# Metadata of the chapter that will be visualized online

| | | |
|---|---|---|
| Chapter Title | On the Compatibility of Big Data Driven Research and Informed Consent: The Example of the Human Brain Project | |
| Copyright Year | 2016 | |
| Copyright Holder | Springer International Publishing Switzerland | |
| Corresponding Author | Family Name | **Christen** |
| | Particle | |
| | Given Name | **Markus** |
| | Suffix | |
| | Organization | University Research Priority Program Ethics, University of Zurich |
| | Address | Zollikerstrasse 117, 8008, Zurich, Switzerland |
| | Email | christen@ethik.uzh.ch |
| Author | Family Name | **Domingo-Ferrer** |
| | Particle | |
| | Given Name | **Josep** |
| | Suffix | |
| | Division | Department of Computer Engineering and Mathematics UNESCO Chair in Data Privacy |
| | Organization | Universitat Rovira i Virgili |
| | Address | Av. Països Catalans 26, 43007, Tarragona, Catalonia, Spain |
| | Email | josep.domingo@urv.cat |
| Author | Family Name | **Draganski** |
| | Particle | |
| | Given Name | **Bogdan** |
| | Suffix | |
| | Division | Laboratoire de Recherche en Neuroimagerie, Department of Clinical Neurosciences |
| | Organization | Centre Hospitalier Universitaire Vaudois |
| | Address | Mont Paisible 16, CH-1011, Lausanne, Switzerland |
| | Email | bogdan.draganski@gmail.com |
| Author | Family Name | **Spranger** |
| | Particle | |

| | | |
|---|---|---|
| | Given Name | **Tade** |
| | Suffix | |
| | Division | Faculty of Law/Institute of Science and Ethics |
| | Organization | University of Bonn |
| | Address | Bonner Talweg 57, 53113, Bonn, Germany |
| | Email | spranger@iwe.uni-bonn.de |
| Author | Family Name | **Walter** |
| | Particle | |
| | Given Name | **Henrik** |
| | Suffix | |
| | Division | Division of Mind and Brain Research, Department of Psychiatry and Psychotherapy |
| | Organization | Charité – Universitätsmedizin Berlin, Campus Mitte |
| | Address | Charitéplatz 1, D-10117, Berlin, Germany |



Abstract

Big Data research is usually explorative, meaning that not all possible hypotheses are known that one may wish to test when data is made available. For the case of biomedical data this poses a significant challenge, as the originators of the data – patients or research participants – have to provide informed consent for using their data. The typically obtained "closed" or "narrow consent", i.e. consenting to use the data in a well-defined research project, is conceptually incompatible with the explorative nature of Big Data driven research. Therefore, "open" or "broad consent" is proposed as an alternative. Nevertheless, open consent cannot justify any type of data use, but requires an "information framework" that separates legitimate from illegitimate Big Data research. For example, consent is given associated with established disease categories: a patient diagnosed with early-onset Alzheimer's disease may consent to his personal medical information being used for any research enhancing our understanding of this particular disease. In our contribution, we address the question whether and how Big Data driven research may undermine this "information framework" of informed consent using the example of the Human Brain Project (HBP). Within the HBP, a Big Data infrastructure is currently being developed to access a multitude of clinical data related to brain diseases based on the conviction that many neurological and psychiatric disorders and diseases are ill-defined in terms of underlying mechanisms. We analyse the interrelation between effects of Big Data research and informed consent and we evaluate ethical and practical consequences.

# On the Compatibility of Big Data Driven Research and Informed Consent: The Example of the Human Brain Project

**Markus Christen, Josep Domingo-Ferrer, Bogdan Draganski, Tade Spranger, and Henrik Walter**

**Abstract** Big Data research is usually explorative, meaning that not all possible hypotheses are known that one may wish to test when data is made available. For the case of biomedical data this poses a significant challenge, as the originators of the data – patients or research participants – have to provide informed consent for using their data. The typically obtained "closed" or "narrow consent", i.e. consenting to use the data in a well-defined research project, is conceptually incompatible with the explorative nature of Big Data driven research. Therefore, "open" or "broad consent" is proposed as an alternative. Nevertheless, open consent cannot justify any type of data use, but requires an "information framework" that separates legitimate from illegitimate Big Data research. For example, consent is given associated with established disease categories: a patient diagnosed with early-onset Alzheimer's disease may consent to his personal medical information being used for any research enhancing our understanding of this particular disease. In our contribution, we address the question whether and how Big Data driven research may undermine

