuals from underrepresented populations on study teams, and work with community organizations and patient advocacy groups that have the patient's preferences in mind. Identifying methods to reduce risks and barriers of recruitment for each underrepresented population will improve their involvement in studies and that participation is worth potentially reductions in the value of certain data types.

Meeting Summary

Welcome and Meeting Overview

John Ngai, PhD, and Saskia Hendriks, MD, PhD, NINDS

Drs. Ngai and Hendriks welcomed meeting participants and emphasized that the goals of this meeting are to (1) explore meaningful ways to categorize human brain data by potential risks of data sharing, and (2) consider any resulting differences in how researchers must treat and use those data. The 2025 Brain Research Through Advancing Innovative Neurotechnologies® (BRAIN) Strategic Plan, released in 2014, was updated and enhanced in 2019 by: "The BRAIN Initiative 2.0: From Cells to Circuits, Toward Cures" and "The BRAIN Initiative and Neuroethics: Enabling and Enhancing Neuroscience Advances for Society." Collectively, these reports highlight the importance of developing guidelines for data sharing. The BRAIN Initiative has sought to forecast and proactively plan for needs related to sharing and using human brain data.

Dr. Hendriks presented four questions to guide discussions during the meeting:

- What are brain data?
- Are there ethically salient differences between brain data and other types of biomedical data?
- How should the research community balance the benefits of data sharing with potential risks?
- When should strategies be applied that reduce risks, knowing that, in some cases, these strategies may reduce the value of shared data?

Introduction and Background: Ethics of Sharing Human Brain Data Collected in Biomedical Research

Saskia Hendriks, MD, PhD, NINDS

Data sharing is a major priority across the biomedical research community. Data sharing can benefit many different groups, including the researchers who collect and share data, researchers who re-use the shared data, research participants, future patients, and society. Sharing research data collected can accelerate scientific advancement and improve quality of care for patients by enabling the generation of new hypotheses and metanalyses, increasing research transparency and accountability, enhancing collaboration and interdisciplinary research, maximizing the value of research (e.g., greater returns on investment and value gained from participant and researcher contributions), and improving trust in science.

The National Institutes of Health (NIH) recently released a Data Management and Sharing (DMS) policy to promote the sharing of scientific data. Under this policy, NIH encourages investigators and research institutions to (1) plan and budget for the management and sharing of data, (2) submit a DMS plan for review in funding applications, and (3) comply with the

approved DMS plan. The BRAIN Initiative has also released a data sharing policy, which requires that funding applicants submit their data to a BRAIN data archive for sharing, include specific required elements in the Resource Sharing Plan, and account for data preparation and submission costs in their application.

In addition to the benefits, data sharing also poses potential risks to individuals, communities, research teams, and institutions. Study participants providing data expose themselves to risks of psychological, socioeconomic (e.g., discrimination), and legal harms, as well as potential threats to privacy, personal and financial security, and dignity. Data sharing can also expose participants to unwanted attention and result in their data being used for purposes inconsistent with personal values.

Potential community harms may include the leveraging of data to fit private agendas, which could lead to a negative impact on health and non-health interests of groups (e.g., stigma, discrimination, or uses that the community finds objectionable) and undermine trust in research. Risks to sponsors, research teams, and institutions may include repercussions from data misuse, harm to economic interests related to intellectual property, and complications related to career advancement (e.g., lack of recognition for work). With these risks in mind, experts at a recent National Academies of Sciences workshop recommended sharing clinical trial data using approaches that maintain incentives to develop new therapies and facilitate future trials. However, some data sharing safeguards and protections, such as deidentification and controlled access, can reduce the scientific value of the data. Thus, researchers must identify methods to apply sufficient protections without unduly limiting scientific benefit.

Dr. Hendriks then presented the results of a survey of 397 investigators involved in active BRAIN, National Institute of Neurological Disorders and Stroke (NINDS), or National Institute of Mental Health (NIMH) research about sharing human research data in neuroscience. Most investigators reported collecting medical, survey/interview, behavior, and neural data. Most common types of neuroimaging data collected include functional magnetic resonance imaging (fMRI), structural brain imaging, and electroencephalography (EEG). Neural data included neural correlates of behavior, emotion, and decision-making, as well as predictive data. Researchers indicated that some studies included vulnerable participants, including minors, adults who cannot consent, and individuals with stigmatized conditions.

Approximately 29 percent of investigators reported that, after deidentification procedures, reidentification is "definitely not possible," whereas the majority of researchers indicate that it is "probably not" possible. In addition, most respondent investigators reported that perceived likelihood of someone using collected data to cause harm is "extremely unlikely" or "unlikely." When asked about the amount of harm that could be done with collected data, nearly 40 percent responded that they were "not at all concerned, and an additional 40 percent reported feeling "slightly" or "somewhat concerned." Researchers noted that scientists, clinicians, and industry professionals would be the groups most interested in human brain data. Additional interested parties include insurers, education systems, marketing companies, legal systems, employers, law enforcement, and foreign governments. Approximately half of investigators

indicated that certain types of data in their field should be exempt from data sharing to protect participants. For example, data related to HIV status, illegal drug use, gene mutations with clinical implications, and abortion history, as well as data collected without explicit consent, may require exemptions from broad sharing. Lastly, researchers emphasized the importance of developing best practices for data sharing protections.

Panel 1: The data that is collected and stored by neurotechnologies and the inferences that can be made from these data

Benefits and risks of sharing functional MRI data

Susie Huang, MD, PhD, Harvard Medical School

fMRI measures assess changes in blood oxygenation levels related to neuronal activity through neurovascular coupling. fMRI measures evaluate the functional activity of voxels in gray matter over time with both temporal and spatial resolution. fMRI methods can also measure thermal noise, physiological noise (e.g., cardiac, respiratory, and intravascular fluctuations), and subject-specific effects. Researchers can also use fMRI to generate task-based activation maps (i.e., areas of the brain activated after certain tasks relative to control state) and resting state maps of spontaneous low-frequency fluctuations in the blood-oxygen-level-dependent (BOLD) signal. Integration of fMRI data with other data types (e.g., anatomical magnetic resonance imaging [MRI], diffusion MRI, physiological, behavioral) can enhance the interpretation of fMRI data. fMRI data can help identify brain regions of interest that are active during perception, cognition, and action and can lead to discoveries of neural correlates of decision-making, attention, language, and emotions. However, one limitation to fMRI data is that the observed hemodynamic responses are only an indirect measure of neural activity. Other limitations of fMRI data include spatial and temporal resolution, as well as scan reproducibility. Sharing fMRI data can help provide researchers with larger sample sizes to achieve greater precision, address heterogeneity, and enhance reproducibility of findings. Because fMRI data are highly personal and could reveal sensitive information, researchers must consider privacy, misinterpretation, and stigmatization complications when sharing these data.

Dr. Huang and her team have developed Connectome 2.0, a next-generation ultra-high gradient strength human MRI scanner, which is optimized to study neural microstructures and connectional anatomy across scales. Connectome 2.0 surpasses current parameters of other comparable methods, including a maximum slew rate of 600 Tesla per meter per second (T/m/s), maximum gradient strength of 500 millitesla per meter (mT/m), 64-channel ex vivo whole brain coil, and a 72-channel head coil. These higher gradient strengths and slew rates enable higher spatial resolution to infer brain structure and function on the mesoscopic scale.

