Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2023 February 16.

Published in final edited form as:

Neuron. 2022 February 16; 110(4): 600–612. doi:10.1016/j.neuron.2021.11.017.

International Data Governance for Neuroscience

Damian O. Eke[1], **Amy Bernard**[2], **Jan G. Bjaalie**[3], **Ricardo Chavarriaga**[4], **Takashi Hanakawa**[5], **Anthony J. Hannan**[6], **Sean L. Hill**[7], **Maryann E. Martone**[8], **Agnes McMahon**[2], **Oliver Ruebel**[9], **Sharon Crook**[10], **Edda Thiels**[11], **Franco Pestilli**[12]

[1]Centre for Computing and Social Responsibility, De Montfort University, UK/HBP;

[2]The Kavli Foundation, Los Angeles, CA, USA;

[3]University of Oslo, Norway;

[4]Center for Artificial Intelligence, School of Engineering, Zurich University of Applied Sciences, Zurich, Switzerland;

[5]Kyoto University Graduate School of Medicine, Kyoto, Japan;

[6]Florey Institute of Neuroscience and Mental Health, University of Melbourne, Australia;

[7]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, Canada;

[8]University of California, San Diego, CA, USA;

[9]Scientific DataDivision, Lawrence Berkeley National Laboratory, Berkeley, CA, USA;

[10]School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ, USA;

[11]National Science Foundation, Alexandria, VA, USA;

[12]Department of Psychology, Center for Perceptual Systems, Center for Theoretical and Computational Neuroscience, and Institute for Neuroscience, The University of Texas, Austin, TX, USA

Abstract

As neuroscience projects increase in scale and cross international borders, different ethical principles, national and international laws, regulations, and policies for data sharing must be considered. These concerns are part of what is collectively called data governance. Whereas neuroscience data transcends borders, data governance is typically constrained within geopolitical boundaries. An international data governance framework and accompanying infrastructure can assist investigators, institutions, data repositories, and funders, with navigating disparate policies. Here, we propose principles and operational considerations for how data governance in neuroscience can be navigated at an international scale, and highlight gaps, challenges, and opportunities in a global brain data ecosystem. We consider how to approach data governance in

Correspondence: Franco Pestilli, Department of Psychology, The University of Texas at Austin, 108 E. Dean Keeton St., Austin, TX 78712, pestilli@utexas.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a way that balances data protection requirements and the need for open science so as to promote international collaboration, through federated constructs such as the International Brain Initiative (IBI).

In Brief

Eke et al. describe the state and challenges of data governance in neuroscience. They identify the need for International Data Governance (IDG). Recommendations are made to advance IDG in major areas: prioritization, principles development, practical tool implementation and educational activities. [40/40]

Introduction

The growing availability of shared neuroscience data from large and small-scale projects is driving unprecedented research and innovation. As a result of a welcomed move towards open sharing of neuroscience data, data are often crossing the legal and national borders from where they originate. The future of understanding the brain depends on developing a robust research ecosystem that facilitates bringing together data across diverse organismal sources, including human and non-human animals, collected under different jurisdictions. As a result of the international nature of many projects, neuroscience is creating novel opportunities for data sharing and discovery, while also generating new *technical*, *legal*, and *ethical* challenges. These novel challenges depend, in part, on different laws and regulations across nations, states, institutions, and funders alike. As of today, the lack of global data governance coordination across countries often places the responsibility associated with data sharing on individual researchers and their institutions, increasing researchers' risk and liability or limiting the potential for discovery (2019a). Institutions that fear liability may err on the side of caution and interpret general regulations in a way that impedes sharing of scientific data (Use Case 1). There is a critical need to define and clarify neuroscience data governance across international borders. To facilitate scientific discovery, mitigating risks and data usage safety concerns while minimizing liability to individual researchers should be made a top priority by researchers, institutions, professional societies, policy makers, industry, funders and other stakeholders.

Data governance has been defined as the "overall management of the availability, usability, integrity, quality, and security of data in order to ensure that the potential of the data is maximised whilst regulatory and ethical compliance is achieved within a specific organisational context" (Fothergill et al., 2019). Historically, data sharing has been defined in a project-centric fashion and in general, projects have been organized and managed within a single country or region. Importantly, we emphasize that data management is different from data governance, although the two are highly interconnected. Here, we define data governance as the principles, procedures, frameworks, and policies that ensure acceptable and responsible processing of data in each stage of the data lifecycle; from collection, storage, processing, curation, sharing and use, to deletion. These procedures help to maintain data integrity, quality, availability, accessibility, usability and security, and define data controllership (or stewardship) and other responsibilities related to the data. Data governance is rooted in existing laws, regulations and ethical principles but extends beyond

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to include policies and interpretations within organizations and specific projects (Stahl et al., 2018). International data governance therefore encompasses both the standards and practices for ensuring that data are FAIR: Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016) and the ethical principles, policies, recommendations, laws, and regulations that apply to processing, sharing and using data within and across borders.

This paper lays out the case for why a robust international data governance (IDG) is necessary for neuroscience and makes recommendations on what parameters an IDG should cover. An effective IDG should be compatible with the open-sharing needs of the neuroscience research community while respecting the diversity of ethics, cultures, and privacy around data sharing across nations. Critically, IDG should prioritize the ability of researchers around the world to work collaboratively and share data to better understand the brain. A key goal of an IDG is to maximize sharing and impact generated from data and minimize the risk that researchers and institutions assume when sharing data. To achieve this goal, a key task of an IDG is to clarify international policies and to help implement governance plans that facilitate research and respect the individuals (both investigators and the study participants). As of today, the foundations of IDG are not established, and as a result, best practices for implementing IDG are also not agreed upon. Because of this, the discussion presented hereafter can benefit not only neuroscience research, but also other fields in biomedical, behavioral, cognitive, and biological sciences. Likewise neuroscience would benefit from any ongoing discussion in other scientific domains.

This paper arose from the Data Sharing and Standards Working Group established by the International Brain Initiative (IBI - https://www.internationalbraininitiative.org; (International Brain Initiative, 2020), a collective created to unify emerging national and regional large-scale neuroscience endeavors. The goal of this paper is to raise awareness of the need for clearer data governance frameworks for the neuroscience community and its stakeholders, including researchers, institutions, professional societies, publishers, funders, and policymakers, and to propose recommendations on how an IDG can be established and managed.

The importance of international data sharing in neuroscience

In recent years, neuroscience data sharing has finally been made a priority in the community (Ascoli et al., 2017; Avesani et al., 2019; Ferguson et al., 2014; McDougal et al., 2016; Milham et al., 2018; Nichols et al., 2017; Poldrack and Gorgolewski, 2014; Poline et al., 2012; Teeters et al., 2015). There are two primary drivers for this need. First, there is a critical need to increase the size of the data sets available, beyond what can be collected in individual laboratories and to address the needs of emerging fields of research involving large neuroscience data sets (e.g., Artificial Intelligence (Kietzmann et al., 2019; Marblestone et al., 2016)). Second, there has been a community-driven need to address reproducibility (McDougal et al., 2016), openness, and FAIR-ness (Wilkinson et al., 2016). The movement for data sharing and openness has been strong, consistent, and successful (Klapwijk et al., 2021; Milham et al., 2018). As a result, the requirements and expectations for data management have moved from reluctance to open sharing and publication of datasets (Gorgolewski et al., 2016; 2019b).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Mechanisms for promoting the sharing of data have been or are currently being developed. As successful community standards and schemas such as the Brain Imaging Data Structure (BIDS (Gorgolewski et al., 2016)); https://bids.neuroimaging.io/), the Open Metadata Initiative for Neuroscience Data Structures (OpenMINDS; https://github.com/HumanBrainProject/openMINDS), and Neurodata Without Borders (NWB; https://www.nwb.org/) emerge and gain traction, these efforts facilitate the distribution and reuse of data. The International Neuroinformatics Coordinating Facility (INCF; https://www.incf.org/) has played a key role in the development and the harmonization of the technical standards for the international neuroscience community (Abrams et al., 2021). In turn, the Institute of Electrical and Electronics Engineers (IEEE) has made efforts to identify standardisation priorities for neurotechnologies (2020). A distributed set of well-managed data archives have been established that can house neuroscience data of multiple types, e.g., brainlife.io, EBRAINS, Canadian Open Neuroscience Platform, dandiarchive.org, OpenNeuro.org, National Data Archive (nda.nih.gov), and NeuroMorpho.org. Funders are requiring or recommending posting data products on these archives. See Supplementary Table 1 for IBI partner initiatives' data sharing policies and what they cover (Jwa and Poldrack, 2021).

