
Review

The role of ethics in data governance of large neuro-ICT projects

Bernd Carsten Stahl,¹ Stephen Rainey,² Emma Harris,³ and B Tyr Fothergill¹

¹Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK, ²Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK and ³Communications Department, Max Delbrück Centre for Molecular Medicine, Berlin, Germany

Correspondence Author: Bernd Carsten Stahl, Centre for Computing and Social Responsibility, De Montfort University, The Gateway, Leicester LE19BH, UK; bstahl@dmu.ac.uk

Received 16 October 2017; Revised 17 February 2018; Editorial Decision 26 March 2018; Accepted 2 April 2018

ABSTRACT

Objective: We describe current practices of ethics-related data governance in large neuro-ICT projects, identify gaps in current practice, and put forward recommendations on how to collaborate ethically in complex regulatory and normative contexts.

Methods: We undertake a survey of published principles of data governance of large neuro-ICT projects. This grounds an approach to a normative analysis of current data governance approaches.

Results: Several ethical issues are well covered in the data governance policies of neuro-ICT projects, notably data protection and attribution of work. Projects use a set of similar policies to ensure users behave appropriately. However, many ethical issues are not covered at all. Implementation and enforcement of policies remain vague.

Conclusions: The data governance policies we investigated indicate that the neuro-ICT research community is currently close-knit and that shared assumptions are reflected in infrastructural aspects. This explains why many ethical issues are not explicitly included in data governance policies at present. With neuro-ICT research growing in scale, scope, and international involvement, these shared assumptions should be made explicit and reflected in data governance.

Key words: data governance, neuroscience, ICT, ethics

INTRODUCTION

Brain-related conditions and diseases are an immense burden on individuals and health care systems.¹ Neuroscience offers opportunities to address some of these issues. Such research is increasingly based on information and communication technology (ICT), partly because the enormous complexity of the human brain requires large-scale data collection, and partly because human brain research faces ethical limitations, which may be partially overcome through ICT (e.g., through the use of simulation or emulation of brain functions). It is therefore unsurprising that neuroscience research programs and associated ICT tools are diverse and proliferating.

These initiatives can be seen as complementary approaches to improving our understanding of the brain. To be successful, they must exchange data and insights, which raise significant scientific challenges with regard to standards, formats, and technologies. One key challenge arises from ethical and legal questions relating to data governance.^{2–4} By data governance, we mean all processes related to the collection, storage, processing, curation, use, and deletion of data. Neuroscience data can be collected from animals, humans, and technical experiments such as simulation. Ethical and legal questions can refer to research ethics (e.g., animal protection, human subject research and informed consent), but also to data use, where questions of data protection or unintended data use can arise.

The assumption motivating our research is that data governance must account for ethical concerns to ensure that neuroscience data can be used broadly across projects and disciplines. The research question guiding this paper is therefore how large neuro-ICT projects integrate consideration of ethical issues within data governance. This research question is in the first instance descriptive and aims to elicit how data governance reflects ethical questions. However, this descriptive interest gives rise to a normative concern. If the future success of neuroscience depends on large-scale ICT-enabled collaboration, then an appropriate integration of ethical concerns is not only a condition of successful collaboration, but of the sustained progress of neuroscience as a discipline. We therefore use this paper to develop recommendations to assist large neuro-ICT projects in developing data governance that facilitates collaboration. A key interest here is whether data governance structures facilitate international and interdisciplinary collaboration.

Neuro-ICT is of particular relevance with regards to data governance because of the fast growth of large-scale projects in the US, Europe, and China as well as many other countries. Ethical concerns are of primary concerns due to the subject matter of the brain which renders data subjects potentially vulnerable when involved in neurobiological and psychiatric studies. And finally, neuro-ICT combines various different data sources (human data from volunteers, patient data, animal data, and technical data) which raise vastly different ethical issues. However, most of the questions discussed in this paper will have some relevance in other biomedical or ICT research.

This paper constitutes a unique contribution to knowledge by providing the first published overview of ethics in data governance across major neuro-ICT initiatives. It shows that some ethical issues are well covered by some policies, whereas others are largely ignored. Furthermore, it demonstrates that there is a large amount of variation between projects. Understanding this is important for individual scholars who will be familiar with the practices in the projects they are involved in, but who lack an overview of the field. Finally, it is critical to consider the overall landscape in defining the next steps of developing data governance principles and practices that are conducive to collaboration across national, cultural, and disciplinary boundaries.