M. Christen (✉)
AQ1 University Research Priority Program Ethics, University of Zurich, Zollikerstrasse 117, 8008 Zurich, Switzerland
e-mail: christen@ethik.uzh.ch

J. Domingo-Ferrer
Department of Computer Engineering and Mathematics UNESCO Chair in Data Privacy, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
e-mail: josep.domingo@urv.cat

B. Draganski
Laboratoire de Recherche en Neuroimagerie, Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois, Mont Paisible 16, CH-1011 Lausanne, Switzerland
e-mail: bogdan.draganski@gmail.com

T. Spranger
Faculty of Law/Institute of Science and Ethics, University of Bonn, Bonner Talweg 57, 53113 Bonn, Germany
e-mail: spranger@iwe.uni-bonn.de

H. Walter
Division of Mind and Brain Research, Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Campus Mitte, Charitéplatz 1, D-10117 Berlin, Germany


B.D. Mittelstadt, L. Floridi (eds.), *The Ethics of Biomedical Big Data*, Law, Governance and Technology Series 29, DOI 10.1007/978-3-319-33525-4_9

this "information framework" of informed consent using the example of the Human Brain Project (HBP). Within the HBP, a Big Data infrastructure is currently being developed to access a multitude of clinical data related to brain diseases based on the conviction that many neurological and psychiatric disorders and diseases are ill-defined in terms of underlying mechanisms. We analyse the interrelation between effects of Big Data research and informed consent and we evaluate ethical and practical consequences.

## 1 Introduction

Modern biomedical research as well as the ongoing digitalization of healthcare systems is creating an enormous amount of data that has the potential to significantly change our understanding of various diseases. Previous examples of scientific milestones achieved through advances in information technology include the steadily growing number of Internet accessible sequence databases in molecular biology since the early 1980s with its emanation – The Human Genome Project. Neuroscience[1] has clearly taken a similar direction, which is illustrated by several new initiatives for data sharing and common databases. Such initiatives are deemed to be necessary given the massive output of this field. It is estimated that more than 100,000 papers a year are published in neuroscience (Grillner 2014) – most of them involving the analysis of data of various kinds, from genetic data and electrophysiology measurements up to imaging and behavioural data. Compared to other fields like molecular genetics, however, the large majority of neuroscience data sets are still small due to the complexity of the research needed for generating them.[2] Furthermore, data-sharing standards are often lacking. Such small data sets have been referred to as "long-tail" data and may in the future become an important source of new findings (Ferguson et al. 2014).

This traditional focus of neuroscience on "small science" and "small data" comes increasingly under pressure due to recent "big neuroscience" initiatives (Christen et al. accepted). Several Big Data projects are underway to access both small and big data sets generated through research in neuroscience – a development that is exemplified by the "Big Data" issue of *Nature Neuroscience* in November 2014. While many of these efforts focus on model animals, Big Data is also being generated from humans. For example, the amount of openly available and shared neuroimaging data has increased substantially in the last few years (Poldrack and Gorgolewski 2014; Thompson et al. 2014). Even larger data sets concern

[1] In the following, we use a wide understanding of neuroscience, including also medical fields that deal with neurological or brain diseases like neurology, neuropsychology or psychiatry.

[2] Examples include morphological reconstructions of neurons (which is very time-consuming), research with nonhuman primates (which is highly regulated and expensive) or neuroimaging research (which requires a costly infrastructure).

whole-genome sequencing data and the increasing use of technologies for creating large transcriptomic and epigenetic data sets from brain tissue (Shin et al. 2014).