While neuroscience has made great strides in establishing the basic infrastructure for data sharing and integration, these efforts focus mostly on technical standards. Still missing is a comprehensive IDG framework to guide a global neuroscience data ecosystem (data ecosystem meaning the raw measurements, the metadata, the software for analysis and management). As archives for neuroscience data are distributed globally, new mandates from both funders and publishers for data sharing to utilize these archives will increase the flow of data across international borders into data archives and back out again to the worldwide community. Thus, neuroscience is in the process of transforming from primarily a local, geographically limited and lab-centric endeavor to an international data-centric activity where neuroscience data within these infrastructures may come from disparate sources, sites, or projects subject to different national and regional regulations, socio-cultural principles, and theoretical perspectives (Kellmeyer, 2018; Paninski and Cunningham, 2018; Teeters et al., 2015) (see Use Case 2). Furthermore, neuroscience research transcends not-for-profit scholarship to applied clinical or product-based outcomes (for example, medical devices and consumer brain-technologies, etc; (Statt, 2017; View all posts by Tim Urban →, 2017; Wexler and Reiner, 2019)).

The evolving definition of neuroscience data

We consider here the question of whether an IDG for neuroscience is covered under the broader issue of data governance for biomedical data (and data from other biological sciences) or whether there are unique aspects of the sharing of neuroscience data that require special consideration. Neuroscience is perhaps distinguished from other domains by its highly multidisciplinary nature, bringing together researchers with diverse expertise including physiology, molecular biology, anatomy, medicine, behavior, cognitive science and computational science. Each of these disciplines is served by its own research communities with their own standards and model systems, which leads to a significant number of silos to cross when attempting to build infrastructures or forge collaborations. Moreover,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

neuroscience is, at the same time, also deeply integrative, and brings together multiple disciplines, scales of biological organization, and data modalities, to gain an in-depth picture of the nature and functions of the brain and other neural systems. Neuroscience data is characterized by a constant technological flux as funders and scientists seek to develop new techniques that will push the traversal of scales and modalities, leading to new data types and infrastructures constantly being required. Neuroscience data can be gigantic. For instance, typical neuroimaging protocols produce nearly 50 GB of data per participant at a single visit, approaching a petabyte-order dataset from a large longitudinal cohort. Coordinating standards and infrastructures in the face of evolving technologies becomes extra challenging both within and across national boundaries. While other fields certainly must navigate cross-disciplinarity, neuroscience is uniquely reliant on integrating and compiling heterogeneity to both characterize the complexities of the brain and keep pace with the rapid rate of discovery.

Neuroscience has also yet to agree on what data should be shared. Historically, neuroscience data have been defined as raw measurements of nervous system structure, operational properties, and function (Figure 1a). A modern definition of neuroscience data transcends raw measurements to comprise derived data as well as metadata that describe the full set of processing steps and analyses used to produce derived data (Amunts et al., 2019; Avesani et al., 2019; Halchenko and Hanke, 2012). These measurements can be collected with a wide variety of techniques spanning all the way from genetic, molecular and cellular approaches to imaging, physiological, and electrophysiological approaches, laboratory analysis, audio/visual recordings, and behavioural observations. Yet, most raw data are unfit for research; they require curation and preprocessing. This derived data can be essential for gaining understanding of the brain compared to raw data, yet derived datasets cannot unequivocally be separated from the complex series of processing steps used to generate them (Figure 1b). Fortunately, the ever-expanding collaboration between neuroscience, engineering, and computer science has created opportunities to track and capture derived data and processing steps so as to support an expansion of the notion of data in neuroscience.

The growing needs and challenges for sharing neuroscience data

The nature of neuroscience data creates challenges for sharing that are not only technical in nature but also economic, ethical and legal. Similar to human genomic data, heightened sensitivity with neuroscience data comes, in part, from the connection it has to human identity, identification, and personhood. Neuroscience research and innovation have been noted to provide “unprecedented possibilities for accessing, collecting, sharing and manipulating information from the human brain” (Ienca and Andorno, 2017) in a way that uniquely challenges human rights principles. Possibilities of neuroprediction through neuroimaging studies (Haynes et al., 2007; Schreiber et al., 2013), neuromarketing (McClure et al., 2004), and pervasive neurotechnologies have informed considerations of Neurorights – a set of rights being proposed to be added to the Universal Declaration of Human Rights (UDHR) (Ienca and Andorno, 2017; Yuste et al., 2017). These issues highlight deeper concerns related to the continued convergence of neuroscience research and artificial intelligence (AI) that may impact our understanding of human identity, freewill, and privacy.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

As in other fields, effective access to and utilization of neuroscience data creates tension between two fundamental community needs: maximizing data access and reducing risk to the subjects and researchers (Figure 2). On the one hand, there is a need for increasing openness, sharing and re-utilization of data to advance scientific understanding and discovery (Ascoli et al., 2017; Avesani et al., 2019; Eglen et al., 2017; Ferguson et al., 2014; Milham et al., 2018; Nichols et al., 2017; Poldrack and Gorgolewski, 2014). On the other hand, there is a need for safeguarding study participants, reducing risks associated with sharing identifiable information, and limiting potential breaches in privacy (White et al., 2020). Additionally, the increasing use and sharing of neuroscience data in industry raise significant tensions related to commercialization and benefit sharing.

While issues regarding ethics and subject protections are usually thought of in the context of human research, animal research also presents challenges, as public attitudes, rules, and ethical guidelines for experimenting on animals differ across countries (see Use case 3). However, unlike human neuroscience data that are increasingly regulated in many countries and regions, animal data are not yet regulated raising possibilities of ignoring potentially problematic sharing of animal data (especially when the data are generated from countries with weak animal welfare regulations). Globally accepted governance mechanisms for animal data via IDG can therefore help to harmonize procedures and processes for animal experimentations to meet scientific standards and also societal expectations.

In principle, neuroscience could benefit from international data governance instruments and frameworks developed in other domains of biological sciences. In practice, after reviewing a wide range of subfields within the biomedical sciences, no established international data governance framework was found. A set of data-related tools for international data governance has been proposed in genomics research. The Global Alliance for Genomics & Health (GA4GH) has developed useful data tools such as data use ontology v1 and GA4GH passports v1, which address the specific needs of genomic data. Although neuroscience could benefit from these tools, neuroscience data are likely to present additional technical and ethical challenges due to its complexity and scale. In light of the recent restrictive regulations spearheaded by the European Union and Australian government, the growing need for sharing data across laboratories in different countries, an increasing legal burden is affecting both investigators and institutions. At the same time, an improved ethical scrutiny for the legitimate use and re-use of neurodata shared across countries is necessary but lacks foundations (Hallinan et al., 2021). Neuroscience presents an excellent example for governments, international organisations, and other agencies to consider when developing and implementing data sharing policies. In the following sections, we outline challenges in international data sharing and make recommendations in the context of neuroscience, recognizing that these issues may hold for other domains as well.

We touched on some of the broad technical, legal, and ethical challenges above, but there are also practical difficulties imposed by differences in language, cultural practices, and the multiplicity of technical standards. For example, during the course of this work we gathered the data sharing policies governing the large brain projects or initiatives that are members of the IBI (Supplemental Table 2; (OECD, 2017)). These documents were produced for national constituencies and so were not always written in a language understandable to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

those outside of these jurisdictions. Table 1 summarizes some of the types of challenges encountered in neuroscience data sharing that can be addressed by an IDG framework. In the following, we consider ethical and legal challenges in more detail.

Underlying a responsible IDG framework are binding regulations and ethical principles. While many of the legal issues are addressed by the disparate data protection laws available in different regions and countries, some of the ethical issues, such as privacy and confidentiality, informed consent, and autonomy have a long history in bioethics and technology ethics. Thus, their consideration often relies on soft-law governance instruments (see below). In the face of novel uses of globally available large and complex biomedical data occasioned by advancements in technology, a growing body of literature is emerging on the considerations of ethics of data in biomedical sciences (Ienca et al., 2018; Knoppers and Thorogood, 2017; Mittelstadt and Floridi, 2016; Salerno et al., 2017; GA4GH.org). This attention to ethics is important not only because of the expanding nature and role of data, but also because of the potential risks of advancing research in direction breaching ethical principles not covered in law or the potential risk of lack of progress stemming from fear of breaching unclear ethical and legal principles.