Question and Answer Session

Participants discussed whether MRI-based methods are resulting in data and inferences that are more identifiable than previous iterations due to higher resolution. Dr. Huang confirmed that methods optimization has led to increased resolution to enable inferences at the individual level, instead of the population level. Fast MRI techniques can record images on the order of

milliseconds, which enables researchers to view physiological variations that are not observable with conventional MRI methods.

Human EEG Data

Lorna Quandt, PhD, Gallaudet University

Human EEGs generate continuous waveforms representing the ongoing electrical activity in the brain and can be used to observe task-specific responses. Currently, EEG data is used to infer diagnoses of neurological conditions and characterize resting state- or task-related neural oscillations. Deep learning models can be applied to EEG data in order to identify unique characteristics of brain activity; however, such analyses could enable personal identification within or across studies. Typically, EEG data are used to authenticate study participants and perform emotional characterization. However, sharing EEG data has potential privacy risks, and the use of public data may lead to falsified authentication methods; these risks must be considered when sharing EEG data, particularly data from marginalized populations.

Question and Answer Session

Participants discussed identification and authentication using EEG data, noting that authentication is possible with current methods but identification of participants is not. Dr. Quandt added that more research is needed to optimize resting-state authentication methods.

Benefits and Risks of Sharing Structural Brain Imaging

Douglas Greve, PhD, Harvard Medical School and Massachusetts General Bingham Hospital

Structural neuroimaging provides researchers with three-dimensional (3D) visualizations of the whole head with 1mm resolution; these images indicate the state of the brain at a given time and place but can be obtained over time to observe longitudinal changes. Structural brain imaging enables the anatomical segmentation of brain features, allowing researchers to make inferences on the relationship of specific structures and disease. For example, researchers have identified that hippocampal volume is significantly lower in individuals with Alzheimer's disease (AD), compared to unaffected individuals. Structural imaging analyses can now diagnose AD with sensitivities and specificities of 95-98 percent. These approaches to diagnosis have been applied to many other diseases. However, many neurobiological diseases and disorders may lead to similar impacts on brain structure because the changes are subtle.

Dr. Greve highlighted several biobanks with robust MRI data, including the UK Biobank, the Human Connectome Project (HCP), OpenNeuro, and the Enhanced Neuroimaging Genetics Meta Analysis (ENIGMA) consortium. The benefits of sharing structural MRI data include increasing sample sizes (particularly for underrepresented subjects), improving machine learning (ML) methods, and enabling re-analyses to test new hypotheses and algorithms.

Another major consideration related to structural brain imaging sharing is identifiability because faces can be easily constructed using MRI data. Brain imaging data are normally deidentified prior to release, and one deidentification method used is defacing, which involves the changing of image pixels so that the original face cannot be recreated. However, other

information within MRI data could be used to identify an individual, including unusual demographics (e.g., an age of 110 years old) and metadata (e.g., date, place of scan). In addition, images of an individual's skull and teeth may also enable identification of an individual. All of these potential sources of identifiable information must be considered when releasing brain imaging data.

Question and Answer Session

Dr. Greve confirmed that current structural brain imaging methods can inform inferences about disease states and progression, but not about behavior or mental states.

BCIs and Data Sharing

Jennifer Collinger, PhD, University of Pittsburgh

Dr. Collinger's laboratory uses bidirectional sensorimotor brain-computer interfaces (BCIs) to record neural activity from the motor cortex. Dr. Collinger's team uses intracortical implant devices, which collect data from single or small groups of neurons, thus providing limited spatial coverage. Larger and more superficial devices, such as EEG, have broader spatial coverage but reduced resolution. The type of device used greatly impacts the data collected and resultant inferences. The intracortical devices collect neural data, device information, movement kinematics, computer usage statistics, and sensory responses. These data can be stored as raw voltage data or binned action potential data. Inferences that can be made with this data include movement intention, language, and expected sensory perception. Inferences that may be possible include environmental information, behavior, health biomarkers, cognitive states, actual sensory perception, and modulation of intention or affective state. Observing movement intention in the brain requires that researchers place electrodes in appropriate areas of the motor cortex and calibrate the BCI during structured tasks with defined movement-related goals (e.g., controlling a robotic arm, computer cursor, or speech decoding). While the device is used, researchers can decode movement parameters in real time and estimate movement intention. Transferring sensory information to the brain requires similar procedures, starting with the targeted placement of electrodes for stimulating neural activity. Then, researchers can deliver patterns of microstimulation to convey sensory information and document perceptual quality, psychometrics properties, and performance. These methods can be used to help restore movement and function to patients.

Privacy is a primary concern for the use of BCI data. Dr. Collinger noted that patients must have the right to choose which data are saved and with whom. Potential for harm is another concern because the future capabilities and possible nefarious intentions of BCI users can be difficult to predict. Further, the potential to reidentify patients through BCI-derived data is poorly understood because of the limited number of BCI studies to date.

Question and Answer Session

Dr. Collinger noted that the processed and binned action potential data contains most of the valuable information from the raw data that is needed to facilitate BCI goals (e.g., restore function to patients), but is less cumbersome to store than the large-scale raw data.

Ethics of Sharing Individual-Level Human Brain Data Collected in Biomedical Research

Doris Wang, MD, PhD, University of California at San Francisco

Dr. Wang's laboratory studies gait and balance impairments associated with Parkinson's disease (PD). These impairments include postural instability, freezing, and challenges with gait execution and adaptation. Very few current treatments for PD symptoms address gait and balance impairments, but levodopa treatment can improve these symptoms over time.

Gait is a complex and dynamic motor process that requires precisely coordinated movements across multiple limbs while maintaining balance. Neural control of this process is poorly understood. Thus, Dr. Wang and her team seek to identify signals associated with normal and pathological gait patterns and assess the ability of adaptive neurostimulation to restore normal gait patterns. In particular, they use electrocorticography paddle devices and globus pallidus internus (GPi) deep brain stimulation (DBS) methods. Field potential recordings obtained through DBS electrodes can measure neuronal synchronization, which can help determine the connectivity of specific brain regions. Dr. Wang's laboratory also uses force-sensitive resistors and a goniometer to measure gait kinematics (e.g., stride length, swing time); they also use wireless ankle monitors to collect these measurements in real-world environments. These technologies using sensors can recreate a 3D skeleton of an individual walking while capturing GPi and motor cortex data. Such data can help researchers understand how neural networks change during walking for an individual with PD. Dr. Wang and her team found that increased neural synchrony is associated with improved gait, particularly when an individual is receiving PD medications. They have also identified specific gait cycle biomarkers based on the swing phases of the right and left legs, as well as biomarkers that differentiate walking versus sitting. Future directions for this work include decoding more features of walking (e.g., stride time, step length, symmetry) from neural signals and using technologies to better understand intention to walk, stop, and turn, as well as lower body motor control and coordination.

Sharing data collected from these studies is important to support scientific rigor and reproducibility, as well as direct-to-consumer applications, and to promote scientific progress and discoveries on the mechanisms of brain function in both health and disease. Risks to sharing such data include the disclosure of personal information, reverse inference of an individual's cognitive and affective states, and discrimination.

Question and Answer Session

Dr. Wang confirmed that, in her laboratory, some data is directly uploaded to a Cloud environment, but some are only uploaded once a patient device is connected to a computer.

Discussion

Moderators: Sameer Sheth, MD, PhD, Baylor College of Medicine; Nina Hsu, PhD, NINDS

Patient Education

As patients become more involved in clinical studies (e.g., recording data from home), researchers must provide more education regarding how their samples and data can be used, as well as the risks and benefits associated with participation in a study.