In the following, we will focus on particular Big Data initiatives that are integrated in the Human Brain Project (HBP). The HBP was announced in January 2013 as one of two flagship projects funded by the European Commission's Future and Emerging Technologies Programme. The matched funding for the HBP of about 1.16 billion Euros over 10 years provided by European Union (EU) and partners shall enable a concerted effort to "lay the technical foundations for a new model of ICT (information and communication technologies) based brain research, driving integration between data and knowledge from different disciplines, and catalysing a community effort to achieve a new understanding of the brain, new treatments for brain disease and new brain-like computing technologies" (HBP Report 2012, 3). A major goal of the project involves data integration, for which the HBP is developing six ICT-based platforms dedicated, respectively, to Neuroinformatics, Brain Simulation, High Performance Computing, Medical Informatics, Neuromorphic Computing, and Neurorobotics (for detailed information: https://www.humanbrainproject.eu/). Those platforms are intended to allow sharing research data of all levels of neuronal integration (related, e.g., to ion channel structures, synapse distributions, neuronal microcircuits, brain connectivity patterns, or functional imaging data), methods and models (e.g., in form of computer programs, respectively code) and accessing databases that contain a multitude of clinical data related to brain diseases – the latter will be provided by the Medical Informatics platform and is described in more detail in Sect. 4.

There are certainly many ethical issues associated to data generation (e.g., animal experimentation) and data sharing in neuroscience (e.g., allocation of scientific credit when publishing results originating from shared data). But our focus here is on the problem of informed consent when the data emerges from human subjects, which researchers are required to obtain by current data protection legislation in European countries. Traditionally,[3] "closed" or "narrow consent" is provided, i.e. patients or research participants consent to only one or a few specific uses of the data in a well-defined research project. This, however, is conceptually incompatible with the nature of Big Data driven research that seeks patterns in data based on hypotheses that are often not known when the data has been collected. Therefore, a growing number of researchers and legislators propose "open" or "broad consent" as an alternative, meaning that consent is given to using data for broader research fields or – as a maximum – for any form of research (for an example in genetics, see Lunshof et al. 2008). As we will outline below, such a broad consent poses ethical challenges. These are increased in the case of human brain data, as such data

[3] Seen from a broader historic perspective, (closed) informed consent is a rather recent phenomenon, but can now be considered as standard at least in research settings in industrialised countries. In this contribution, we refrain from outlining the history of informed consent and of international differences in the understanding of informed consent.

is by nature sensitive even if it does not contain healthcare information, because it contains information about the organ of the mind and thus to a certain extent also about the mind itself.

In our contribution we are particularly interested in a potential conflict that is posed by Big Data research in neuroscience, especially when the research is related to neurological or psychiatric diseases. On the one hand, despite the consent being "open", it requires specifying *some information* about what the person is consenting to; otherwise the consent cannot be called "informed". Thus, any form of *informed* consent is embedded in an "information framework" that outlines the general context in which the data is generated, what kind of data is actually obtained, and – although not exhaustively and still in rather general terms – what kind of results could be expected through analysing the data. A plausible and frequent way of generating this information framework is by referring to disease categories – we call this the disease space ontology. For example, a patient diagnosed with early-onset Alzheimer's disease may consent to his personal medical information – health record data, genetic data, neuroimaging data etc. – being used for any type of research enhancing our understanding of Alzheimer's disease.

On the other hand, there is as long-standing discussion in neurology and psychiatry that many current neurological and psychiatric disorders and diseases are ill-defined in terms of underlying mechanisms (Owen 2014; Thagard 2008). On the example of Major Depressive Disorder representing a separate disorder category according to DSM-5, there may be a different classification with a number of subtypes depending on a variety of underlying biological mechanisms. Some types of depressive syndromes may in fact turn out to be other disorders, whereas some might turn out to be subsumed under a disease category with known causes and mechanism and not just a syndrome, i.e. a heterogeneous cluster of symptoms (Monroe and Anderson 2015). Taking these two developments together, it could be that the standard way of providing an "information framework" through disease categories is likely to be shattered through research that necessarily relies on Big Data approaches, in particular in case of brain diseases. We take this apparent paradox as a starting point to explore the connection between the information framework of informed consent and Big Data research that may affect this framework.

This question will be approached in our contribution from various angles. First, we briefly outline the problem of neuroscience-informed disease categorisation with a particular focus on psychiatric diseases. This should motivate the claim that changing the disease space ontology could have an effect on the practice of giving informed consent. Second, we describe in detail the current setup of data collection and informed consent practice within the HBP intended to improve and change our understanding of disease categories in neuroscience. In this way we want to outline that significant changes with respect to our understanding of brain diseases are not a mere theoretical scenario. Third, we discuss the legal problems of open informed consent practices and their dependence on an information framework. In this context, specific attention is paid not only to existing data protection law, but also to legislation aiming at the protection of research participants (e.g. the Council of Europe´s Convention on Human Rights and Biomedicine). Fourth, we evaluate the

underlying moral justifications for upholding or transgressing certain "information borders" in terms of information spheres following the proposals of Nissenbaum (2004) and van den Hoven (2008). Finally, we sketch novel technological solutions for addressing this problem by referring to concepts like traceability of data use and verifiable anonymisation.