A recent publication by the IBI Neuroethics working group discussed extensively many of the ethical issues associated with neuroscience data (such as informed consent, agency and autonomy, privacy, equitable access and benefits, misuse, dual use, and animal protection) and how different social norms, cultural practices, and religious values influence the way they are defined, operationalized, and enforced (Global Neuroethics Summit Delegates et al., 2018). These differing perspectives on the nature and meaning of ethics give rise to tensions in the context of global data sharing, especially in the presence of competing values, interests, and commitments. For most researchers, interactions with ethics compliance remains within local, regional, institutional, or project-specific processes that are different in jurisdictions in which their data are shared/received. Fothergill et al. (2019) also pointed out that the apparent dominance of European and North American perspectives in the field of neuroscience creates an imbalance in a research ecosystem that is increasingly global. Thus, the challenge for an IDG framework in neuroscience is to harmonize these different cultural and ethical perspectives in a way that will enhance understanding, advance collaborations, and facilitate responsible sharing of data. Leaving ethics behind in the global governance of neuroscience data may likely result in missed opportunities for collaboration and sharing, including benefit sharing, a concept emphasized by the Nagoya Protocol on access and benefit-sharing (Biosafety Unit, 2021).

Data-related governance instruments

As indicated above, international data sharing is typically subject to a set of normatives comprising national and international legislation, recommendations, policies, and agreements in need of harmonization. These regulations are jurisdictionally constrained and fragmented but can have international implications (See Use Cases). Although some non-binding instruments of international significance exist, there is no single global, enforceable regulatory framework that shapes IDG discourse. As data move across countries and continents, available national or regional regulations are rendered less efficient,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

inadequate, and inconsistent. In the case of national collaborations, researchers may rely on the assumption that legal normatives that apply to their institutions will also hold for those of their collaborators. However, this assumption is only partially valid and often not applicable when collaborations involve multiple countries. Even though research organisations typically provide some kind of support on the development of data management plans and advice on legal aspects for technology transfer, this type of support is insufficient as the former is focused on national legislation and directives of funding agencies, whereas the latter mainly concerns advanced stages of research. Instead, given the multiple challenges related to neuroscience data sharing, it is necessary to create mechanisms that inform and support researchers throughout all stages of the research process. Below we present a general and not exhaustive description of the different governance instruments that may apply to researchers in neuroscience. Descriptions are categorized depending on the type of governance instruments and the breadth of their scope.

International soft law

International soft law are quasi-legal instruments that are not legally binding but are encouraged as a matter of principle. Applicable soft laws include general instruments such as the Universal declaration of human rights (United Nations) or the declaration of Helsinki (WMA). Other examples of soft law that are more specific to neuroscience include the OECD recommendations on health data governance (OECD Website) and responsible innovation in neurotechnology (OECD Website). These declarations are the result of significant international consensus processes and, despite not being legally binding, have strong influence in the practice of organizations across the world. For instance, the European Human Brain Project (HBP) opinion on data protection makes reference to the National Commission for the Protection of Human Subjects (1979); The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research, and the UNESCO Principle of Respect for Human Vulnerability and Personal Integrity of 2013.

National legislations on intellectual property

National legislations on intellectual property, privacy, security, and trade also define the applicable framework for data sharing. International collaborations are particularly sensitive to discrepancies among legal frameworks in different countries. An illustrative example of this case is the GDPR. While research occupies a privileged position within the GDPR, it nevertheless highlights the discrepancies in the regulations for data sharing with respect to countries outside the European Union, which are often in conflict with principles governing open science (Bovenberg et al., 2020; Townend, 2018). For instance, legitimate personal data transfers from the EU to the USA were based on a legal agreement called the *Privacy Shield*. On July 16th 2020, the Court of Justice of the EU issued a landmark decision in the *Schrems II* case that invalidated the Privacy Shield decision (Fantin, 2020). The Privacy Shield was part of a list of possible legal justifications outlined in articles 44–49 of the GDPR for the transfer of personal data from the EU to third countries. This list of justifications for cross-border transfer include *adequacy decisions, bilateral or multilateral agreements, specific situations on the basis of consent, public interest, vital interest of the data subject and legitimate interests pursued by the controller not overridden by the interests or rights of the data subject*. In the absence of the *Privacy Shield* or any other form of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

adequacy decision, data transfers from the EU to the USA can now only happen on the basis of the other justifications on the GDPR list (Hallinan et al., 2020) and with appropriate safeguards that are mostly regarded as complex processes.

Other examples of relevant national legislations that have considerable influence on how data is shared include the USA Health Insurance Portability and Accountability Act (HIPAA) that sets standards for protection of sensitive patient health information in the USA. As illustrated in Use Case 2, data repositories hosted at American Institutions such as the Open Data Commons for Spinal Cord Injury (odc-sci.org) require that all deposited data be compliant with HIPAA regulations. However, as shown in the Supplementary Figure 1, there are multiple levels and meanings of de-identification and pseudonymisation. How do researchers outside of the USA who have collected data under local regulations know that their data can be legally submitted to an archive like ODC-SCI? In this category we also find clinical regulation where different entities may apply. These legislations have disparate understandings of pertinent issues such as Intellectual Property (IP) and core issues such as the length of time privacy must be protected. HIPAA, for example, extends privacy protection for 50 years after someone dies, whereas under the GDPR, privacy protections stop at the death of the individual.

National legislation regarding data exchange can be strongly influenced by international relations. In particular, security concerns have been raised as motivation to restrict data sharing across countries. International scientific collaborations similarly may be impacted because of fraying political relationships between countries or changes in data protection regulations (see Use Case 1).

General and specific sponsor-driven policies on data sharing and governance: Research data sharing is also shaped by policies that come from funders. More and more funders are issuing requirements for data sharing that impact all grantees. In other cases, certain funding programs may come with specific requirements for data sharing and governance. For instance, the US NIH BRAIN Initiative has issued a specific requirement for data sharing for fundees even specifying into which repositories data must be deposited. The National Institute of Mental Health has a data archive (NIMH Data Archive (NDA)). The NDA has specific policies with challenging requirements for data sharing mandating federal-wide assurance numbers for access. Additionally, the U.S. National Institute of Health has recently published new data management and sharing policies. In the HBP, these requirements include the European Commission strategy for data and Guidelines on FAIR Data Management in Horizon 2020. Supplemental Table 2, shows some of the current policies and requirements for data sharing across programs in different national and geographic regions. The table focuses especially on the IBI partner programs/initiatives.

Institutional policies and contracts: Researchers wishing to share data across international borders may also encounter policies established by their home institutions that impact aspects of data governance. In some cases, institutions may have agreements in place with foreign institutions that can facilitate international collaborations. Similarly, when such agreements are not in place, the researcher may have to expend considerable effort to try

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to negotiate memoranda of understanding (MOUs) and non-disclosure agreements (NDAs) between institutions.

Project specific policies: Individual projects or consortia may be governed by data management and sharing rules they establish through agreement among consortium members. For example, a consortium may agree to a specific license under which data must be shared. Often consortia work through member committees to establish these agreements.

The above discussion illustrates the intricacies of the current state of brain data governance. Taking all these policies together, they represent a complex regulatory burden on the various stakeholders that must comply with them. Given the increasingly international nature of neuroscience research, researchers need training and resources that enable them to gain better understanding of the value and process of international data sharing, and to navigate the various regulatory and ethical requirements associated with such activity (Figure 3).

What does an International Data Governance Framework for Neuroscience look like?

The IDG framework for neuroscience may best be thought of as a function that provides a responsible and holistic approach for all stages of the data lifecycle - from collection, processing, curation, archiving/preservation, application and utilization, and sharing to deletion (Table 2). The IDG should facilitate neuroscience data sharing for scientific discovery and technology transfer, by reducing the burden for individual researchers or institutions. A nimble IDG framework can allow sharing while respecting rights of the subject, and cultural and ethical values. The IDG will need to shape aspects of governance related to data quality (which includes its accuracy, completeness, relevancy, validity, timeliness, and consistency), security, integrity, usability, attribution, accessibility, and ultimately its trustworthiness.