DATA GOVERNANCE AND DATA-SPECIFIC CONCERNS

Many approaches and attitudes toward collaboration in large-scale scientific projects exist.⁵⁻⁸ It is therefore important to understand the regulatory and normative context of these collaborations, particularly in areas concerning data.^{9,10} The use of big data analytics is transforming research across disciplines.¹¹⁻¹³ It can raise (and overcome) various social and ethical issues.¹⁴⁻¹⁶ This is especially accurate for neuroscience, which holds great promise for identifying the causes and treatment of brain-related diseases. There are, however, limits presented by personal health data use, even with amenable and consenting user groups.¹⁷⁻¹⁹ More specifically, neuroscientific research can raise ethical concerns, e.g., volunteers or patients as data subjects, or the *in vivo* use of animals.²⁰ Finally, ethical considerations fundamentally limit neuroscience because they severely restrict certain types of research, e.g., basic research on humans.²¹ The use of technical tools can facilitate large-scale neuroscientific data collection, analysis and data mining, or brain process simulation,

for example,^{13,22} which overcomes certain central ethical considerations, but may simultaneously raise others. One obvious example is privacy and human data protection, but many others exist.^{9,23,24}

These issues are discussed in neuroethics,²⁵⁻²⁸ but the discourse focuses on neuroscience rather than neuroscientific research organization in large data-intensive projects. Data governance is central to the creation and implementation of large-scale projects linking neuroscience and ICT. “Data governance refers to who holds the decision rights and is held accountable for an organization’s decision-making about its *data assets*”.²⁹ It covers the use of ICT, decision-making entities, and which data are considered relevant and valuable. It is linked to both organizational and technical aspects of research. In this paper, we focus on the question of how data governance affects or is affected by ethical aspects of the research.³⁰

In neuro-ICT research, several parts of the research process are easily identified as ethically-relevant, including data acquisition; storage, destruction, and access control; curation and the retrieval process; and processing leading to scientific insights.³¹ Questions of attribution and publication also arise,³² and these are also related to data governance.³³

Data governance is of clear ethical relevance, because it determines whether and how ethical issues such as privacy and data protection are approached. Furthermore, data governance is of high importance with regard to multi-level regulatory and legislative compliance. The complexity of these questions is quickly exacerbated by the international nature of many data-intensive projects, often involving collaboration from initial data collection through to data distribution, publication and dissemination.³⁴⁻³⁷ Attention to data governance is therefore not just part of good research practice, but potentially key in sharing data across disciplinary, organizational, and national boundaries. In this sense, one can argue that appropriately addressing ethical issues in data governance is a condition of the overall success of neuro-ICT as an enterprise, which significantly impacts the likelihood of further progress of neuroscience as a whole.

These questions are relevant because existing research and funding mechanisms incorporate ethical concerns in various ways. Both human subject research and animal research are strictly regulated in most research systems. Adherence to ethical principles is typically a condition of research funding as well as publication of findings. In large collaborative projects, however, it is not always clear which ethical principles form the baseline and which behavior counts as adhering to ethical principles. In neuro-ICT research, for example, some researchers will have a biomedical background and be intimately familiar with the biomedical principles of research ethics, whereas others will come from an engineering or computing background where ethical issues are interpreted and dealt with in fundamentally different ways.³⁸ In many cases, collaboration in neuro-ICT projects is mediated via shared use of data. It is therefore important to understand how ethical questions are expressed or embedded in data governance structures.

METHODS

To address this, we outline the relevant components of data governance, the inclusion criteria for projects to be part of this study, and our approach to data analysis.

Components of Data Governance

To frame our textual analysis, we ask *what* specific ethical issues appear relevant in current governance approaches (specific issues), and *what is proposed* as an approach to mitigate them (policies and implementation).

Specific issues

In order to analyse current data governance structures and processes in existing projects, a level of agreement on which aspects could or should be covered is necessary. This starts with the ethical or social issues that can reasonably be expected to arise during neuroinformatics research, which can initially be divided according to the data source. While animal data mostly raises issues during its creation, human data raises issues during creation (e.g., informed consent, especially where vulnerable users are concerned),^{39,40} and during the data-processing and analysis phases, where questions of data protection become relevant. Further issues can arise downstream (e.g., publication), and retrospectively, e.g., withdrawal of previously available participant data.⁴¹

Other issues of potential relevance include dual use, i.e., using research findings for undesirable military purposes, misuse for non-intended or criminal activities,⁴² and issues around benefit sharing or data source exploitation.⁴³ Important questions of data ownership and intellectual property may also arise. Attribution, i.e., who receives various benefits from the provision or use of data, is relevant in academically-oriented research. It is important whether the data governance structure explains where these issues arise, e.g., during data production, data-processing, data use, etc.

Policies and implementation

In order to understand how these issues are addressed, we look at policies and their implementation. Policies are specific ways in which particular issues are addressed and data governance is realized. One example with far-reaching implications is data access policy.^{44,45} This determines who may access which data for which purposes. The definition of such an access policy requires further detail, including the definition of user types, which we would expect to find in the documentation to be analyzed. An important aspect of policies is the question of enforcement, which may require the collection of additional data (e.g., logging data). A central question is how sanctions for the infringement of data governance principles are determined and implemented, which may differ according to the type of issue or behavior in question. Another important factor with bearing on enforcement sanctions is audits.⁴⁶

Identification of relevant projects

This analysis requires a broad overview of current practice and thus identification of the main relevant projects in the field. To accomplish this, we specified the necessary criteria that a project should fulfil. We expect that the project:

1. Is IT-based
2. Has a focus on neuroscience
3. Uses data or makes it available
4. Is collaborative
5. Is open to external users

Applying these criteria ensured that the projects have sufficient similarity with regard to size, topic, and scope, in order to render the analysis valid and interesting.