## 2 Disease Categorisation in Psychiatry from a Neuroscientific Point of View

The human brain is among the most complex structures that are object of scientific investigation and it is therefore not surprising that brain diseases are hard to understand. Broadly construed, neurological and psychiatric diseases can be defined as disorders of the brain. There is a continuum of disorders with respect to the degree of their scientific understanding. In some cases, the neurobiological cause is simple and known (e.g. a specific genetic aberration on chromosome 4 in Huntington's disease). Other disorders are diagnostically well-defined and there is a considerable body of knowledge available regarding their underlying mechanisms (e.g. neurodegeneration of dopaminergic neurons in Parkinson's disease). Yet other disorders are difficult to diagnose (in particular in the early phase) and competing theories are available regarding the pathophysiological mechanisms (e.g., Alzheimer's disease). Finally, in many frequent disorders although neurobiological knowledge is available, but rather limited and their definitions today still rely on clinical signs, symptoms and duration (most psychiatric disorders like schizophrenia or depression). In the following, we will focus on the relation between disease categorisation and Big Data driven research for psychiatric disorders, as strong hopes, even promises, have been raised that those approaches can improve knowledge and subsequently therapy (Owen 2014; Wang and Krystal 2008).

According to the two most influential manuals for categorising psychiatric disorders, the IDC-10 of the World Health Organization and the DSM-5 of the American Psychiatric Association, the diagnosis of a psychiatric disorder rest only on clinical features, i.e. on the presence of a specified number of certain symptoms for a specified duration and the exclusion of certain specified causes, like a "organic" disease or an intoxication. The disorder concept of DSM and ICD is categorical: either you have the disease or you don't – although disorder are characterised by different degrees of severity, e.g. for depression. ICD-10 as well as DSM-5 do not rely on underlying pathophysiological mechanisms as most of them are not known, heavily debated, or can only be diagnosed post-mortem.

Between the publication of DSM IV (released in 2004) and DSM-5 (released in May 2013) it was hoped that the new DSM-5 would advance the field considerably with respect to two issues: integrating dimensional approaches (i.e. use constellation of symptom dimensions instead of categories for example for the diagnosis of personality disorders) and integrating neurobiological criteria (genetic, molecular, neuroimaging) for making diagnoses. Suggestions in this direction were intensively

discussed by the DSM-task force of the American Psychiatric Association over several years. At the end, however, none of these conceptual changes were included in DSM-5. This was largely because it was felt that neurobiological knowledge was not (yet) reliable enough, but also due to the fact that the DSM is much more than a medical nosology: it also serves a central societal role by providing the basis for mental health care and thus is conservative in nature as changes would immediately affect millions of patients and carefully balanced systems of providers and consumers.

This missing integration of neurobiological knowledge frustrated many mental health scientists. In fact, 3 weeks before the official release of DSM-5, Thomas Insel, at that time director of the National Institute of Mental Health (NIMHM),[4] the largest research institute for mental health in the western world, launched a considerable attack on DSM-5 by declaring in his blog that "the weakness (of DSM-5) is its lack of validity. Unlike our definitions of ischemic heart disease, lymphoma, or AIDS, the DSM diagnoses are based on a consensus about clusters of clinical symptoms, not any objective laboratory measure. In the rest of medicine, this would be equivalent to creating a diagnostic system based on the nature of chest pain or the quality of fever. Indeed, symptom-based diagnosis, once common in other areas of medicine, has been largely replaced in the past half century as we have understood that symptoms alone rarely indicate the best choice of treatment" (Insel 2013). The apparent lack of availability of reliable biomarkers for mental disorders was explained by Insel as a conceptual rather than as an empirical problem: it would be equivalent to rejecting the usefulness of the electrocardiogram (ECG) as a diagnostic tool, only because many patients with chest pain do not have ECG changes. In fact it was the ECG which allowed differentiating chest pain due to specific heart problems from other forms of chest pain, i.e. the tool helps to categorise the disorders by measuring physiological processes. And, according to Insel, the same should be done in psychiatry by "collecting genetic, imaging, physiological, and cognitive data to see how all the data – not just the symptoms – cluster and how these clusters are related to treatment response" (Insel 2013). Such an approach is only possible using Big Data techniques, as we will outline in the next section.