From a technical perspective, IDG for neuroscience should clarify and simplify the ethical, cultural, and legal issues across the different stages of the data lifecycle (see Table 2) and propose a simple workflow for addressing issues and implementing research. Whereas some issues are specific to one or two workflow stages, many of them are intricately linked. For instance, how data were collected or processed can affect the way they are shared. Sharing initiatives concerning data collected from human subjects are influenced by informed consent (Spence et al., 2018; White et al., 2020). This can also be impacted by processing activities such as anonymization, de-identification or pseudonymization. Furthermore, sharing of both human and non-human subjects can be impacted by curation (standardization, identification of data controllers) and archiving (e.g., data security). Furthermore, there is a growing body of research literature on whether anonymization, pseudonymization or de-identification can work or not for neuroimaging data (Eke et al., 2021; Song et al., 2015; White et al., 2020). IDG could help to clarify the degree to which issues such as the potential risks of re-identification should be considered or whether they are normal within the larger research goals of the good of the society. A critical issue in the sharing stage (Table 2) is licensing of data for commercial purposes. This is implemented using different approaches across the current data repositories (Jwa and Poldrack, 2021).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Underlining these commercial licensing policies are legal and ethical issues that are also different across jurisdictions. Critically, an IDG framework would clarify and communicate to the scientific community legal responsibilities and liabilities associated with sharing data cross-borders that impact researchers and repository owners. As of today, none of this has explicitly been worked out by the neuroscience community, and as a result risks fall on individual researchers and institutions (see Use Cases 1 and 2).

Conclusion and Critical Considerations

As technological advancements in AI and machine learning continue to expand the nature, scope, and utility of neuroscience data, we propose that the development and implementation of IDG for global neuroscience research be guided by the following considerations.

1. Make International data governance a priority: IDG should not be an afterthought.

Data governance should be addressed before (planning stages), during (project execution stage), and after a project (dissemination and exploitation stages). When data governance is an afterthought, critical issues in the different stages of the data lifecycle are missed. Again, all stakeholders (researchers, project coordinators, institutions, professional societies, publishers, infrastructure providers, funders, and other stakeholders) have a responsibility in ensuring that governance mechanisms are considered proactively. Governance mechanisms can be described in grant proposals or at the beginning of a collaboration. It is particularly important that a defined data governance framework that aligns with accepted IDG be established at the beginning of any project involving international collaborators or with the potential to share data internationally. Infrastructure and institutional policies can also play an instrumental role in ensuring that IDG is not an afterthought. While any changes may require the identification of funding and resources that are required to develop informed IDG plans, concrete next steps to addressing policy changes and raising awareness among appropriate stakeholders are included below.

Suggestions for practical actions:

- Expand the funding initiatives to support the development and implementation of IDG tools and services for neuroscience. Programs at the scale of international coordination as well as multidisciplinary work would be ideal. Efforts to develop and implement such tools would require leadership and support for researchers, as well as the involvement of legal, security, economic, ethics, and technology experts across national borders.
- Integrate IDG into research project planning similar to how data management plans are required in grant applications, would facilitate responsible IDG within research.
- Establish offices of data governance at institutions, data repositories, neurotechnology companies, research projects, funders, and other relevant bodies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Develop principles for international data governance.

There is a need for simple and clear international governance principles to maximize openness and access to data for the good of scientific progress. These principles should cover ethical, scientific, technical, legal, and sharing requirements. The development of the IDG principles should be developed by identifying and involving the multiple stakeholders involved (e.g., researchers, civil society groups, funders, journals, institutional representatives, and technologists). Given the differences in culture and legal systems across nations, and given the many stakeholders involved in processing neuroscience data (researchers, institutions, infrastructure providers, funders, and local governments) IDG for neuroscience would need to be defined by involving multiple communities and representatives from various nations and cultures (Fothergill et al., 2019; Stahl et al., 2019). There is a need to acknowledge and be responsive to the many different needs and socio-cultural and political dynamics that shape the diverse ethical principles and laws associated with data. Therefore, inclusive dialogues with different stakeholders from different cultures and disciplinary backgrounds should characterize the development of IDG principles for neuroscience.

Suggestion for practical action:

- Initiate multi-stakeholder partnerships/collaborations to develop legal, ethical, technical, organizational and cultural principles that can shape IDG for neuroscience data. Primary stakeholders include transnational neuroscience researchers, organizations (e.g., OECD and UNESCO), technical societies (IEEE, INCF), archives, projects and platforms involved in data standards, sharing and analysis (OpenNeuro.org, brainlife.io, conp.ca, ebrains.eu, PRIME-DE) and scientific organizations associated with brain research.

3. Develop practical tools and guidance for streamlined IDG.

Developing simple, easy-to-use tools that allow researchers or other stakeholders to navigate IDG on a case-by-case basis (i.e., when sharing data between pairs of countries) is critical. These tools should help clarify the different and competing ethical and legal principles involved in cross-border transfers for easier implementation. Researchers and funders alike need practical answers to complex recurrent issues (e.g consent, anonymization, access control, human rights and security). Examples of such would be quick guides on ‘How To Share Data’ between pairs of countries, what to share or receive, how to establish consent forms that would allow open sharing of data (see the open brain consent initiative (Bannier et al., 2021)). These tools will ensure that relevant information is available to the scientific community when they are planning an international project. Ideally, this effort would include the development of semi-automated methods for the analysis of regulatory documents and ethical perspectives across countries, and it would involve the input and guidance from experts in navigating international data sharing issues. Individuals with such expertise exist, but they are generally scattered and may be hard to find. Fostering a network of these individuals and integrating them into neuroscience meetings can help to provide guidance to those trying to work across borders.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Suggestions for practical actions:

- Create a global alliance (a consortium) that will guide the development of technical standards and establish data governance best practices. The new consortium should be organized at grass-root level led by scientists and endorsed by neuroscience organizations. The consortium would also need to involve funders and policy-makers to represent the wide range of interests in neuroscience data. The consortium would be synergistic but not overlapping with the issues already addressed by the Global Alliance for Genomics and Health (GA4GH.org).
- Develop a sustainable federated data catalog of neuroscience data from the brain initiatives and other resources around the world, establish governance to maintain metadata standards for FAIR data sharing, and develop and disseminate tutorials, training materials, and educational activities for dataset publishing and data reuse.
- Develop a Neuroscience Technical IDG Toolkit. Among other things, the toolkit would guide researchers on IDG across the data lifecycle steps (Table 2).
- Develop a Regulatory and Ethical IDG Toolkit. This would be defined by a series of documents that would serve the researchers to navigate issues related to IDG.

4. Increase awareness and education on IDG.

There is a need to promote a cultural shift in the scientific community so as to increase awareness of the importance of IDG, given the increasingly international nature of neuroscience research. Issues of cultural and social diversity across nations are especially important, including inclusion of race, age, and gender when defining policies and planning research projects. Such changes in awareness and attitude will require establishing educational mechanisms for responsible open neuroscience data sharing (Choudhury et al., 2014). Developing a set of educational resources is necessary not only to engender cultural change but also to educate investigators on how IDG works and provide key examples using specific research data samples and types. These educational resources should explain both the value of sharing data and the ethical, and legal responsibilities of the parties involved. Furthermore, the educational resources could promote the utilization of data for educational purposes as well as research. This process of data upcycling (Avesani et al., 2019) is critical to training a new generation of scientists with a global mindset on the globalized nature of the research enterprise and creating a culture that attracts and retains a diversity of thinking, heritage, and skill-sets in the neuroscience community. Furthermore, this will effectively accelerate scientific discovery by attracting a multitude of opinions and by bringing higher education closer to the most cutting-edge research data.

Suggestions for practical actions:

- Integrate IDG into educational and professional training activities organized by professional societies and educational institutions (e.g., neuroscience curricula for graduate and undergraduate programs, professional development for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

postdoctoral scholars), including formal IDG certification educational training courses.

- Encourage neuroscience conference organizers, funders, and other initiatives to consider and support targeted training programs dedicated to IDG.
- Offer efficient, accessible, up-to-date data governance training and associated materials for neuroscientists, especially those working in countries with less economic capacity, through international scientific organizations and societies.