Table 1. Neuro-ICT Projects Whose Data Governance was Included

1. 1000 Functional Connectomes	2. ADNI
3. Allen Institute	4. BAMS
5. BMI	6. Brain Biodiversity Bank
7. Braincloud	8. BrainMap
9. Brede database	10. Cancer Imaging Archive
11. CRCNS	12. Human Brain Project
13. HCP	14. IEEG
15. LONI	16. National Institutes of Health
17. National Library of Medicine	18. Neuromorpho
19. NITRC	20. OASIS
21. OpenfMRI	22. SchizConnect
23. UKBioBank	24. Whole Brain Atlas

The process of identifying suitable projects began with the authors brainstorming known projects, followed by a snowballing sampling system, complemented by input from other researchers and experts in the area. This was further supplemented with internet sources via the search terms:

1. neuro-ICT research project
2. neuroscience ICT project

This strategy led to the identification of several lists of projects which were included in the population of investigated projects. Each of these projects was then individually assessed against the inclusion criteria. All available information on the data governance of the selected projects was collected and analyzed. An initial set in excess of 40 candidate sources was found, and subsequently reduced to 24 relevant analysis sources as shown in [Table 1](#).

We do not claim that this list is comprehensive, but that is not necessary for the intended purpose of this paper. We are interested in current mainstream data governance in neuro-ICT, which is why we chose to include some high profile organizations that do not focus exclusively on neuro-ICT, such as the NIH or the UK BioBank, and have a much broader remit. The important point is that they reflect what is currently seen as good practice in the field, which we believe that this sample of projects achieves.

In our analysis, we focused on publicly available descriptions of data policies, their implementation, and enforcement. Such documents are of high value for research because they represent the official position of the research projects, their investigators, and consortia. They are normally subject to a process of review and confirmation that ensures that different positions are adequately reflected and that the policies correspond to the state of the art. There is a well-established stream of research based on document analyses of policies, for example in the area of information systems security,^{47,48} which has been shown to be of relevance to understanding ethical issues.⁴⁹ For the purposes of this exploratory study that aims to provide a first account of the role of ethics in data governance, the published policy documents provided a strong data set.

Data analysis

All analyzed projects possess extensive websites with general overviews and typically access to data, including information on their data governance approach. For our research, we downloaded the data governance documentation, which often required registering on the website.

Once collected, the data governance documentation was analyzed using a thematic qualitative data analysis approach.^{50,51} All documents were uploaded to our analysis tool (NVivo server,

Table 2. Specific Issues

Issue	Example source references	Further information/Comments	Covered by projects
<i>Misuse</i> : the issue of the data being used improperly by those accessing it	“Users may not use the Data Sets for high risk activities such as the operation of nuclear facilities, air traffic control, or life support systems, where the use or failure of the Services could lead to death, personal injury, or environmental damage” (HBP, Terms and Conditions)	The source is one of the few sources that lists explicit rather than implicit issues of misuse.	HBP HCP OASIS UKBIOBANK
<i>Data Protection</i> : ensuring that the data was acquired ethically and is not used in a manner which could lead to it being misused	“I will not further disclose these data beyond the uses outlined in this agreement and my data use application” (HCP, Data Agreement)	While the sources do prohibit using the data in ways which would lead to misuse this is not considered in any depth - see below.	HCP ADNI HBP LONI NIH OASIS OpenfMRI SchizConnect UKBIOBANK
<i>Privacy</i> : how the anonymity of the participants from whom the data is collected are protected	“To protect and assure the confidentiality and privacy of all participants, the Recipient granted access to these data is expected to adhere to the specifications of this DUA” (FCON1000, Data Use Agreement)	The anonymity of the participants is raised but the responsibility is moved back to the user rather than the project	FCON1000 Allen Institute BMI BrainCloud CRCNS HBP OASIS SchizConnect UKBIOBANK

version 10) and coded according to the specific issues listed above. The issue list was open to revision during the data analysis process, so that new, unpredicted issues could be added. Analysis was undertaken by two authors to ensure inter-subjective reliability. We must emphasize that we used an exploratory and interpretive approach to the data,^{52,53} which implies that alternative readings are possible.

Findings

We used the analytical components of data governance (i.e., specific issues, policies and implementation) to organize this section.

By “specific issues,” we mean the aspects of data and data ethics which the projects identify as significant in their data governance documents. Several of these were identified in multiple cases: misuse, data protection, and privacy. These were treated as ethical issues potentially arising from data use. Table 2 shows the specific issues that were identified:

The most striking observation with regard to specific ethical issues, however, is the general absence of some concerns. We did not find evidence of explicit attention to ethical issues related to the generation of data, such as protection of animals or informed consent of participants. There were no references to incidental findings, to possible misuse or dual use of the research, or exploitation of data from developing countries without compensation.