Data Integration and Reidentification Risk
Data is increasingly collected in multimodal fashions and integrated with other data types (e.g., behavioral, demographic) to increase the utility of these datasets for data discovery purposes. Without linking multiple data types, neuroimaging and neurotechnology data have limited applicability. On the other hand, linking data types can also increase the risk of reidentification.

Level of Access
Participants noted that risks associated with data greatly depend on how these data are being shared. Significantly more risk is associated with sharing data publicly than in an access-controlled platform with regulations that users must follow.

Virtual Reality
Virtual reality (VR) devices have gained popularity for a variety of functions. These devices capture eye-tracking data but many device users do not realize they are providing such data to the device's maker. In the future, companies may integrate EEG sensors into VR technologies, which could provide valuable data but also pose potential privacy risks.

Panel 2: Inferences to be drawn from data and their implications for data sharing

OpenBCI Technology and Survey

Conor Russomanno, OpenBCI

Mr. Russomanno and his team facilitate studies aimed at tracking emotions and activities. Previously, he developed ROB3115, a neuro-immersive narrative game that measures the level of user engagement with the game to result in enhanced consciousness of the humanoid robot main character (i.e., more engagement enables the character to be more human, whereas disengagement defaults the character to robot characteristics). Following graduate school, Mr. Russomanno and colleagues established OpenBCI to build open-source tools for the neuroscience field. To date, OpenBCI has built hardware and software for neural interfacing and has sold more than 40,000 units across more than 100 countries, resulting in over 300 citations. Recently, the company has focused on integrating its technologies into VR and augmented reality (AR) methods. Future directions in this field are to create and refine closed-loop neural optimization technologies that receive both conscious input and feedback, as well as subconscious input and feedback. Mr. Russomanno emphasized that cognitive liberty must be considered the primary design constraint when developing new tools.

When releasing their Galea hardware and software platform, OpenBCI also facilitated a survey on the use of BCIs and biometrics. Approximately 85 percent of respondents noted that they use devices that leverage biometrics for security and 53 percent reported feeling more secure using biometrics. The majority of respondents (69 percent) reported being highly likely to use biometrics to create new device experiences or improve mental and physical health. Even more respondents (73 percent) noted interest in using devices that leverage brain waves to improve health, compared to 23 percent that reported no interest. However, most respondents (72

percent) also reported concern about using biometrics and this percentage increased to 81 percent when asked about the use of brain wave-related biometrics data. Respondents indicated most concern about specific private companies using their data (Facebook, Google, Apple, Microsoft, Amazon), followed by other entities, including the government, insurance companies, and doctors.

Decoding Language Using Noninvasive Brain Imaging

Alexander Huth, PhD, University of Texas at Austin

Dr. Huth's laboratory focuses on decoding language without movement using noninvasive brain imaging data and technologies. In one study, Dr. Huth and his team assessed individuals via fMRI while they listened to 75 stories from The Moth Radio Hour, and the research team used these collected data to fit voxelwise encoding models. They use Generative Pre-Trained Transformer (GPT) technology to review the story being played to predict what may happen in the human brain. Decoding language involves the identification of stimuli given the fMRI responses; however, building decoding prediction models can be challenging. The challenges of building these models can be overcome using Bayesian decoding methods, which enables the use of an encoding model and language model. Dr. Huth presented several examples of transcripts from The Moth Radio Hour and the decoded stimulus, noting that the model's word error rate was high (90 percent) but some key exact words and phrases were maintained. Overall, the quantitative performance is highly significant for all metrics, particularly for meaning-oriented metrics such as the BERTscore. This study led Dr. Huth to consider whether researchers can also decode language from imagining a story or movie watching. In order to test these theories, training data is required for each individual subject.

Overall, neural network language models, like GPT, are very effective at predicting cortical responses to natural language. These models can already decode language from fMRI signals and this ability will improve as better language and encoding models, as well as closed-loop decoders, emerge.

Question and Answer Session

Dr. Huth noted that the GPT model processes both individual words and phrases.

Implantable Neurotechnologies: Ethical Considerations in Neural Data Recording, Use & Sharing

Michael Young, MD, Massachusetts General Hospital

Between 2008 and 2016, more than 10,000 patents were executed for neurotechnologies, as well as several U.S. Food and Drug Administration (FDA) approvals for clinical use (including DBS and BCIs). Implantable neurotechnologies can facilitate several functions, including:

- Recording/sensing (e.g., capturing information about or from the nervous system)
- Stimulating/actuating (e.g., stimulating or modulating the nervous system)
- Closed-loop interactions (e.g., combining recording/sensing and stimulating/actuating to control the nervous system)

- Direct physical and biological modification (e.g., physically altering the nervous system by modifying physiology)
- Augmenting/facilitating (e.g., supporting or expanding the existing functions of the nervous system)

Using an implanted sensor and associated computing program, BCIs aim to facilitate neurorecovery and neurorehabilitation to improve various functions, such as movement and communication. Currently, BCIs are being investigated in several clinical trials for various diseases, including locked-in syndrome, quadriplegia, and amyotrophic lateral sclerosis (ALS). Ethical considerations related to data privacy for each of these clinical trials should account for decision-making capacity, personal identity, health disparities, and access to technologies.

New technologies have enabled researchers to collect data at higher resolutions and spatial coverages, collect data from neurons across brain regions, and capture more data overall. With these improved technologies, intracranial neurophysiology data may reveal information about an individual's memories, preferences, mood, sleep habits, psychoactive medications, beliefs, likelihood of future conditions, and more. Combining neural data types poses potential privacy complications, and forecasting what other privacy complications may arise in the future is difficult. However, Dr. Young emphasized that the neurotechnology field is discussing ownership of data, as well as reporting standards for *in vivo* research. Major principles of neuroethics to consider for future studies include (1) autonomy/agency (e.g., bespoke models of informed consent, participant-centered approaches to neural data ownership), (2) beneficence/non-maleficence (e.g., compensation for data collection, sharing expected and unexpected effects of data collection, post-trial responsibilities), (3) justice (e.g., equity in access to neural data, discrimination, ownership, privacy), and (4) philosophical issues (e.g., impact of discoveries, responsibility, impact on understanding of volition, reconciliation of physical and phenomenal experiences). Future directions in these areas include the Standards Ecosystem for Neural Interface Consent and Neuroethics (SENICa), which aims to create an open-access laboratory with broadly accessible tools and ethics responses to foster responsible translation of neurotechnologies and data sharing across the development life cycle.

Question and Answer Session

Dr. Young noted that intracranial neurophysiology data can shed light on a variety of philosophical dimensions (e.g., meaning to be human, self-personhood, identity), particularly when combined with other types of data, such as behavioral data.

Discussion

Moderators: Jim Eberwine, PhD, University of Pennsylvania; Saskia Hendriks, MD, PhD, NINDS

Challenges

Participants acknowledged that reidentification using brain data may not be entirely preventable. Participants also emphasized that more significant challenges may be related to the current informed consent process and potential uses of data that do not align with patient preferences.

Sharing Data and Findings with Study Subjects
In addition to sharing data broadly within the research community, meeting participants also emphasized the importance of returning data to individuals who donated data and samples, as well as sharing discoveries made using individuals' data. Transparently discussing future uses of data, as well as any uncertainty of these uses, with study participants is also critically important.

Safeguards
One participant noted that language translators are required to accurately convey the meaning of the communication they are translating without bias, adding that such safeguards should be incorporated into newer neurotechnologies (e.g., BCIs).