The establishment of an IDG guided by the above considerations will require dedicated efforts and significant resources. Fortunately, neuroscience has established internationally-focused organizations like IBI, INCF and societies like the International Brain Research Organization (IBRO), that can facilitate discussions and work towards the development of an IDG. The IBI serves as a coordinator across large-scale neuroscience initiatives, with the aim to create impact that broadly benefits neuroscience. The above recommendations will be pursued by participants in the IBI network, especially in the early stages, but most points outlined here will require a broader community effort. The recommendations would be then developed and implemented in partnership with neuroscience-specific organizations (e.g., scientific societies such as the Society for Neuroscience, IBRO, coordinating bodies such as INCF), data standards and sharing projects (e.g., OpenNeuro.org, brainlife.io, conp.ca, ebrains.eu, NeuroMorpho.org, PRIME-DE), the private sector and professional organizations (e.g., industry, IEEE), data policy experts, and transnational bodies (e.g., OECD, UNESCO).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors' views are personal views and do not necessarily represent those of the organizations with which they are affiliated such as NSF, NIH or of the U.S. Federal Government. Kenji Doya, Megumi Maruyama and Nargis Akter for administrative support. The Kavli Foundation to AM and AB. European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 945539 (Human Brain Project, SGA3) to DE and JB. National Institute Of Neurological Disorders And Stroke of the National Institutes of Health under Award Number U24NS120057 to OR. MM is a founder and has equity interest in SciCrunch.com, a tech start up providing services in support of Research Resource Identifiers and reproducible science. Principal Research Fellowship and Ideas Grant from the National Health and Medical Research Council (NHMRC), and by the DHB Foundation (Equity Trustees) and Flicker of Hope Foundation to AJH. Krembil Foundation funding to SH. Strategic International Brain Science Research Promotion Program (21dm0307003h0004 and 21dm0307004h0003) and Brain Mapping by Integrated Neurotechnologies for Disease Studies (21dm0207070h0001) from the Japan Agency for Medical Research and Development to TH.). NSF IIS 1636893, NSF IIS 1912270, NSF BCS 1734853, NIH NIBIB 1R01EB029272, NIH NIMH 1R01MH126699, and a Microsoft Investigator Fellowship granted to F.P.

References

Abrams MB, Bjaalie JG, Das S, Egan GF, Ghosh SS, Goscinski WJ, Grethe JS, Kotaleski JH, Ho ETW, Kennedy DN, et al. (2021). A Standards Organization for Open and FAIR Neuroscience: the International Neuroinformatics Coordinating Facility. Neuroinformatics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ALLEA, EASAC and FEAM (2021). International Sharing of Personal Health Data for Research (The ALLEA, EASAC and FEAM joint initiative on resolving the barriers of transferring public sector data outside the EU/EEA).

Amunts K, Knoll AC, Lippert T, Pennartz CMA, Ryvlin P, Destexhe A, Jirsa VK, D'Angelo E, and Bjaalie JG (2019). The Human Brain Project—Synergy between neuroscience, computing, informatics, and brain-inspired technologies. PLoS Biol. 17, e3000344. [PubMed: 31260438]

Ascoli GA, Maraver P, Nanda S, Polavaram S, and Armañanzas R (2017). Win–win data sharing in neuroscience. Nat. Methods 14, 112–116. [PubMed: 28139675]

Avesani P, McPherson B, Hayashi S, Caiafa CF, Henschel R, Garyfallidis E, Kitchell L, Bullock D, Patterson A, Olivetti E, et al. (2019). The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. Sci Data 6, 69. [PubMed: 31123325]

Bannier E, Barker G, Borghesani V, Broeckx N, Clement P, Emblem KE, Ghosh S, Glerean E, Gorgolewski KJ, Havu M, et al. (2021). The Open Brain Consent: Informing research participants and obtaining consent to share brain imaging data. Hum. Brain Mapp

Biosafety Unit (2021). The Nagoya Protocol on Access and Benefit-sharing.

Bovenberg J, Peloquin D, Bierer B, Barnes M, and Knoppers BM (2020). How to fix the GDPR's frustration of global biomedical research. Science 370, 40–42. [PubMed: 33004505]

Eglen SJ, Marwick B, Halchenko YO, Hanke M, Sufi S, Gleeson P, Silver RA, Davison AP, Lanyon L, Abrams M, et al. (2017). Toward standard practices for sharing computer code and programs in neuroscience. Nat. Neurosci 20, 770–773. [PubMed: 28542156]

Eiss R (2020). Confusion over Europe's data-protection law is stalling scientific progress. Nature 584, 498. [PubMed: 32843731]

Eke D, Aasebø IEJ, Akintoye S, Knight W, Karakasidis A, Mikulan E, Ochang P, Ogoh G, Oostenveld R, Pigorini A, et al. (2021). Pseudonymization of neuroimages and data protection: Increasing access to data while retaining scientific utility. Neuroimage: Reports 1, 100053.

Fantin S (2020). Data protection commissioner v Facebook Ireland limited, Maximillian schrems: AG discusses the validity of standard contractual clauses and raises concerns over privacy shield (C-311/18 schrems II, opinion of AG saugmandsgaard Øe). Eur. Data Prot. Law Rev 6, 325–331.

Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, and Martone ME (2014). Big data from small data: data-sharing in the "long tail" of neuroscience. Nat. Neurosci 17, 1442–1447. [PubMed: 25349910]

Fothergill BT, Knight W, Stahl BC, and Ulnicane I (2019). Responsible Data Governance of Neuroscience Big Data. Front. Neuroinform 13, 28. [PubMed: 31110477]

Global Neuroethics Summit Delegates, Rommelfanger KS, Jeong S-J, Ema A, Fukushi T, Kasai K, Ramos KM, Salles A, and Singh I (2018). Neuroethics Questions to Guide Ethical Research in the International Brain Initiatives. Neuron 100, 19–36. [PubMed: 30308169]

Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data 3, 160044. [PubMed: 27326542]

Halchenko YO, and Hanke M (2012). Open is Not Enough. Let's Take the Next Step: An Integrated, Community-Driven Computing Platform for Neuroscience. Front. Neuroinform 6, 22. [PubMed: 23055966]

Hallinan D, Bernier A, Cambon-Thomsen A, Crawley FP, Dimitrova D, Medeiros CB, Nilsonne G, Parker S, Pickering B, and Rennes S (2020). International Transfers of Health Research Data Following Schrems II: A Problem in Need of a Solution.

Hallinan D, Akintoye S, Stahl BC, and Eke D (2021). Neuroexceptionalism: Framing an emergent debate.

Haynes J-D, Sakai K, Rees G, Gilbert S, Frith C, and Passingham RE (2007). Reading hidden intentions in the human brain. Curr. Biol 17, 323–328. [PubMed: 17291759]

Ienca M, and Andorno R (2017). Towards new human rights in the age of neuroscience and neurotechnology. Life Sci Soc Policy 13, 5. [PubMed: 28444626]

Ienca M, Vayena E, and Blasimme A (2018). Big Data and Dementia: Charting the Route Ahead for Research, Ethics, and Policy. Front. Med 5, 13.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

International Brain Initiative. Electronic address: j.g.bjaalie@medisin.uio.no, and International Brain Initiative (2020). International Brain Initiative: An Innovative Framework for Coordinated Global Brain Research Efforts. Neuron 105, 212–216. [PubMed: 31972144]

Jwa A, and Poldrack R (2021). The spectrum of data sharing policies in neuroimaging data repositories.

Kellmeyer P (2018). Big Brain Data: On the Responsible Use of Brain Data from Clinical and Consumer-Directed Neurotechnological Devices. Neuroethics.

Kietzmann TC, McClure P, and Kriegeskorte N (2019). Deep Neural Networks in Computational Neuroscience. In Oxford Research Encyclopedia of Neuroscience, (Oxford University Press),.

Kissner L (219AD). Deidentification versus anonymization.

Klapwijk ET, van den Bos W, Tamnes CK, Raschle NM, and Mills KL (2021). Opportunities for increased reproducibility and replicability of developmental neuroimaging. Dev. Cogn. Neurosci 47, 100902. [PubMed: 33383554]

Knoppers BM, and Thorogood AM (2017). Ethics and Big Data in health. Current Opinion in Systems Biology 4, 53–57.

Lankau EW, Turner PV, Mullan RJ, and Galland GG (2014). Use of nonhuman primates in research in North America. J. Am. Assoc. Lab. Anim. Sci 53, 278–282.

Liu Z, Li X, Zhang J-T, Cai Y-J, Cheng T-L, Cheng C, Wang Y, Zhang C-C, Nie Y-H, Chen Z-F, et al. (2016). Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. Nature 530, 98–102. [PubMed: 26808898]

Marblestone AH, Wayne G, and Kording KP (2016). Toward an Integration of Deep Learning and Neuroscience. Front. Comput. Neurosci 10, 94. [PubMed: 27683554]

McClure SM, Li J, Tomlin D, Cypert KS, Montague LM, and Montague PR (2004). Neural correlates of behavioral preference for culturally familiar drinks. Neuron 44, 379–387. [PubMed: 15473974]

McDougal RA, Bulanova AS, and Lytton WW (2016). Reproducibility in Computational Neuroscience Models and Simulations. IEEE Trans. Biomed. Eng 63, 2021–2035. [PubMed: 27046845]

Milham MP, Craddock RC, Son JJ, Fleischmann M, Clucas J, Xu H, Koo B, Krishnakumar A, Biswal BB, Castellanos FX, et al. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. Nat. Commun 9, 2818. [PubMed: 30026557]

Mittelstadt BD, and Floridi L (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In The Ethics of Biomedical Big Data, Mittelstadt BD, and Floridi L, eds. (Cham: Springer International Publishing), pp. 445–480.