In fact, the NIMH started a research program some years ago which is now known under the name Research Domain Criteria (R-DOC; Morris and Cuthbert 2012). The basic idea of this approach is to achieve a dimensional characterisation of mental illness as mentioned above in order to discover, refine or reclassify mental disorders. For this purpose, it is suggested to study diseases based on a two-dimensional grid based on current neurocognitive and molecular approaches and knowledge. One dimension consists of five core domains of mental functioning ("systems") that have been determined by consensus conferences of active scientists

[4] Interestingly, in particular with respect to the increasing role of ICT for (mental) health, Thomas Insel announced in September 2015 after 13 years serving as director of the NIMH that from November 2015 on he will move to Alphabet, the umbrella organization of Google in order to help to develop mobile health technologies.

from the field, i.e. systems for negative valence, positive valence, cognition, social processes and arousal. Each of these domains has subdomains, e.g. the system for negative valence comprises the subdomains active threat ("fear"), potential threat ("anxiety"), sustained threat, loss and frustrative non-reward. The other dimensions refer to levels of organisation on which the constructs within the domains can be measured: from genes, molecules, cells, circuits, physiology, to behaviour, self-reports, and paradigms. By filling this 2-dimensional grid with scientific results, it will, in the long run, be possible to characterise mental disorders on a sound empirical basis and detect patterns leading to the discovery of new disorders or reclassification of new ones. These discussions on a new understanding of mental disorders as disorders of neurocognitive domains are also referred to as the "third wave of biological psychiatry" (Walter 2013).

However, for this approach to being realised, a revolution, or at least a reform of disease concepts is required. It also would entail Big Data neuroscience on mental health: only if you have obtained enough high dimensional data from many domains of many subjects together with clinical data, this approach might become successful. But standard DSM-based research uses the (not-so) gold(en) standard of symptom-based categories and will thus make no progress. Therefore, Insel has announced that the NIMH will in the future not fund research based on "old" still gold-standard disease categories, but rather RDOC-oriented, dimensional research. To take a simple example, it would not fund neurobiological research on alcohol addiction, but rather neurobiological research on impulsivity as a contributor to alcohol drinking.

But what would such a change induce on the level of actual researchers who have to interact with patients and research subjects and obtain their informed consent for using their data? Consenting to the use of data in research obviously requires a basic understanding on the context in which the data has been generated and in which it is likely to be used. Lay people like patients usually do not have the competences needed to assess the detailed hypotheses of research in which they are involved, e.g. when they are asked to participate in a clinical trial for testing a new medication. Although such detailed information is not required, as the main interest of the patient probably is to obtain information on possible health risks and benefits – this type of information is still presented in a context framed by the disease from which the patient is suffering. Taking the simple example from above, a patient with a severe drinking problem would probably expect that the research in which he is involved relates to alcohol addiction and not to some research on impulsivity, as the person may consider impulsivity (to some degree at least) as a legitimate aspect of his personality. Thus, the specific disease along with a laymen understanding of what, e.g., a depression or Alzheimer's dementia involves, is crucial for putting the informed consent into a context.

This context also affects the moral significance of diseases. A disorder caused by a genetic factor (e.g. Huntington's disease) is associated with specific types of moral problems (e.g., related to inheriting the disease) that are not perceived to be present in neurodegenerative disorder. Some brain disorders are associated with a stronger stigma than others (e.g. schizophrenia versus epilepsy). Yet some disorders are

understood to be clearly "brain based" (e.g., Parkinson's disease), whereas others are much more associated to "external" (e.g., social or cultural) causes, although it is likely that changes in the brain play an important role in the disease course (e.g., anorexia nervosa) – and such "external causes" involve a different responsibility relation (e.g. by avoiding certain social settings or by generating an imperative to change certain societal aspects through policy interventions). Consenting to use data related to one disease may thus mean something different than consenting to contribute data related to another disease or to broader spectrum of diseases relevant to specific domains of functioning.