Panel 3: The potential risks to communities

Confronting exclusionary practices within neuroimaging

Jocelyn Ricard, Yale University

The vast majority of neuroscience study and clinical trial participants are non-Hispanic white participants, with less than 10 percent of recruited individuals representing other races and ethnicities. These disparities are also evident when reviewing large-scale neuroimaging collection efforts, such as the UK Biobank. The UK Biobank has collected data on nearly 500,000 individuals, of which 94 percent are white. Black individuals compose only 1.6 percent of this repository's sample size, with Indian individuals representing 1.2 percent, and remaining groups representing less than 1 percent. Further, the HCP and the Adolescent Brain Cognitive Development (ABCD) studies are also primarily comprised of White Americans (76.1 percent and 56.0 percent, respectively), with only 13.6 percent and 11.9 percent representing Black participants. Ms. Ricard emphasized that these large-scale data collection efforts must begin overrecruiting individuals from underrepresented groups.

In addition to recruitment efforts, neuroimaging methodologies are also not equally accessible to all races and ethnicities. For example, MRI methods are not as amenable to Black individuals with natural hairstyles, such as an afro or hair styled with metal tracks. Braids and other hairstyles commonly worn by Black individuals can cause MRI images to display significant artifacts because of materials, such as clay or beeswax with high iron oxide content, used in the braiding process—these artifacts can significantly hinder data analyses and model predictions. Studies are underway to optimize EEG electrodes for use on individuals with coarser hair.

Disparities within recruitment and methodologies must take greater priority within the neuroscience research community. From the 1970s to the 2010s, only 5 percent of publications highlighted racial demographics of participants (1,511 of 26,380). Ms. Ricard emphasized the importance of providing study participants with accessible and diverse study materials, as well as removing methodological and financial barriers to study participation, when possible. For example, one study found that providing study participants with rideshare services led to an increase in recruitment and retention of participants from underrepresented populations. Overall, such barriers to recruitment and involvement in studies must be addressed more broadly throughout the neuroscience research community.

Question and Answer Session

Ms. Ricard emphasized that the full extent of current exclusionary practices is likely not known but may have downstream consequences. For example, one study used prediction models to identify findings related to suicide risk for White and Asian individuals, but not Black or Native American/Alaska Native individuals. These findings could inform new interventions for some groups and not others, amplifying inequities. Exclusionary practices are evident in all types of neuroscience and imaging studies and thus new practices to mitigate these complications must be developed. Ms. Ricard further emphasized the need for more representation of underrepresented groups within research teams to help gain trust within the community.

Indigenous Data Sovereignty and Brain Data

Krystal Tsosie, PhD, MPH, MA, Arizona State University

Despite efforts to increase diversity in genomic studies over the past decade, Indigenous people still constitute less than 1 percent of research participants. The reasons underlying why Indigenous individuals may choose to not engage in research are complex and linked to distrust, particularly factors that are cultural (e.g., differences in Indigenous and western values, concerns related to donating biospecimens, and lack of interest in research by non-Indigenous researchers) or political (e.g., "blood quantum," unforeseen effects on federally recognized rights, reidentification, stigmatization, racial profiling, and biopiracy and biocolonialism of Indigenous biomarkers). During the onset of several large-scale genome diversity projects, Indigenous peoples have cited concerns about co-optation and biocolonialism and asked the United Nations to halt these projects. In addition, the Human Genome Diversity Project (HGDP) recently published on its successful recruitment of 51 global populations, including Indigenous groups; however, U.S.-based Indigenous groups were not sampled and the Indigenous DNA samples obtained were extracted from disempowered global communities. Indigenous peoples who participated in the HGDP reported feeling "duped, lied to, and exploited," while not receiving the incentives (i.e., medicine) that led them to join the study.

Indigenous genomic sovereignty is the right of Indigenous people to exercise autonomy to protect their interests related to genomic data, and this right is intrinsically defined by the community, not colonially. Because most Indigenous peoples are not engaging in genomic studies, the majority of indigenous data in current projects may have been extracted from legacy samples that pre-date Tribal Research Review Boards, unprotected groups, Indigenous peoples that live outside the Tribal jurisdictions, and public health datasets with few opt-out options. Thus, Dr. Tsosie emphasized the need to rethink informed consent procedures, which are inadequate for Indigenous communities and other small and easily re-identifiable groups. Individuals from these vulnerable populations should be fully informed about how genomic information can be used, and risks associated with research engagement must be adequately disclosed. In one study, Tribal community members express the most concern regarding how unconsented data would be used in research studies. One solution is to institute more organizations like the Native BioData Consortium (NBDC), which enables Indigenous participants to deposit their DNA and store it in an Indigenous-led repository on Tribal lands. Another example is the Australian Traumatic Brain Injury National Data Project, which is using

knowledge interface methods to ensure Indigenous data sovereignty and that the concerns and interests of Indigenous peoples are adequately considered.

In addition to informed consent procedures, study designs must also be re-evaluated to ensure that one Indigenous person participating in a study is not treated as representative of an entire tribe, because tribes can be heterogeneous and can encompass multiple sub-identities. Study designs must employ Findable, Accessible, Interoperable, and Reusable (FAIR) and Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) principles. Researchers need to (1) ensure that studies do not harm Indigenous communities, (2) partner with those communities, and (3) caution against making broad generalizations about Indigenous peoples without adequate sampling and cognizance of heterogeneity and the fact that Indigeneity is a sociopolitical construct—not a biological one.

Question and Answer Session

Dr. Tsosie noted that, even just five years ago, conversations about improving engagement with diverse populations were nonexistent. One notable driving force behind these conversations are new Indigenous researchers and scientists from underrepresented groups joining the neuroscience research workforce.

Improving informed consent processes aims to build trust but also enhance education related to the use of patient samples and data. NBDC, for example, enables study participants to receive periodic summaries that describe how their samples were used and also update their preferences for the use of their data as well. NBDC also provides educational materials to study participants to help them better understand their role and rights in the research process.

Discussion

Moderators: Nita Farahany, JD, PhD, MA, Duke University; Saskia Hendriks, MD, PhD, NINDS

Community Harms

Participants discussed potential community harms, particularly those related to possible stigmatization, discrimination, and continued exploitation of underrepresented communities. The criminal justice system has begun using samples and new technologies to identify perpetrators of crimes; participants noted concern that these practices may eventually expand to other systems, enhancing discrimination and stigmatization. Dr. Tsosie noted that one major community harm is the retention of samples from Indigenous groups who believe that an individual must be whole before entering the spirit world. Initiation of several programs successfully shared samples back with respective study participants, but others have not been as successful. For example, the O'odham (also known as Pima) people engaged in a diabetes-focused study with NIH and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) but later withdrew themselves from the project due to a lack of perceived benefit and concern of potential harm. However, researchers continued to use samples collected through this study for decades after withdrawal. One possible mitigation strategy to help trust between researchers and underrepresented groups involves connecting underrepresented groups with community organizations. These community organizations can

provide education and guidance to help underrepresented groups make informed decisions about research participation and data sharing.

Transparency
Dr. Tsosie noted that transparency about sample collections can be complex. For example, in the paleontology field, museums have expressed a reluctance to share their collection inventories because descendant communities may attempt to regain their ancestral artifacts. This tension may also exist with brain biobanks.

Characterization
Dr. Tsosie recommended that researchers reevaluate how different racial and ethnic groups are characterized and whether the reference groups used for comparisons are also appropriately characterized. Many groups may be more heterogenous than researchers have accounted for within their studies.