These negative findings are of crucial importance. They demonstrate that many of the ethical issues that are well established in biomedical research in general and in neuroscientific research in particular are not considered worthy of inclusion in data governance structures and policies. In an increasingly international and interdisciplinary research environment, we believe that ignoring such issues may jeopardize successful collaboration. In the discussion section,

we will return to possible explanations of why the authors of the policies may think that this lacuna is justified.

Unlike the possible ethical issues, the documents we analyzed contained much more detail on mandatory user policies, many of which were shared, as can be seen from Table 3.

This table demonstrates that data policies are detailed and cover a wide array of activities and practices. They cover many of the crucial aspects that need to be taken into consideration in order to facilitate data sharing. It is not always obvious, however, how they relate to the ethical issues underlying research in neuro-ICT.

Discussion and Conclusion

Our description of data governance in neuro-ICT projects is unlikely to surprise those familiar with the field. There is a great resemblance across a number of the documents we reviewed, suggesting that authors examine each other’s work for inspiration on good practice. However, the analysis of data governance regimes provides interesting insights with regard to underpinning assumptions and beliefs. These, in turn, are crucial to understanding the implications of data governance for ethics and vice versa.

An important initial point is that the focus of the majority of analyzed documents is the distribution of data, with scant attention given to the generation of data. This is easily explained in most cases, as the data made available via the systems are typically produced by the host projects or organizations and questions of data generation, including questions of research ethics, are dealt with elsewhere. These organizations therefore know the details of the data generation, including any ethical issues that may be involved and how these have been addressed. However, this information is generally not provided, suggesting that it is not perceived as important for potential users.

Table 3. Policies and Implementation

Terms/concepts	Example source references	Further Information/comments	Covered by projects
<i>Access:</i> the requirements for users being allowed to look and use the data	“Prior to accessing the SchizConnect data (‘Data’), Recipient must agree to the following terms and conditions regarding its use” (SchizConnect, Data Use Agreement)	Complexities of access range from having to actively submit supporting evidence to default acceptance of the data use agreement through the act of accessing the data on the site	Allen Institute FCON1000 ADNI BAMS Brain Cloud Brain Map HBP HCP EEG LONI NIH National Library of Medicine NITRC SchizConnect UKBIOBANK
<i>Attribution/Identification:</i> the rules for acknowledging the source of the data in publications	“I will acknowledge the HCP project as a source of data and include language similar to the following: Data collection and sharing for this project was provided by the Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD).” (HCP, Data Agreement)	This is a major theme in the documentation which suggests it is a priority in these projects	Whole Brain Atlas SchizConnect UKBIOBANK OASIS NITRC Neuromorpho National Library of Medicine HBP HCP CRCNS Cancer Imaging Archive Brain Cloud Brain Map Brain Biodiversity Bank Allen Institute FCON1000 ADNI
<i>Ownership:</i> who ultimately owns the data and associated information	“The content, organization, graphics, design, compilation, magnetic translation, digital conversion and other matters related to the Platform are protected under applicable copyrights, trademarks and other proprietary (including but not limited to intellectual property) rights.” (HBP, Terms and Conditions)	Ownership is mainly restricted to copyright notices	Whole Brain Atlas SchizConnect UKBIOBANK OASIS NITRC Neuromorpho National Library of Medicine HBP HCP CRCNS Brain Cloud Brain Map Brain Biodiversity Bank Allen Institute FCON1000
<i>Scope of Use:</i> the specific ways in which the data can and cannot be utilized	“Investigator(s) will use requested datasets solely in connection with the research project described in the approved Data Access Request for each dataset” (BrainCloud, Data Use Agreement)	This policy is used to cover a number of issues such as provenance and retention - who can use and how.	Whole Brain Atlas SchizConnect UKBIOBANK OASIS NITRC Neuromorpho National Library of Medicine HBP HCP CRCNS Cancer Imaging Archive Brain Cloud Brain Map Brain Biodiversity Bank Allen Institute

(continued)