If now a research program is installed that seeks connections between neuronal diseases that lay people consider rather different, should they be informed on these possible links? For example: should a Parkinson's disease patient be informed that analysing her data may help to understand schizophrenia or depression – and in this way implicitly given her some reason to suspect that she might suffer also from one those diseases? Actually, the re-conceptualised disease space ontology may look very different compared to the disease space that frames the current social handling of these diseases in terms of physician-patient relation, health insurance, or stigmatisation. Here, we try to sketch possible ethical consequences of such a change in the disease space. But before that, we outline the actual possibility that such a change could happen (Sect. 4) and the current legal setting related to informed consent (Sect. 5).

## 3 Data Collection, Informed Consent and the Human Brain Project

Every day, an impressive amount of data related to brain health and disease are produced in clinical and research establishments across Europe. Usually, these data are in the format of descriptive clinical data, laboratory results or brain images that serve to help medical decision-making. They are viewed mostly only once before being archived on departmental or laboratory servers for a finite number of years. This mass of data constitutes an enormous research resource that is currently largely unused. Though the data are collected at different sites, it has now been demonstrated that the variance introduced by analysing data from multiple imaging platforms or clinical chemistry laboratories is much smaller than the variance that is attributable to the disease (Stonnington et al. 2008). In other words, variability through differences in methodological practice can be controlled. This fact suggests an opportunity to use archived data for the pathophysiological, anatomical and medical studies on a population basis. This is a major motivation of the data integration strategy of the Human Brain Project (HBP) in the medical informatics platform.

Recent advances in computing and commercially available algorithms for federating data from local databases that work unobtrusively in the background in real

time make such a project practicable and cost-effective. The Medical Informatics platform of the HBP proposed an initial programme based on federation of data related to brain diseases to establish feasibility, sharing protocols, data usage agreements, access protocols and other issues. This idea represents a quantum leap from the path trodden out in the past by successful database initiatives such as the *Alzheimer's Disease Neuroimaging Initiative* (ADNI, see www.adni-info.org), which is used by many researchers world-wide although it is much smaller in scope and more expensive because the technology was not available at the time of its setup.

From an ethical point-of-view, the mass of brain health and disease related information collected in hospitals, clinics and research establishments is grossly underused at present, which represents an extraordinary waste of resources. With advances in modern information technology, especially in terms of massive data storage and access to hardware, analysis tools and data mining techniques, these data can be used to carry out a range of studies of social and medical importance. The range of possible investigations is enormous, if the data can be systematised and intelligently mined. The main goal of the Medical Informatics platform of the HBP is to federate and integrate clinical and basic science research together with information technology and establish new ways for open access to shared aggregate data in order to ask hypothesis driven questions, to mine data, to carry out epidemiological, genetic and other surveys. But certainly, the question emerges whether the practice of large-scale access to this data is compatible with the informed consent given by the patients from which this information emerges. We will come back to this point later and we first outline the technical procedures of data collection.

The complexity of clinical data especially in the field of neuroscience makes evident the need for a coherent framework for integrating the multiple temporal and spatial scales of data to facilitate its interpretation. Ongoing large-scale projects (e.g., ENIGMA, Human Connectome Project, Allen Brain ATLAS, GENSAT) demonstrate that brain imaging data capturing *in vivo* anatomical and functional information about the brain can serve as a backbone for developing a viable framework for research data integration. From a clinical perspective, the more prominent examples are the recent developments in the neuro-epidemiology of dementia based on differential patterns of cortical atrophy associated with cognitive decline; the development of biomarkers from analysis of scans and subsequent cognitive outcome or neuro-pathological examination; population wide genetic association with *in vivo* pathology studies, as demonstrated by image-derived brain tissue characterisation.

To give a very specific example that illustrates what could become possible, there is a pressing demographic and economic need to answer questions about the preclinical stage of dementia, in particular the incidence and natural history of pathological change, early detection and diagnoses based on brain measures rather than behavioural expression, and how to monitor the rate of pathological brain changes on sufficient numbers of people such that the results are generalizable. The repercussions of the results will be important because there is preliminary evidence

to suggest that the dementias can be differentiated by early distribution of brain atrophy. It should therefore be possible to identify purer cohorts of the different dementia-associated diseases than is now possible to identify, and to test and develop new specific disease-modifying drugs. This type of research could eventually lead to a new classification of dementia-associated diseases that is quite distinct from today's understanding. This would be an example of a re-conceptualized disease space through Big Data research.