Nichols TE, Das S, Eickhoff SB, Evans AC, Glatard T, Hanke M, Kriegeskorte N, Milham MP, Poldrack RA, Poline J-B, et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci 20, 299–303. [PubMed: 28230846]

OECD (2017). Coordination and support of international research data networks (Organisation for Economic Co-Operation and Development (OECD)).

Okano H, Sasaki E, Yamamori T, Iriki A, Shimogori T, Yamaguchi Y, Kasai K, and Miyawaki A (2016). Brain/MINDS: A Japanese National Brain Project for Marmoset Neuroscience. Neuron 92, 582–590. [PubMed: 27809998]

Paninski L, and Cunningham JP (2018). Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. Curr. Opin. Neurobiol 50, 232–241. [PubMed: 29738986]

Poldrack RA, and Gorgolewski KJ (2014). Making big data open: data sharing in neuroimaging. Nat. Neurosci 17, 1510–1517. [PubMed: 25349916]

Poline J-B, Breeze JL, Ghosh S, Gorgolewski K, Halchenko YO, Hanke M, Haselgrove C, Helmer KG, Keator DB, Marcus DS, et al. (2012). Data sharing in neuroimaging research. Front. Neuroinform 6, 9. [PubMed: 22493576]

Poo M-M, Du J-L, Ip NY, Xiong Z-Q, Xu B, and Tan T (2016). China Brain Project: Basic Neuroscience, Brain Diseases, and Brain-Inspired Computing. Neuron 92, 591–596. [PubMed: 27809999]

PRIMatE Data Exchange (PRIME-DE) Global Collaboration Workshop and Consortium. Electronic address: michael.milham@childmind.org, and PRIMatE Data Exchange (PRIME-DE) Global Collaboration Workshop and Consortium (2020). Accelerating the Evolution of Nonhuman Primate Neuroimaging. Neuron 105, 600–603. [PubMed: 32078795]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rosenbaum S (2010). Data governance and stewardship: designing data stewardship entities and advancing data access. Health Serv. Res 45, 1442–1455. [PubMed: 21054365]

Salerno J, Knoppers BM, Lee LM, Hlaing WM, and Goodman KW (2017). Ethics, big data and computing in epidemiology and public health. Ann. Epidemiol 27, 297–301. [PubMed: 28595734]

Salles A, and Farisco M (2020). Of Ethical Frameworks and Neuroethics in Big Neuroscience Projects: A View from the HBP. AJOB Neurosci. 11, 167–175. [PubMed: 32716744]

Schreiber D, Fonzo G, Simmons AN, Dawes CT, Flagan T, Fowler JH, and Paulus MP (2013). Red brain, blue brain: evaluative processes differ in Democrats and Republicans. PLoS One 8, e52970. [PubMed: 23418419]

Shi L, and Su B (2019). A transgenic monkey model for the study of human brain evolution. Zool Res 40, 236–238. [PubMed: 31011135]

Song X, Wang J, Wang A, Meng Q, Prescott C, Tsu L, and Eckert MA (2015). DeID - a data sharing tool for neuroimaging studies. Front. Neurosci 9, 325. [PubMed: 26441500]

Spence O. 'mareen, Onwuchekwa Uba R, Shin S, and Doshi P (2018). Patient consent to publication and data sharing in industry and NIH-funded clinical trials. Trials 19, 269. [PubMed: 29724236]

Stahl BC, Rainey S, Harris E, and Fothergill BT (2018). The role of ethics in data governance of large neuro-ICT projects. J. Am. Med. Inform. Assoc 25, 1099–1107. [PubMed: 29767726]

Stahl BC, Akintoye S, Fothergill BT, Guerrero M, Knight W, and Ulnicane I (2019). Beyond Research Ethics: Dialogues in Neuro-ICT Research. Front. Hum. Neurosci 13, 105. [PubMed: 30983981]

Statt N (2017). Kernel is trying to hack the human brain — but neuroscience has a long way to go.

Teeters JL, Godfrey K, Young R, Dang C, Friedsam C, Wark B, Asari H, Peron S, Li N, Peyrache A, et al. (2015). Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. Neuron 88, 629–634. [PubMed: 26590340]

Townend D (2018). Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Hum. Genet 137, 657–664. [PubMed: 30120573]

United Nations Universal Declaration of Human Rights | United Nations.

View all posts by Tim Urban → (2017). Neuralink and the brain's magical future — wait but why.

Wexler A, and Reiner PB (2019). Oversight of direct-to-consumer neurotechnologies. Science 363, 234–235. [PubMed: 30655433]

White T, Blok E, and Calhoun VD (2020). Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. Hum. Brain Mapp

Wiener M, Sommer FT, Ives ZG, Poldrack RA, and Litt B (2016). Enabling an Open Data Ecosystem for the Neurosciences. Neuron 92, 617–621. [PubMed: 27810004]

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. [PubMed: 26978244]

Yuste R, Goering S, Arcas BAY, Bi G, Carmena JM, Carter A, Fins JJ, Friesen P, Gallant J, Huggins JE, et al. (2017). Four ethical priorities for neurotechnologies and AI. Nature 551, 159–163. [PubMed: 29120438]

Zhang X-L, Pang W, Hu X-T, Li J-L, Yao Y-G, and Zheng Y-T (2014). Experimental primates and non-human primate (NHP) models of human diseases in China: current status and progress. Dongwuxue Yanjiu 35, 447–464. [PubMed: 25465081]

(2019a). European data law is impeding studies on diabetes and Alzheimer's, researchers warn.

(2019b). The Data-Sharing Problem in Neuroscience.

(2020). IEEE Industry Connections (IEEE-IC) Standards Roadmap: Neurotechnologies for Brain-Machine Interfacing. IEEE Industry Connections (IEEE-IC) Standards Roadmap: Neurotechnologies for Brain-Machine Interfacing 1–100

Elements of a new ethical framework for big data research – Washington and lee law review.

WMA declaration of Helsinki – ethical principles for medical research involving human subjects.

OECD Recommendation on Health Data Governance - OECD.

OECD Recommendation on Responsible Innovation in Neurotechnology - OECD.

GA4GH The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework..

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Use Case 1: Data Protection Regulatory Challenges for International Collaborations

The lack of clarity surrounding the European Union's General Data Protection Regulation (GDPR)"s requirements, and their varied interpretations, have disrupted international data sharing collaborations. In 2019, Finland's National Institute for Health and Welfare stopped all data sharing with the laboratory run by National Institutes of Health Director Dr. Francis Collins on account of GDPR-related concerns. This action disrupted decades of collaborative work on a project studying Type 2 diabetes, which previously utilized 32,000 shared DNA samples. Similarly, the GDPR has been cited as the reason behind the International Genomics of Alzheimer's Project restricting the sharing of data between partners outside of the EU. The consortium now runs isolated analyses - which ultimately reduce the value of data, limits research, and costs additional money and time (Eiss, 2020; 2019a). In another report (ALLEA, EASAC and FEAM, 2021) it is estimated that in 2019, about 5,000 collaborative projects involving the US NIH were affected by the implementation of the EU GDPR.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Use Case 2: Challenges to Sharing Spinal Cord Data Across National Borders

Neuroscience is characterized by multiple data archives, usually run by researchers, that cover specific data types or serve particular communities. Many of these repositories are recommended by journals and funders as a place to publish data and are therefore likely to serve an international clientele. The Open Data Commons for Spinal Cord Injury (odc-sci.org) is a community platform hosted in the United States at the University of California for sharing data in spinal cord injury. The majority of data is derived from translational research but some de-identified human data are also hosted. Recently, a non-US researcher submitted de-identified human data to ODC-SCI. The curators were concerned whether the data were de-identified according to the Health Insurance Portability and Accountability Act (HIPAA), as according to the data submission policy, data must be HIPAA compliant. The burden is on the researcher to comply, which means that they will have to become familiar with standards outside of their own country or region. As indicated in Table 1 and Supplementary Figure 1, concepts such as de-identification and anonymization may not have the same meaning across jurisdictions. Challenges as described in this use case will be encountered more frequently as data sharing through recognized repositories becomes more mainstream.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Use Case 3. Sharing data obtained from non-human primate (NHP) research