Education
Many research institutions require staff and students to participate in trainings related to research, ethics, and compliance (e.g., Collaborative Institutional Training Initiative [CITI] Training). Meeting participants recommended that staff and students be required to take more trainings focused on ethics, as well as trainings suggested by specific underrepresented groups.

Panel 4: The potential risks to individuals of sharing different types of human brain data

Risk of Reidentification in Neuroimaging
Russell Poldrack, PhD, Stanford University

Utility theory can frame discussions of risk of reidentification as a function of the likelihood of reidentification and the magnitude of potential harm. In the context of brain imaging, facial structure information from structural MRI data can enable 3D rendering of the head and face. However, the quality of these renderings depends on the quality of the structural MRI data. Researchers can use defacing tools to erode parts of facial structure information from MRI data to prevent 3D renderings of facial reconstructions that can result in reidentification. However, deep learning algorithms can impute some of this missing facial structure data, which may result in 3D renderings capable of reidentification. While research studies that collect structural MRI data do not usually collect facial photos, personal photos may still be available elsewhere for matching with 3D renderings.

Dr. Poldrack noted that published research likely overestimates the likelihood of reidentification from structural MRI data. Current publications use relatively small datasets and report a reidentification rate of 28 to 38 percent, depending on the initial defacing method used. Therefore, Dr. Poldrack is generating larger structural MRI datasets to test reidentification accuracy with artificial neural networks and assess whether reidentification accuracy declines in larger datasets. A simulation of 500,000 research participants results in a reidentification

accuracy of less than 1 percent. Therefore, Dr. Poldrack considers the risk of identification low enough for structural MRI data to be classified as non-identifiable.

Potential harms of reidentification depend on the types of data paired with structural MRI data. For example, in OpenNeuro, participant data includes brain imaging during performance of a cognitive task with little additional metadata; thus, the potential harms of reidentification are relatively low. Dr. Poldrack considers claims in current literature regarding the ability to decode future disease risk or cognitive ability from raw data are significantly inflated due to problematic research practices. The greatest potential for harm to research participants is likely the emotional harm that could arise if someone were to reidentify and then blackmail them with overhyped, inaccurate claims of decoding actionable information.

Notably, pairing additional data types with MRI data can increase the risk of reidentification and contain information about diagnoses that can cause reputational harm, as well as insurance and employment discrimination. Rather than immediately heightening data restrictions—which can reduce the impact of shared data and reduce reproducibility—Dr. Poldrack recommended addressing these potential harms through regulatory means to prevent misuse of neuroscience data, similar to the Genetic Information Nondiscrimination Act (GINA) for genetic information.

Question and Answer Session

Unlike DNA or fingerprints, brain data cannot be collected surreptitiously. An existing connectome is required to reidentify someone based on brain data that does not include 3D facial renderings, and relatively few connectomes exist.

Certain defacing methods can impact co-registration of data and downstream secondary analyses. Earlier analyses suggested small effects of defacing on cortical thickness and other downstream measures, but more recent analyses did not indicate significant impacts of defacing on downstream analyses. Defacing can impede generation of head models needed for other measures such as EEG. Therefore, researchers may still need to share intact structural MRI data in a more controlled way or pre-generate these head models. A more powerful strategy for preventing reidentification from structural MRI involves skull stripping, but these methods would further restrict co-registration of additional data types and should only be used if absolutely necessary.

Complicating Notions of Individuals, Neuro Data, and Data Sharing

Sara Berger, PhD, IBM

Dr. Berger's research using various data types (e.g., fMRI, smartphone assessments, wearable technologies, at-home sensors, voice analytics, ML techniques) to track, predict, and treat pain has enabled her to observe a variety of ethical, social, and legal issues, particularly related to data sharing. When she and her team reflect on broader implications of their research and develop resources to address these implications, they frame these discussions based on assumptions made when using terms like individuals, brain data, and data sharing. In challenging these assumptions, Dr. Berger suggests expanding the overall definition of harms that can result from brain data sharing.

In research studies, individuals are assumed to be neatly separated from groups or communities, but these individuals are often simultaneously situated within various communities and social groups. Data on these communities and groups (e.g., gender, race, education, age, income) are collected in demographic forms and used to classify participants for downstream comparison to reference datasets. In addition, researchers assume that brain data can stand alone, separate from other data and societal factors. However, various political, historical, and socioeconomic factors dictate what data are worth generating. Even when discussing the process of data sharing, stakeholders consider data sharing as separable from larger data processes and social systems. However, any data being shared were at some point collected and used for other purposes. Therefore, Dr. Berger views data sharing as a socio-technical system of sharing when developing intervention strategies to improve research, data quality, and harm prevention.

Dr. Berger provided examples of harms that may arise due to data sharing. Study participants who consent to data sharing are exposed to risks of unintentional uses of their data. For example, a participant may consent to a biomarker study, and these data could be used by another group to advocate for using the biomarker to detect a pre-existing condition to deny patient insurance or deny worker's compensation. In addition, data sharing policies and use of shared data can result in harms to broader communities. Shared data may consist of cleaned data, down-sampled data, select data types, or these data may only contain data from certain individuals due to missing data, ineligibility, or outlier values. Therefore, secondary uses of these data may lack important context. In addition, researchers may only share initial inferences of these data, and sharing of these inferences may not be covered under initial informed consent. A common practice in secondary data analysis involves aggregation of multiple datasets which can enable harmful or stigmatizing inferences not covered by original consent forms.

To combat potential harms of data sharing, Dr. Berger proposed the following actions: (1) increase transparency of data use in consent forms and processes; (2) assist study participants with understanding nuances of data and potential consequences of data sharing; (3) provide researchers with additional data sharing training; (4) uphold existing strong strategies to secure data and regulate access but remain open to adjusting standards; and (5) provide opportunities for researchers to provide feedback and report concerns about secondary data use. Data repositories can consider data use restrictions that require formal requests and limit duration of data access. She concluded her presentation by posing two questions to frame future data sharing discussions:

- Are we trying to protect data or people?
- Is a data type itself risky or harmful, or are the methods for data creation and use responsible for risk and harm?

Question and Answer Session

Dr. Berger emphasized that data sharing protections should focus on data collection and use rather than data type, and considerations of data use will become increasingly important with

the aggregation of different data types, including those from social media. In addition, researchers need to better understand what data and for what purposes people are comfortable with data sharing. Dr. Berger indicated that her main concerns with data sharing involve unsubstantiated claims using under-sampled datasets and the potential scalability of any harms that may result.

Discussion

Moderators: L. Syd M. Johnson, PhD, SUNY Upstate Medical University; Saskia Hendriks, MD, PhD, NINDS

Data Usage Controls

Datasets that require formal requests for access can still be defined as open datasets. Rather than complete unrestricted data use, researchers need to carefully consider and justify specific data uses. Therefore, researchers require additional training on regulatory practices and data sharing ethics as well as funding incentives for proper data use practices.

However, too many data restrictions and limitations can result in decontextualization of datasets, which impacts research findings. Data restrictions in the European Union have resulted in minimal data sharing and frustration in the scientific community. Impeding data sharing to this degree impairs the ability of other researchers to compare results to confirm validity across larger sample sizes.