Table 3. continued

Terms/concepts	Example source references	Further Information/comments	Covered by projects
			FCON1000 ADNI NIH
<i>Commercial Exploitation</i> : restrictions regarding using the data for economic gain.	“Recipient agrees not to sell the data in any form to any entity or individual or to distribute the data to anyone other than his/her research staff who will also agree to the terms within this DUA”.(FCON 1000, <i>Data Use Agreement</i>).	The limited mentions of commercial exploitation are focused on prohibiting commercial applications of the data.	FCON1000 ADNI Allen Institute Cancer Imaging Archive HBP National Library of Medicine NITRC UK BIOBANK
<i>Retention/Destruction</i> : how long and in what manner the data is stored, and whether it must be destroyed after use	“When finished using the data, Recipient may retain one copy of the data with access permitted solely for archival and compliance purposes. Any other copies of the data must be destroyed, or otherwise disposed of properly, as permitted by law.” (FCON 1000, <i>Data Use Agreement</i>).	No further guidance is offered on these matters. This is particularly surprising in regard to the retention/destruction of data as this is a procedure that may require some technical instructions about how to ensure all files are removed from the memory of a computer(s)	FCON1000 ADNI Allen Institute HBP NIH National Library of Medicine OpenfMRI UK BIOBANK
<i>Provenance</i> : an understanding of the flow of data	“You may use them for any purpose which will not interfere with their use by others. We do ask that you SECURE OUR PERMISSION, so that we can track the uses being made.” (Brain BioDiversity Bank, Copyright)	Controlling the flow of data, is only alluded to at a very superficial level, usually implicit in other policies such as retention and scope of use	ADNI Allen Institute Cancer Imaging Archive HBP HCP National Library of Medicine Open fMRI Schizconnect UK BIOBANK

This indicates that the systems are built for particular types of users and applications, who are assumed to have certain characteristics. The intended audience is clearly professional scientists who can make sense of the data and have been explicitly socialized to scientific norms and expectations, with the tacit assumption that they will behave accordingly. There is an implicit reliance on shared norms. To a large extent (with the exception of EU-based initiatives), the assumption driving data governance seems to be that the US American legal framework is applicable to potential users. This would explain why many of the ethical issues related to data are not explicitly reflected in data governance. The US Animal Welfare Act, for example, covers the treatment of animals in research. Patient data and possible privacy concerns are dealt with in the Health Insurance Portability and Accountability Act. Corresponding pieces of legislation for European projects include the Directive 2010/63/EU on the protection of animals used for scientific purposes and the European data protection framework, notably the General Data Protection Regulation (EU) 2016/679). A key problem even in the context of these well-developed legal frameworks is that of interoperability, as there is no simple way of assessing whether compliance with one automatically implies compliance with the other.

The specific requirements for researchers to deal with ethical issues are regulated as the “Responsible Conduct of Research” which is implemented in Section 7009 of the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science (COMPETES) Act. It is unlikely that individual researchers are familiar with all of the details of these various federal

acts and even less so with more detailed regulation on the state or local level. However, the research ethics infrastructure that has been developed to implement these requirements, in particular the general coverage of all research through institutional review boards (IRBs^{54,55}) has been comprehensive. The US IRB system is contentious,^{56,57} but explains why neuroscientists seem to believe that ethical issues are not in need of further attention.

The sanctions offered for various breaches of data governance policies predominantly involve withdrawal of access to resources. The law is invoked infrequently as a punishment, generally in relation to copyright infringement, rather than a broader concern with the protection of data-subjects. A notable exception here is the legal duty to inform of data security breaches, which is widely mentioned throughout policy documentation and is another provision of US federal law.

The actual use of data does not appear to be a major concern. User behavior is prescribed, but few or no attempts are made to ascertain whether users work within permissions and expectations. This may again reflect the self-policing nature of the scientific system where infringements of norms can lead to severe formal and informal sanctions.

Overall, it is clear that the main purpose of the vast majority of the systems we investigated is data publication. This needs to be understood in the context of an academic system predicated upon the imperative of “publish or perish,” which explains the significant attention to citation guidance which exceeds most other components of data use principles. Policy adherence monitoring, performance or

effectiveness is not overtly discussed. Rather, logging of user activity is mentioned as a way to monitor compliance. Primarily, such discussions are concerned with the proper attribution of datasets and sources. This is interesting, as potentially complicated questions of data ownership are likely to become more prominent in international collaborative research, in particular when such research leads to marketable projects, and are generally not reflected in data governance documents.

This state of affairs arguably characterizes the current system of neuroscientific work. Individual and collective esteem is gained through accrual of high citation counts. Ethical issues are an aspect of the data generation process, but once dealt with, no longer warrant attention. These principles have served the field well so far, but are perhaps insufficient for it to move forward.

The present environment of neuro-ICT research is characterized by a growing number of large, international collaborative efforts. Simultaneously, work is moving out of the realm of exclusively fundamental research into numerous application areas beyond the biomedical field that include applications of potential commercial or military interest, e.g., the uses of neuromorphic computing or artificial intelligence (AI) for novel types of control systems.