Although non-human primates (NHP) are generally considered a powerful animal model for addressing particular neuroscience questions, research in NHP presents serious cultural and socio-ethical concerns for many. The close phylogenetic relationship with humans underlies its use in research: human brain disorders such as autism (Liu et al., 2016) and genetic editing for inclusion of human genes (Shi and Su, 2019) to study genetic mechanisms can inform human neurological changes. Some experiments raise significant concerns, shaped by cultural, social, legal, and ethical differences across international boundaries. In recent years, as a result of changes in legislation largely informed by public pressure, the use of NHP in research has been reduced in the European Union (European Commission, 2010) and the United States (Lankau et al., 2014), while NHP research continues to be a staple of neuroscience research in countries in East Asia (Okano et al., 2016; Poo et al., 2016; Zhang et al., 2014). Researchers seeking to engage in NHP research may perform experiments in a permissive locale, and then transfer data to more strict jurisdictions. Therefore, NHP data generated under different legal and ethical frameworks raises concerns of how such data can responsibly be shared with partners or deposited in archives hosted in areas in which such research would not have been permitted. As Rommelfanger, *et al.* (Global Neuroethics Summit Delegates et al., 2018) ask; Should a country accept or use data collected elsewhere in a fashion that is not considered locally ethical and legal? Does this dilemma require international consensus regarding minimum standards? These are questions that border on ethics but also must be considered when developing best practices in data sharing governance. The PRIMatE Data Exchange (PRIME-DE) initiative is one effort raising needed awareness of this issue in NHP imaging where funds from some agencies cannot be used to process shared data for which animal care practices, standards and regulations were not sufficiently known or in compliance with policies (PRIMatE Data Exchange (PRIME-DE)).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

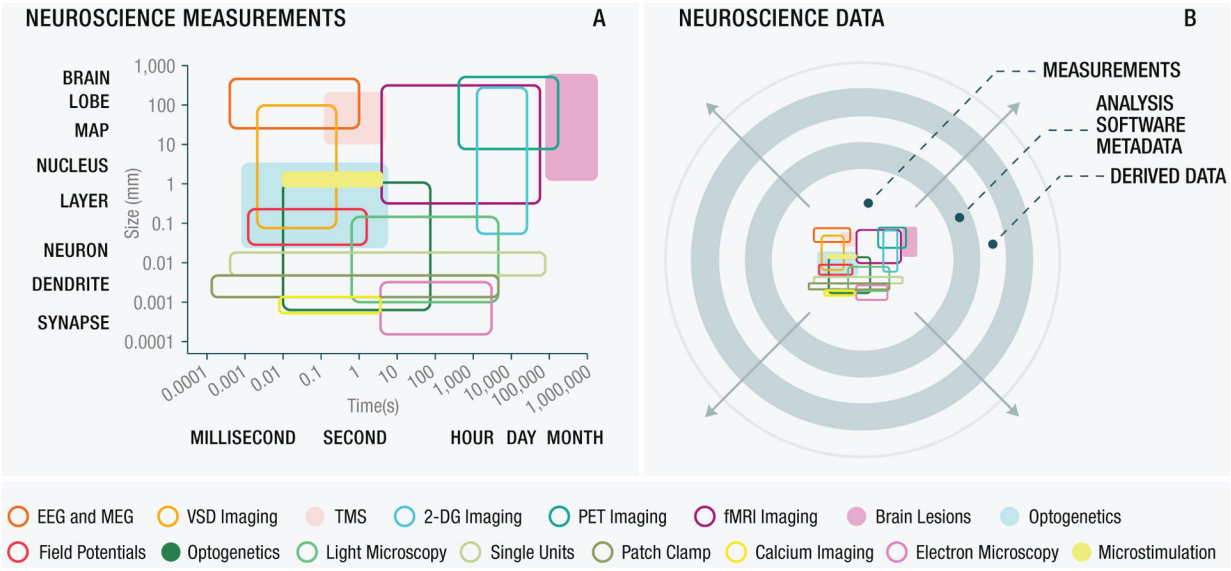

Figure 1. Neuroscience is accompanied by the challenge of managing measurements and derived data across scales.

A. Neuroscience measurements. Measurements in neuroscience have grown over the years. New image modalities and dimensions of measurement have contributed to understanding the brain. **B. Neuroscience data.** A modern definition of data in neuroscience is not limited to measurements, but also encompasses, derived data and the analysis software with all the associated metadata necessary to track the operations performed on the measurements to make them suitable for scientific projects.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

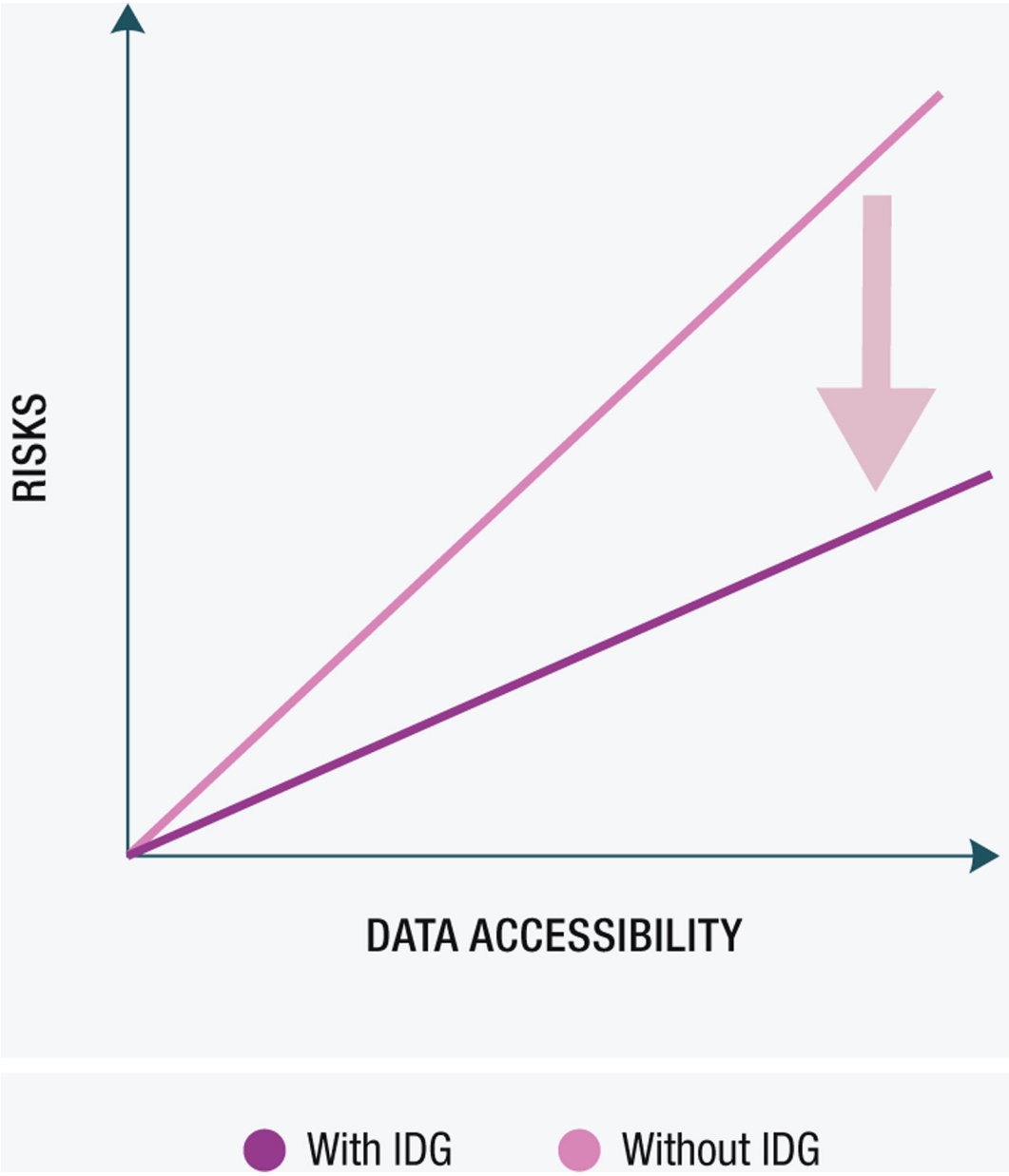

Figure 2. Data accessibility versus risks associated with sharing.