Risks of Sharing Structural MRI Data

Analyses of larger datasets are required to better understand how realistic the risk of study participant reidentification. In addition, the feasibility and predictive values of deriving actionable data from structural MRI, such as AD risk, remain unclear. Overestimating of the information derivable from imaging data can result in unsubstantiated harmful claims and applications of these data. For example, whereas researchers have identified biomarkers that enable fairly accurate ascertainment of acute thermal or pressure pain, the utility of these biomarkers in assessing chronic pain remains unclear. However, despite the lack of scientific support, certain individuals are already asking to use these biomarkers for chronic pain measurements in courtrooms, which can result in real-life consequences.

Use of Consent Forms to Communicate Current Risks of Data Sharing

Current consent forms are insufficient to obtain true informed consent. Consent forms should be more study-specific and more granular to allow participants to opt in or out of data use for specific purposes as well as submission of data to specific repositories. While dynamic consent would be difficult in the United States due to the current data sharing landscape and research practices, this type of consent would allow participants to change their preferences regarding authorized uses of their data. The overall consent process needs to be explicit about data use and storage and should include conversations between investigators and participants to address participant questions and concerns.

Use of Consent Forms to Communicate Future Risks of Data Sharing

Data stored and available in repositories after conclusion of the original study can be analyzed through new approaches in the future. Consent form language can provide transparency about potential future uses and actions used to prevent reidentification, but this language cannot guarantee that shared data will never result in reidentification or improper data use. One strategy to prevent the application of technologies developed in the distant future to a participant's data would involve data destruction after a certain amount of time, similar to the European Union's "right to be forgotten." However, first-level inferences made from these data may still exist beyond destruction of the original data.

Future Directions for Studying Data Sharing Consent

To make more informed decisions on data sharing policies, future studies are required to track secondary data uses and understand individual participant and community attitudes relevant to specific uses of their data. Notably, research participants are relatively open to data sharing with other researchers for other uses, but many want to know where their data goes and who is using it. Researchers should also document data sharing and consent choices for studies, as well as link data in repositories to original consent forms. Providing researchers with these consent forms will enable reflection on whether a secondary analysis conflicts with original participant's consent.

Panel 5: Research participants' perspectives on sharing human brain data

My Story and My Experience with BCI

Nathan Copeland, Research Participant

Mr. Copeland suffered a catastrophic car accident, resulting in quadriplegia, and he is a research participant and the longest chronic BCI holder in the world. His participation in BCI research has enabled him to shake hands with former president Obama, contribute his experiences and insights to research conferences and publications, operate a robotic arm, and perform cursor-based computer tasks, including drawing and playing video games. He has even been able to sell some of his digital artwork as non-fungible tokens (NFTs).

Mr. Copeland's BCI consists of electrodes that report the activity of 256 individual neurons, and these activity readings are used by the motor cortex connections of the BCI to enable control of a robotic arm. His BCI delivers microstimulations based on inputs from the robotic arm that elicit sensations of pressure, tapping, or warmth to his own hand. These sensory inputs enable better control of the robotic arm via sensory feedback. While the system that enables the robotic arm is too large to use at home, BCI researchers provided Mr. Copeland with a portable system that decodes neuronal activity from his BCI to enable use of a computer cursor.

Mr. Copeland overall is extremely passionate about technological development, especially because it enables him to perform tasks that previously were not possible. He indicated zero concerns about secondary uses of his neuronal activity data, especially because the decoder system needed to understand those activity readings requires daily training and these data likely will not cause him any harm. In addition, he believes that because the BCI interacts with only a small portion of his brain, meaningful insights that could be drawn from these data are proportionally limited. Mr. Copeland's main concern was related to the pairing of his BCI readings with internet activity.

Question and Answer Session

Mr. Copeland explained that because the inclusion criteria for his BCI study is extremely narrow and few eligible individuals are even aware of the study, he felt an obligation to participate and share his data to help accelerate science forward. This participation gave him a new purpose to combat the despair that emerges when a person is told they cannot feasibly do activities they love anymore. Therefore, he is content with his data being shared with anyone and used for any purpose, even though he does not know what applications are possible in the future. A blanket consent form for data sharing was sufficient for Mr. Copeland. While he would not want to consent each time his data was shared, he would be interested in a notification system to inform him of how other researchers were using his data.

Advancing Collective Rights, Interests, and Voices of Patient Communities in Healthcare Technology

Christine Von Raesfeld, People with Empathy

People with Empathy is a patient advocacy and allyship organization active in the rare and chronic disease spaces. This organization is defining technology and digital rights for patient communities and representing the collective rights, interests, and voices, of patient communities. Ms. Von Raesfeld shared several negative experiences due to a lack of data sharing related to her chronic illness. First, she participated in a lupus study but received only raw data that was inaccessible to her and her doctors. Years later, someone managed to interpret these data and found that Ms. Von Raesfeld had multiple mitochondrial mutations that would have been informative to her health care team. In addition, researchers generated genomics data using Ms. Von Raesfeld's samples during her participation in various studies. These data were not readily shared with her care team, which when paired with the results of a pharmacogenomics analysis, would have revealed differences in Ms. Von Raesfeld's metabolism of certain drugs, which had been administered to Ms. Von Raesfeld and may have resulted in brain damage.

Ms. Von Raesfeld explained how the consent processes for receiving experimental care, as well as internet browsing do not provide the information necessary for someone to truly consent. When Ms. Von Raesfeld required implantation of a cardiac pacemaker, the urgency of her severe condition prevented her from truly reviewing all the risks of study participation and data sharing related to the implantation of a Bluetooth-enabled pacemaker. She was presented with a decision that required her to consent to receive this life-saving implant. In addition, when

people consent to share their data to use social media and when they access different webpages, they are not fully informed about ways these data are used and shared. For example, pixel tracking can link a user's internet browsing patterns with their social media page. For example, Ms. Von Raesfeld's internet activity related to researching her chronic illness resulted in targeted ad marketing on her Facebook feed for funeral flowers, which can be very damaging to the mental health of severely ill individuals. In addition, Meta Pixel tracking, developed by Facebook's parent company, is used on hospital websites, even those that require authentication and involve input of protected health information (PHI). This tool enables Facebook to indirectly access PHI; Department of Health and Human Services (HHS) recently issued a warning to hospital systems and telehealth providers about these privacy and security risks, which may violate Health Insurance Portability and Accountability Act (HIPAA) regulations. Data sharing among different webpages and social media websites enables the use of digital dark patterns that exploit cognitive biases and human psychology to nudge users toward actions that may not align with their own best interest.

In addition to brain data, study participants have particular concerns about mental health and reproductive health data sharing. For example, one unethical data sharing practice involved a suicide chatline nonprofit sharing and use of caller data for-profit company to develop chatbots for suicide prevention. Users of this nonprofit chatline did not consent to their data being shared in this way for corporate profits.

Due to gaps in the consent process pertaining to data sharing and future data use, nonprofits across the United States are working to build collective rights for study participants based on opinions on data sharing across different patient communities. One potential strategy to achieve these rights involves the establishment of a data marketplace. Users would have agency over their own data and have the opportunity to select a data broker who would make decisions about how the user's data are used. This model would prevent individuals from signing an excessive number of consent forms while still maintaining control over their data.

Question and Answer Session

Ms. Von Raesfeld proposed that patient advocacy groups serve essentially as trusted data brokers. Reorganization of advocacy groups (i.e., based on specific patient needs, not disease areas) can enable easier and more efficient implementation of such a broker system.