This means that the assumptions revealed by our data policy analysis may no longer be tenable. Users of data will not necessarily be professional scientists bound by the largely informal codes of behavior that regulate the scientific system. Users will originate from a broader range of institutional, national, political and cultural backgrounds, and the presumed views and values upon which these policies are predicated may not be present. The purposes of data use are likely to expand further into areas potentially perceived as ethically problematic. And, finally, if this new environment of neuro-ICT aims to become truly collaborative, it may no longer be sufficient to rely on a relatively straightforward broadcasting model of collaboration where data is made available under defined conditions. Instead, a two-way flow of data or more complex, multi-scalar configurations will be required to allow scientists to assess the various types and quantities of data required for their scientific objectives. The aim of a broad, collaborative, sharing research community is widespread and reflected in the data governance documents, but current data governance regimes are not necessarily conducive to collaboration. If this new collaborative environment emerges, it will pose numerous challenges. Many of these will be scientific and technical, but many will also involve ethics. The ethical issues arising from data generation present a key challenge. For a multilateral exchange of data between multiple organizations in differing jurisdictions to be viable, the metadata requirements will be significant. Funded research must often follow ethical requirements in using collected data. Where research occurs in a homogenous legislative environment, this is unproblematic. However, when data originates from different jurisdictions, it can be unclear whether the ethical requirements in one project are sufficient to meet the requirements of another. As a practical example, US institutions mainly follow 45 CFR 46 and applicable guidelines from the NIH for human research and/or human data (the Human Subjects regulations). To use such data in the EU, for example, it would have to demonstrably comply with European principles and those of the relevant Member State. Equivalent concerns refer to animal data, with different jurisdictions possessing different animal protection regimes that are not always compatible.

While data provenance and its ethical components are an important concern, there are similar issues around data use and the eventual consequences thereof. As an example, the European

Commission's Horizon 2020 research funding programme contains ethics requirements related to 10 different topics. These include research ethics for humans, cells and animals, but also questions of data protection, misuse and dual use (i.e., funding must not be provided for research used for military applications). Permissions to use human subject data are typically linked to informed consent statements which can delineate the purposes for which data can be used. If data subjects were assured that the work would be used for research purposes only, it is unlikely that this data could permissibly be used for commercial purposes.

In practice, large-scale international collaboration may be stifled because of ethics-related concerns. It may be impossible for projects to use valuable data due to uncertainties of ethical status. This does not imply that anything unethical has happened, but simply that the level of awareness and the available information on ethical issues in current data governance regimes may be insufficient to permit deeper collaboration.

If this accurately describes the situation, which we believe to be the case (as it is reflected in our own practice), then the question is: How it could be addressed? Two non-contradictory approaches to the problem are possible: bottom-up and top-down. The former requires the researchers involved in neuro-ICT projects contact other projects to ensure that collaboration is viable. This, no doubt, is already happening to a significant degree. Problematically, it creates numerous ad-hoc approaches that may not be mutually compatible and will therefore not solve the overarching problem of large-scale collaboration. An inclusive, top-down approach, creating shared principles and standards of data governance could potentially achieve this end. The problem here is that it requires international institutions with the expertise and motivation to guide a potentially drawn-out process of negotiation and standardization. This type of development might be driven by associations in the field, e.g., the International Neuroinformatics Coordinating Facility (<https://www.incf.org/>), and could be supported by international standardization bodies and embedded in international political bodies, such as UNICEF or OECD.⁵⁸ Table 4 summarizes the recommendations arising from our research.

The development of data governance for large-scale research projects is a key concern shared by many research funders, research organizations and researchers. Neuro-ICT research is at the forefront of these activities, given its substantial requirements for data and analysis, and due to its potential contribution to novel ICT developments. In this paper, we have argued that the justification for recent, significant global investment into neuro-ICT is, to a great extent, an ethical one. The premise of this paper was that the ethical basis of neuro-ICT needs to be reflected in its data governance. Based on analysis of the data governance principles of 24 large-scale neuro-ICT projects, we found that data protection and the attribution and citation of data are core concerns. Many other aspects, particularly those related to the provenance and generation of data, are much less clear. Our contention is that the neglect of ethical aspects of data governance can constitute a serious impediment to large-scale international collaboration, which in itself would be an ethical issue.

By focusing on published documents concerning data governance, we have gained important insights into the official view of projects concerning their data governance principles and standards. Having shown that these are likely to be insufficient for future large-scale international and interdisciplinary collaboration, further research should be undertaken to better understand how these policies were constructed and how they are interpreted and implemented in practice.

Table 4. Recommendations

Responsible actor	Recommended action
Individual researcher	Be aware of and reflect on ethics of data collection and use
Research project/program	State ethical aspects of data (e.g., ethics approval details) and cover ethics in data governance tools (e.g., data management plan)
Research funder	Encourage researchers to explicitly communicate ethics of data; provide resources for researchers to develop appropriate ways of reflecting on ethics.
International research organizations (e.g., INCF)	Develop standards of data governance that explicitly cover ethical aspects.
International organizations (e.g., OECD, UNICEF)	Highlight the relevance of ethics in data governance; develop principles of international collaboration that are sensitive to different ethical positions.

More detailed understanding of the social reality of policies will also be required to fill the relatively abstract recommendations we have provided with sufficient detail as to render them practically applicable. A further avenue of future research will be to broaden the investigation beyond neuro-ICT and look into closely related areas, such as psychiatry, but also the wider biomedical research environment.