This figure shows the reduction in risks (arrow) associated with data sharing and accessibility as a result of a proper International Data Governance (IDG). Without IDG, the risks increase at a higher rate than with IDG (pink). Clarity and facilitation of understanding of regulations helps mitigate risks to individual researchers and institutions.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

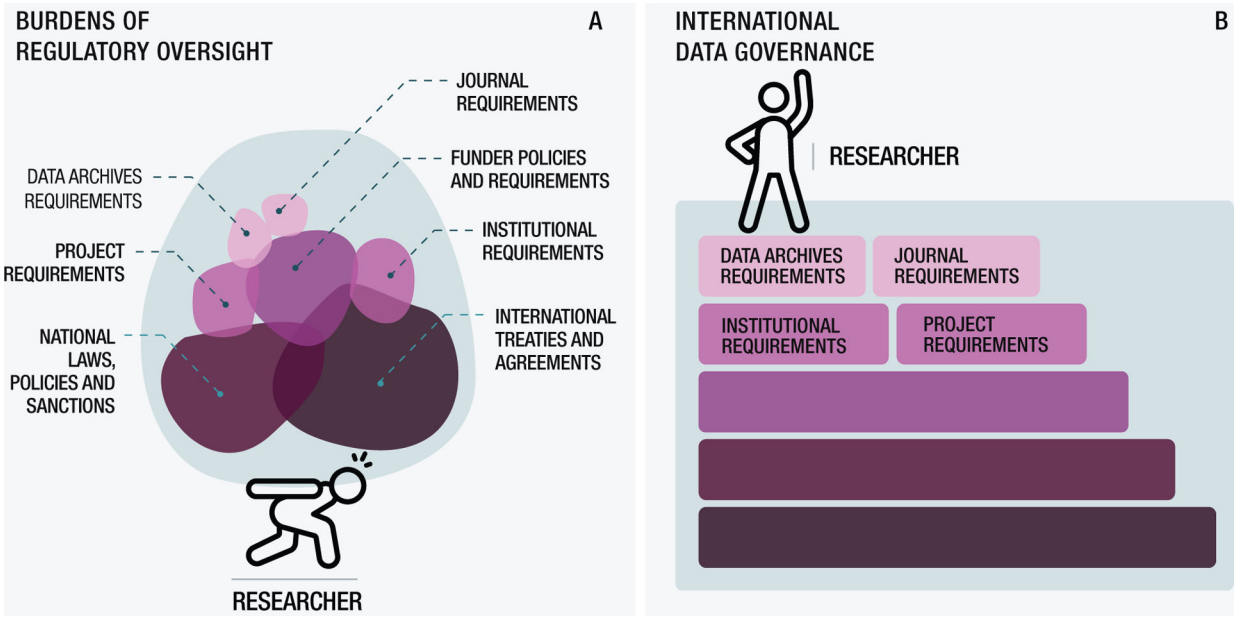

Figure 3. The cascade of regulations, policies, and norms facing researchers engaged in international data sharing.

A. Currently a researcher wishing to share data across borders is burdened by multiple layers of regulatory oversight, some of which may be competing with one another. **B.** With clear practical guidance and tools on international data governance, the researcher can instead stand on a firm foundation of ethical and legal guidelines.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Challenges of international data governance in neuroscience.

Challenge	Description
Ethics	Understanding the ethical imperative for openness on the one hand and the need for data protection compliance on the other; also differences in organizational and cultural values as well as the ethical frameworks and principles underlying the concept of data governance (Fothergill et al., 2019; Salles and Farisco, 2020; Stahl et al., 2018). For example, linking between neural data, cognitive processes, mental states, and mental integrity might have potential benefits but also threats such as manipulation (Yuste et al., 2017) (see also the Neurorights Center at Columbia Universityhttps://nri.ntc.columbia.edu/).
Regulations and Policies	Differences in regulations and policies, including those governing human and animal protections, and different interpretations of regulations and policies. Lack of clarity on regulations and policies overall, and lack of notification of changes to regulations and policies (Rosenbaum, 2010).
Different definitions of Core concepts	Core data concepts such as de-identification, anonymization, and pseudonymization may not mean the same thing in different countries due to varied understanding of personal data (Wiener et al., 2016). Most often anonymization and de-identification are used synonymously in literature. However, anonymization refers to an irreversible process, whereas de-identification gives a room for re-identification which is closer in meaning to pseudonymization than anonymization (Kissner, 219AD; Wiener et al., 2016). Supplementary Figure 1 demonstrates how these are conceptualized and regulated by data protection regulations especially, by the GDPR.
Language	The lack of IDG can create challenges due to differences in language and interpretation between partners. For example, relevant ethical and legal documents that influence data governance are in different languages that the individual researcher may not understand (English, German, French, Japanese, Chinese, Indian, Spanish, Swedish). This highlights one of the problems posed by the increasing internationalization of neuroscience research.
Cultural diversity	In addition to language differences, there are different regional and organizational cultural differences that can affect data sharing. These differences may include social and cultural constructs about the brain and mind, diversity in ethical frameworks and principles, political and regional priorities, as well as approaches to intellectual property management. Sensitivity to these cultural differences is needed for an effective data sharing ecosystem.
Size, complexity and diversity of data	Neuroscience data sets are big and comprise lots of data. In addition to the technical challenges of hosting and harmonizing all of these data, the size and complexity of neuroscience data will likely move the scientific community towards hosting data in accessible environments such as the cloud and bringing computers to the data. There are costs associated with building, and sustaining these infrastructures that may be beyond the reach of researchers in many geographic areas. Should governments develop their own national infrastructure to support big data research or let data be collected outside of government-run infrastructures? If infrastructures are funded by one country, to what extent are they expected to support or subsidize global access to the data hosted by them?

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Needs to consider for International Data Governance (IDG) across the Neuroscience data lifecycle.

Several issues are at stake when data must be considered from an international perspective. Some of these issues are general to any project involving human and animal data. Yet, some specific needs and challenges must be considered when crossing international borders.

Data Lifecycle Stages	Neuroscience Data Governance Needs that affect international sharing
Collection	*Informed consent*. How to collect consent for international sharing and utilization of data *Sampling bias*. How to represent a very heterogeneous population across ethnic groups and cultures *Regulatory differences and legal basis for data collection*. How to understand the different laws for animal welfare or data protection
Processing	*Anonymization, de-identification, and pseudonymization*. How to assure that subjects' privacy can be kept while retaining scientific utility of the data *Regulatory differences and legal basis for data processing*. How to understand the data protection laws in different countries
Curation	*Standardization*. How to understand the different standards for metadata schemes across countries *Data-curation transfer agreements*. How to establish agreements that allow data curation outside of the owner's nation when needed *Security*. How to assure that risks of data breaches are minimized
Archiving and preservation	**1** *Retention policy*. How long shall the data be preserved **2** *Data controllership, stewardship, or custodianship*. Who owns rights on or controls data? **3** *Funding*. Who is paying for data archiving? Should access be free for all users? **4** *Security*. How to ensure risks of data leaking are minimized over time as technology changes
Application and utilization	*Incidental findings*. How to communicate findings that pertain to the health of the study participant *Minimization*. How to ensure studies use the minimal amount of data so as to minimize risks to the participants (e.g., re-identification or privacy break-ins) *Misuse*. How to ensure data are not misused or misapplied for ethically, legally or socially unacceptable purposes. *Biases in analysis and results interpretation*. How to mitigate data analysis bias concerns or misinterpretation of results. *Dual use:* How data can be used responsibly for both civil and military application. *Commercial exploitation:* What restrictions are available regarding using data for economic gain?
Sharing	*Access control*. How to manage access to data, authorization and data use agreements (DUA) across investigators, institutions and countries. *Third party and international sharing*. How to overcome regulatory limitations to sharing data to assure effective scientific impact in international projects. *Risks of re-identification*. How to prevent potential risks of re-identification given advancements in machine learning and AI. *Licensing*. How to approach data licensing and intellectual property concerns when required. *Attribution*. How to cite and keep track of contribution to data collection, processing or curation.
Deletion	*Inappropriate retention*. How to ensure data is retained and deleted responsibly after it has been used. *Loss of data and unintended deletion*. How to ensure resilience to human mistakes.

This table lists some of the most critical aspects that must be considered when embarking on international projects for brain research.