Perspectives on Sharing Data in BRAIN Initiative Studies

Amy McGuire, JD, PhD, Baylor College of Medicine

Dr. McGuire presented empirical data on data sharing, primarily from the BRAIN Initiative. Data sharing involves a tradeoff between potential risks and benefits, and views on this tradeoff differ for different entities and individuals. While risks are associated with data sharing, harms can also occur due to a lack of data sharing. When study participants want their data used to advance science, researchers may disrespect that preference by not sharing these data. In addition, if participants are already exposed to risks involved in sharing data, then the scientific community needs to ensure that researchers make the most of that data.

Dr. McGuire emphasized how different contexts should inform data sharing practices. Research participant opinions on data sharing depend on who is asking for the data and their level of trustworthiness. Generally, research participants tend to trust researchers to use their data in ethical ways, whereas some participants are concerned about data access by third parties, including government and law enforcement agencies or corporations that may use data nefariously (e.g., stealing and ransoming). Therefore, researchers need to develop improved guidelines on the ethics of data use and confidentiality, including stronger data use agreements that specify allowable and unallowable data uses. In addition, enforcement systems are required to hold bad actors accountable when they violate rules and policies with unauthorized data access and use.

Dr. McGuire provided key insights obtained from her studies of study participant opinions on data sharing, summarized in the sections below.

Randomized Controlled Trial of Consent for Data Sharing (2007-2010): This study enrolled participants already enrolled in genomics studies that included data sharing practices. Most participants indicated they were willing to broadly share their genomic data and other data types generated in these studies. Their desires to advance research often outweighed concerns about privacy. However, a minority of participants did express concerns about open data sharing practices and indicated preferences for more restricted access.

Poliseq (2012-2016): Poliseq is a modified policy Delphi study that engaged stakeholders to identify the biggest challenges to clinical integration of next-generation sequencing data. Stakeholders ranked data sharing as the most important policy challenge but the least feasible challenge to address.

InfoCommons (2015-2018): This study investigated the requirements to share data in a responsible and ethical manner. Findings indicated that researchers should actively engage research participants and patient advocates during the data sharing process to provide agency over individual decisions. An alternative to requiring participants to consent to data sharing at each decision point would involve the use of trusted intermediaries, such as data brokers. In addition, researchers need to establish their trustworthiness in relation to privacy, security, transparency, and accountability.

Data Sharing in DBS: Researcher's Perspectives (2017-2022): For this study, 23 researchers were interviewed about DBS-related data sharing. Researchers indicated support for and commitment to data sharing but also noted that data sharing practices are too heterogenous. They also expressed the following concerns about sharing human brain data: privacy and confidentiality, usability of shared data by others, ownership and control of data, and limited resources available for data sharing efforts. Generally, respondents were supportive of broad data sharing within the scientific community and expressed only occasional concerns about data security.

The Sulston Project (2019-2024): This modified Delphi study is investigating challenges related to sharing cancer gene variant data. Stakeholder concerns collected thus far include data

privacy and security; equity, diversity, and inclusion; and data quality. In addition, current values and incentives in the scientific community discourage data sharing.

BRAINshare: Sharing Data in BRAIN Initiative Studies (2021-2025): The overall goal of BRAINshare is to identify challenges and concerns specific to human brain data sharing. Thus far, Dr. McGuire and collaborators have conducted 54 stakeholder interviews to identify challenges and convened a Delphi panel of 31 members of the research community to rank those challenges. The prioritized challenges specific to human brain data sharing include (1) high cost of data sharing, (2) lack of data standards in the brain research community, (3) limited brain data sharing incentives in the research community , and (4) limitations in the current informed consent process on ethical data sharing. Stakeholders recommended that NIH prioritize funding for certain types of data over others, including data with the most reuse potential and data that are the most difficult to collect. This study also includes study participant interviews also aim to understand their attitudes, preferences, and concerns about brain data sharing and privacy. Of the eight study participants interviewed thus far, they had a range of conditions and were generally supportive of data sharing to advance science. However, some expressed concerns about potential discrimination by insurance companies and other abuses by those in power as well as the ethics of contributing to large industry profits while receiving little in return. Some also indicated preference for separating consent to participate in a study from consent to share data. In addition, study participants indicated that other data types were also highly personal beyond just brain data including substance use, sexual activity, family history, and personal internet activity and communication history.

Ultimately, data sharing requires those sharing data to be trustworthy and for study participants to have autonomy over decisions regarding sharing of their data. Dr. McGuire cautioned against putting too much weight on the consent process and that rules and consequences are required to deter bad actors from misusing data.

Question and Answer Session

Dr. McGuire explained that study participation and data sharing are activities that should be separated during the consent process, when possible. When data sharing is just an additional deliverable to the primary objective of a study, then consent to participate should not include consent to data sharing. However, if the primary objective of a study is to generate a data resource, then consent to participate should require consent to data sharing.

Discussion

Moderators: Christine Grady, MSN, PhD, NIH; Nina Hsu, PhD, NINDS

Strategies to Improve Participant Understanding of Data Sharing

Study participants tend to recall discussions about data sharing that occurred during the study consent process, but they cannot always remember the content of those discussions—i.e., how researchers actually will share and use their data. Importantly, investigators need to educate study participants on the present and potential future risks of data sharing while also combatting the overestimation of the amount of actionable data that can be extracted from study data. Investigators should emphasize that risks of data sharing can change over time, and

that while investigators cannot guarantee that participant data will not be used for unethical purposes, they can assure participants that procedures are in place to protect their data as much as possible. In addition, consent forms should be more granular and request consent for different types of data sharing separately from one another. These data sharing consent conversations are an essential part of the education necessary for communities to better understand research and data sharing processes.

Penalties for Unethical Data Usage

Due to the anonymity of hackers, systematically holding these individuals accountable for stealing or ransoming data would be nearly impossible. However, other countries may have stricter rules and penalties that could serve as models for U.S. data sharing policies. When considering penalties, researchers with good intentions who do not follow specific data sharing rules should be treated differently from hackers with intentions to harm individuals or communities. Policymakers should also consider whether criminal or civil penalties are more appropriate for holding bad actors accountable.

Engaging Communities to Improve Informed Consent

Shared data typically lacks data from minority populations due, in part, to lower recruitment and enrollment rates into research studies. For example, current BRAIN-funded adaptive DBS study cohorts are not representative of the diversity of the U.S. population; as a result, Dr. McGuire's research studying cohort member opinions of DBS data sharing will also lack data on opinions from certain communities. To obtain data from underrepresented groups, investigators could administer surveys asking hypothetical questions about data sharing, but these answers would likely differ from answers to questions about actual data sharing. Another alternative could include involving underrepresented communities in discussions about informed consent and data sharing. Community organizations can participate in essential roles of educating people about research studies, consent, and data sharing, as well acting as data brokers for the people within their communities. These practices would return some data sharing autonomy to underrepresented groups themselves.

Breakout Room Reports

Meeting participants separated into breakout groups to discuss various components of data sharing, including technology developments to support ethical data sharing, methods to improve informed consent processes for more diverse communities, strategies to empower participants to make autonomous decisions about their data, and cybersecurity measures to prevent unethical data use.

Technological Developments to Support Ethical Data Sharing

As technical systems used for data sharing evolve, developers should embed within these systems mechanisms that ensure participant rights. For example, research participants can create an individual digital avatar, which can adjust, in real-time, preferences for how their data can be used and shared. Granularity in these preference options can improve study participants' experience when making data-sharing decisions. Study participants could (1) make certain data types permanently available; (2) make some data types available but capable of

ready deletion if the participant withdraws consent; (3) choose specific research topics for which their data can be used; (4) grant time-limited access to specific data types; and (5) grant data access to certain research entities. An alternative to the digital avatar would be a data broker system whereby participants select individuals to represent their data sharing interests.