Despite significant amounts of research still to be undertaken, we believe that this paper makes an important contribution to knowledge through its demonstration of an ethical gap in current data governance practices. Closing this gap will require community efforts from the researchers and institutions affected, and the political will to implement solutions. We believe that by highlighting the importance of attentiveness to ethics in data governance of neuro-ICT research, this paper promotes the necessary discussion and fosters the development of ethics-informed data governance principles and standards.

FUNDING

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1).

COMPETING INTERESTS

None.

CONTRIBUTORS

BS, EH and SR developed the strategy for the paper. EH and SR collected and analyzed data and developed the initial narrative. BS and TF revised and rewrote the draft that was initially submitted and undertook revisions. All authors fulfil the ICMJE criteria for authorship.

ACKNOWLEDGEMENTS

This work was greatly supported by the work of the Data Governance Working Group of the Human Brain Project. We are grateful to two anonymous

reviewers for their valuable feedback. This paper presents the perspectives of the authors, and does not reflect the position of the European Commission or the Human Brain Project.

REFERENCES

1. Wittchen HU, Jacobi F, Rehm J, *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* 2011; 21 (9): 655–79.
2. Greely HT, Ramos KM, Grady C. Neuroethics in the age of brain projects. *Neuron* 2016; 92 (3): 637–41.
3. Nijboer F, Clausen J, Allison BZ, *et al.* The Asilomar Survey: stakeholders' opinions on ethical issues related to brain-computer interfacing. *Neuroethics* 2013; 6 (3): 541–78.
4. Quaglio G, Corbetta M, Karapiperis T, *et al.* Understanding the brain through large, multidisciplinary research initiatives. *Lancet Neurol* 2017; 16 (3): 183–4.
5. Kandel ER, Markram H, Matthews PM, *et al.* Neuroscience thinks big (and collaboratively). *Nat Rev Neurosci* 2013; 14 (9): 659–64.
6. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science* 2007; 316 (5827): 1036–9.
7. Coccia M, Wang L. Evolution and convergence of the patterns of international scientific collaboration. *Proc Natl Acad Sci USA* 2016; 113 (8): 2057–61.
8. Jirotko M, Lee CP, Olson GM. Supporting scientific collaboration: methods, tools and concepts. *Comput Supported Coop Work* 2013; 22 (4–6): 667–715.
9. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 2016; 22 (2): 303–41.
10. Padova Y. The Safe Harbour is invalid: what tools remain for data transfers and what comes next? *Int Data Privacy Law* 2016; 6 (2): 139–61.
11. Mayer-Schönberger V, Cukier K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: Houghton Mifflin Harcourt; 2013.
12. Rothstein MA. Ethical issues in big data health research: currents in contemporary bioethics. *J Law Med Ethics* 2015; 43 (2): 425–9.
13. Toga AW, Foster I, Kesselman C, *et al.* Big biomedical data as the key resource for discovery science. *J Am Med Inform Assoc* 2015; 22: 1126–31.
14. Bollier D, Firestone CM. *The Promise and Peril of Big Data*. Washington DC: The Aspen Institute; 2010.
15. Boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Inform Comm Soc* 2012; 15 (5): 662–79.
16. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books; 2017.
17. Bietz MJ, Bloss CS, Calvert S, *et al.* Opportunities and challenges in the use of personal health data for health research. *J Am Med Inform Assoc* 2016; 23 (e1): e42–8.
18. Chee HL. Ownership, control, and contention: challenges for the future of healthcare in Malaysia. *Soc Sci Med* 2008; 66 (10): 2145.
19. Cheung C, Bietz MJ, Patrick K, *et al.* Privacy attitudes among early adopters of emerging health technologies. *PLoS One* 2016; 11 (11): e0166389.
20. Kilkenny C, Browne W, Cuthill IC, *et al.* Animal research: reporting in vivo experiments: the ARRIVE guidelines. *Br J Pharmacol* 2010; 160 (7): 1577–9.
21. Sugarman J. Human brain research and ethics. In: Clausen J, Levy N, eds. *Handbook of Neuroethics*. Netherlands: Springer; 2015: 1103–6.
22. Fan W, Bifet A. Mining big data: current status, and forecast to the future. *SIGKDD Explor Newsl* 2013; 14 (2): 1–5.
23. Denning T, Matsuoka Y, Kohno T. Neurosecurity: security and privacy for neural devices. *Neurosurg Focus* 2009; 27 (1): E7.
24. Sepai O, Collier C, Van Tongelen B, *et al.* Human biomonitoring data interpretation and ethics; obstacles or surmountable challenges? *Environ Health* 2008; 7 (Suppl 1): S13.
25. Farah MJ. Neuroethics: the practical and the philosophical. *Trends Cogn Sci (Regul Ed)* 2005; 9 (1): 34–40.

26. Farah MJ. *Neuroethics: An Introduction with Readings*. Cambridge, MA: MIT Press; 2010.
27. Illes J, Sahakian BJ. *Oxford Handbook of Neuroethics*. Reprint edition. Oxford, UK: OUP Oxford; 2013.
28. Marcus S, Foundation CAD. *Neuroethics: Mapping the Field: Conference Proceedings*, Dana Press, May 13–14, 2002; San Francisco, CA.
29. Khatri V, Brown CV. Designing data governance. *Commun Acm* 2010; 53 (1): 148.
30. Floridi L, Taddeo M. What is data ethics? *Phil Trans R Soc A* 2016; 374 (2083): 20160360.
31. Hripsak G, Mirhaji P, Low AF, et al. Preserving temporal relations in clinical data while maintaining privacy. *J Am Med Inform Assoc* 2016; 23 (6): 1040–5.
32. Floyd SW, Schroeder DM, Finn DM. Only if I'm First Author™: conflict over credit in management scholarship. *Acad Manage J* 1994; 37 (3): 734–47.
33. Shabani M, Dyke SOM, Joly Y, et al. Controlled access under review: improving the governance of genomic data access. *PLoS Biol* 2015; 13 (12): e1002339.
34. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. *Ann Intern Med* 2016; 164 (7): 505.
35. Evers K. The contribution of neuroethics to international brain research initiatives. *Nat Rev Neurosci* 2017; 18 (1): 1–2.
36. Dove ES, Townend D, Meslin EM, et al. Ethics review for international data-intensive research. *Science* 2016; 351 (6280): 1399–400.
37. Integrating ethics and science in the International HapMap Project. *Nat Rev Genet* 2004; 5 (6): 467–75.
38. Rogerson S. Ethics and ICT. In: Galliers RD, Currie W, eds. *The Oxford Handbook of Management Information Systems: Critical Perspectives and New Directions*. Oxford: OUP Oxford; 2011: 601–22.
39. Kim H, Bell E, Kim J, et al. iCONCUR: informed consent for clinical data and bio-sample use for research. *J Am Med Inform Assoc* 2017; 24 (2): 380–7. doi: 10.1093/jamia/ocw115
40. Kaye J. Abandoning informed consent: the case of genetic research in population collections. 2004: 267412. <https://repository.library.georgetown.edu/handle/10822/503951> Accessed February 20, 2017.
41. Ye C, Giangregorio L, Holbrook A, et al. Data withdrawal in randomized controlled trials: defining the problem and proposing solutions: a commentary. *Contemp Clin Trials* 2011; 32 (3): 318–22.
42. Bezuidenhout L. Data sharing and dual-use issues. *Sci Eng Ethics* 2013; 19 (1): 83–92.
43. Citizen Meetings in the Human Brain Project | Fonden Teknologirådet. <http://www.tekno.dk/article/citizen-meetings-in-the-human-brain-project/?lang=en> Accessed February 22, 2017.
44. Zimmer M. But the data is already public™: on the ethics of research in Facebook. *Ethics Inf Technol* 2010; 12 (4): 313–25.
45. Liyanage H, Liaw S-T, Di Iorio CT, et al. Building a privacy, ethics, and data access framework for real world computerised medical record system data: a Delphi Study. Contribution of the Primary Health Care Informatics Working Group. *Yearb Med Inform* 2016; (1): 138–45.
46. García-Marzá D. Trust and dialogue: theoretical approaches to ethics auditing. *J Bus Ethics* 2005; 57 (3): 209–19.
47. Cavusoglu H, Benbasat I, Bulgurcu B. Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly* 2010; 34: 523–48.
48. Warkentin M, Willison R. Behavioral and policy issues in information systems security: the insider threat. *Eur J Inf Syst* 2009; 18 (2): 101–5.
49. Stahl BC, Doherty NF, Shaw M. Information security policies in the UK healthcare sector: a critical evaluation. *Inform Syst J* 2012. doi: 10.1111/j.1365-2575.2011.00378.x.
50. Aronson J. A pragmatic view of thematic analysis. *Qual Rep* 1995; 2: 1–3.
51. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006; 3 (2): 77–101.
52. Myers MD, Walsham G. Exemplifying interpretive research in information systems: an overview. *J Inf Technol* 1998; 13 (4): 233–4.
53. Trauth EM, Jessup LM. Understanding computer-mediated discussions: positivist and interpretive analyses of group support system use. *MIS Quarterly* 2000; 24 (1): 43–79.
54. Zimbardo PG. On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition* 1973; 2 (2): 243–56.
55. Stark L. *Behind Closed Doors: IRBs and the Making of Ethical Research*. 1st ed. Chicago: University of Chicago Press; 2011.
56. Schrag ZM. *Ethical Imperialism: Institutional Review Boards and the Social Sciences, 1965-2009*. 1st ed. Baltimore: Johns Hopkins University Press; 2010.
57. Klitzman R. *The Ethics Police?: the Struggle to Make Human Research Safe*. 1st ed. Oxford; New York: OUP USA; 2015.
58. OECD. Recommendation of the OECD Council on Health Data Governance. 2017. <http://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf> Accessed March 1, 2018.