Cloud-based systems with built-in analysis tools can provide data access while prohibiting data downloads and restricting how researchers use data. Alternatively, when shared data are available via download, traceability could be built into those files to provide a clear picture of who has downloaded the data and how the data are used.

Practices to Enable Participant Involvement in Data Sharing Decisions

True informed consent to data sharing requires study participants to fully understand current and future implications of their consent. Developing consent forms that better communicate these implications will require assistance from community members and advocates to clearly convey information. By delineating different aspects of consent in these forms, study participants will be empowered to make more informed data sharing decisions. Some of these aspects include (1) consent to data sharing with specific entities; (2) consent to data sharing for specific purposes; (3) separation of consent to participate from consent to share data; and (4) separate consent to generate different data types. By increasing transparency of data sharing during the consent process, investigators can help build trust among the research community and study participants' respective communities.

More research is required to understand underrepresented views of data sharing to ultimately increase the diversity of brain data cohorts. Importantly, when researchers engage diverse communities, some may express significant concerns and resistance to data sharing. Investigators need to engage and earn the trust of communities to build lasting relationships and the infrastructure necessary for sustained engagement, such as community advisory boards. To improve these relationships, study teams should increase their internal diversity: with more diverse research teams, brain researchers can begin employing people in diverse communities to further trust between research and diverse communities.

Data Storage and Cybersecurity Practices to Prevent Unauthorized Data Use

Use of controlled access policies for informatics systems for data sharing can help prevent unauthorized and unethical data activities. The BRAIN Initiative has multiple data archives for human data that are specific to certain data types. Some of these archives are associated with online computing engines that can be used to perform analyses within a cloud-based platform. Such a platform provides computational resources to those who may not otherwise have access while protecting data from being downloaded for inappropriate analyses or data aggregation practices not covered by initial consent to data sharing.

Data sharing systems can be permission-based or trustless (i.e., blockchain). In permission-based systems, prospective users need to request access from a centralized organization that formally reviews these requests. While this process adds burden to researchers, it is more protective of study participant data than simple open access. A trustless system does not

require formal permission to access data; instead, it employs formal processes that were built into the blockchain database at the time of design, enabling intrinsic enforcement of certain data access patterns. Both types of systems add a layer of protection for study participants while still enabling the research community to converge upon specific data analysis approaches using the same datasets.

Study participant metadata can be used to link data from different datasets, increasing the risk of reidentification or aggregation of data in ways not covered by original informed consent. One approach to prevent this process would involve compartmentalization of different data modalities to reduce the risk of aggregation, but this method may further decontextualize data. Data decontextualization can result in researchers misunderstanding dataset nuances and lead them to use inappropriate analysis methods or make inaccurate conclusions.

Notably, as new technologies develop in the artificial intelligence and ML spaces, some latent information within existing datasets may become extractable and used to reidentify study participants. Stakeholders can try to anticipate some of these technology advancements and build methods to protect data, but a nonzero risk of reidentification will remain.

Conclusion

When conceptualizing policies and systems to enable ethical data sharing, stakeholders should continue thinking about the inferences that can be made from different data types, the risks those inferences pose to individuals and communities, and approaches to mitigate those risks. Eventually, different kinds of data could be ranked based on these risks, and data protections can be refined to optimize ethical data sharing practices. The Neuroethics Working Group will build on the insights gained from this workshop and identify how these insights apply to specific issues with data sharing as well as which insights are specific to brain data. These discussions are ongoing, with additional meetings in August 2023 and February 2024. The Neuroethics Working Group will then share these findings with the broader research community.

Appendix: Agenda

All times ET

DAY 1

10:00 AM **Welcome**
John Ngai, PhD, BRAIN, NINDS

10:05 PM **Introduction and Background**
Saskia Hendriks, MD, PhD, NINDS

10:30 PM **Panel 1: The data that is collected and stored by neurotechnologies and the inferences that can be made from these data**

Benefits and risks of sharing functional MRI data
Susie Huang, MD, PhD, Harvard Medical School

Workshop on Ethics of Sharing Individual Level Human Brain Data Collected in Biomedical Research: Human EEG Data
Lorna Quandt, PhD, Gallaudet University

Ethics of Sharing Individual Level Human Brain Data Collected in Biomedical Research
Doris Wang, MD, PhD, University of California at San Francisco

Benefits and Risks of Sharing Structural Brain Imaging
Douglas Greve, PhD, Harvard Medical School

Panel 1: Data from neurotechnologies and the inferences that can be made
Jennifer Collinger, PhD, University of Pittsburgh

12:30 PM **Panel 1 Discussion**
Moderators: Sameer Sheth, MD, PhD, Baylor College of Medicine; Nina Hsu, PhD, NINDS

1:00 PM **LUNCH/BREAK**

2:00 PM **Panel 2: Inferences to be drawn from data and their implications for data sharing**

OpenBCI Technology and Survey
Conor Russomanno, OpenBCI

Decoding Language Using Noninvasive Brain Data
Alexander Huth, PhD, University of Texas at Austin

Implantable Neurotechnologies: Ethical Considerations in Neural Data Recording, Use & Sharing
Michael Young, MD, Massachusetts General Hospital

3:00 PM **Panel 2 Discussion**
Moderators: Jim Eberwine, PhD, University of Pennsylvania; Saskia Hendriks, MD, PhD, NINDS

3:30 PM **BREAK**

3:45 PM **Panel 3: The potential risks to communities**

Confronting exclusionary practices within neuroimaging
Jocelyn Ricard, Yale University

Indigenous Data Sovereignty and Brain Data
Krystal Tsosie, PhD, MPH, MA, Arizona State University

4:25 PM **Panel 3 Discussion**
Moderators: Nita Farahany, JD, PhD, MA, Duke University; Saskia Hendriks, MD, PhD, NINDS

4:45 PM **Day 1 Wrap Up**

5:00 PM **Adjourn**

DAY 2

10:00 AM **Welcome**
John Ngai, PhD, BRAIN, NINDS

10:05 AM **Recap of Day 1**
Saskia Hendriks, MD, PhD, NINDS

10:20 AM **Panel 4: The potential risks to individuals of sharing different types of human brain data**

Risk of Reidentification in Neuroimaging
Russell Poldrack, PhD, Stanford University

Complicating notions of individuals, neuro data, and data sharing
Sara Berger, PhD, IBM

11:00 AM **Panel 4 Discussion**
Moderators: L. Syd M. Johnson, PhD, SUNY Upstate Medical University; Saskia Hendriks, MD, PhD, NINDS

11:30 AM **BREAK**

11:45 AM **Panel 5: Research participants' perspectives on sharing human brain data**

My Story and My Experience with BCI
Nathan Copeland, Research Participant

The Light Collective
Christine Von Raesfeld, People with Empathy

BRAINshare: Sharing Data in BRAIN Initiative Studies
Amy McGuire, JD, PhD, Baylor College of Medicine

12:45 PM **Panel 5 Discussion**
Co-moderators: Christine Grady, MSN, PhD, NIH; Nina Hsu, PhD, NINDS

1:15 PM **LUNCH/BREAK**

2:15 PM **Breakout Rooms/Small-Group Discussions**

3:15 PM **Re-Convene/Breakout Room Report-Outs**

4:00 PM **Day 2 Wrap Up and Next Steps**

4:15 PM **Adjourn**