The use of data mining – the technological precondition for restructuring the disease space – involves the extraction of patterns from large sets both for scientific and business related queries. The use of this technique has exploded in the last few decades in many fields in biomedicine, as outlined in Sect. 2. Considering the remarkable advances in biomedical imaging technology and analysis, data mining offers new opportunities capitalising on the ability to extract characteristic features from abundant and diverse information about human (patho-) anatomy and physiology. The creation of disease-specific neuroimaging data repositories (ADNI, ENIGMA, IMAGEN) represents first attempts to use advanced neuro-informatics methodologies for databases of clinically relevant information. Although offering standardised data processing of anatomical brain images, these databases serve mainly as repositories rather than frameworks for data mining on clinical neuroscience grounds. Data mining approaches are aimed at making use of the large data set in order to extract main predictors that explain variance in the data. An understanding of the nature and extent of inter-subject variation is critical for the characterisation of the neural basis of cognitive processes in healthy subjects and the changes that cause abnormal functioning. Data mining approaches build upon the decomposition of inter-individual differences to create meaningful classifications of subjects and predictions of continuous variables such as behaviour or performance. The principal hypothesis is that characteristic distributions of variability of the structure of the brain and its connectivity patterns will be of diagnostic value through identification of disease discriminative patterns.

The Medical Informatics platform of the HBP (see www.humanbrainproject.eu/medical-informatics-platform; Frackowiak and Markram 2015) is building on a concept for data federation that allows mining all available resources without the need to directly access the original data. Rather than copying, downloading and mirroring data, the current set-up focuses on locally creating data aggregates, which provide a summary of the available data at a particular site. These aggregates are feature-specific and can be queried by the end-user in the form of double-aggregated data. At no instance is there access to individual-specific data, which could open the possibility for data misuse and identification of a given person. The combination of simple database language queries and advanced methodological tools for statistical inference and learning allows harvesting the aggregated data, binding multiple sources of information and extracting characteristic features to answer domain-specific questions. This is a dynamic process that aims to create clinical generative models of specific diseases. As more data is gathered models can be re-evaluated and refined to answer more subtle questions.

The processing of the data to extract features for aggregation is performed locally within the secured systems of hospitals based on the availability of powerful algorithms able to handle vast amounts of data. Several issues surrounding big data analysis on the Medical Informatics platform of the HBP need consideration in relation to data protection. Legally binding laws enforced by the EU authorities stipulate that responsibility for the data and its ownership is transparent (see also Sect. 5). Although our framework does not allow accessing individual data, current laws and regulation apply to data transfer, data processing and data security and the Human Brain Project has to ensure that data management is compliant with data protection law. This also means that beyond strict procedures for data anonymisation, data preparation for mining should be restricted to well-defined workflows that prevent data miners from identifying specific individuals or from uncovering confidential information. This protection of privacy is – from an ethical point of view – the uncontroversial part of the problem. It is also in the focus of current legislation, as we outline below. But our question is, whether privacy is the only and main concern of Big Data driven research in neuroscience.

## 4 Legal Issues of Open Consent and Its Information Basis

The prevailing Data Protection Law applicable in all EU Member States is mainly based on European legal guidelines. Debates over the minutiae of a new EU Data Protection Regulation (anticipated to be passed around 2016) are fully underway.[5] Particularly the question of how this new Law will affect the use of personal data in a scientific context is one of the main aspects in need of clarification. However, there is a large consensus that the mere, indiscriminate adoption of general data protection standards for scientific work could pose an unnecessary and unjustified restriction of the freedom of research.

Furthermore, specific problems arise regarding the practical implementation of so-called informed consent. Originally developed as a bioethical principle, the notion of informed consent can be found in many documents of international and national law (e.g. the Council of Europe's Convention on Human Rights and Biomedicine) today and has thus become an integral part of the Positive Law. In 'classical' research projects such as in clinical trials it has been shown, however, that providing (too much) information to the research subject can occasionally lead to the opposite effect of what the informed consent aims at; excess of information can leave the concerned party unable to make a (truly) informed choice after all. This problem exacerbates in Big Data driven research areas as outlined above: one cannot effectively communicate the potentially enormous range of testable hypotheses to patients. Therefore it has to be examined which models of informed consent can

[5] An overview on the legislation procedure is available here: http://ec.europa.eu/justice/data-protection/

be used or further developed, that protect research participants on the basis of legal compliance and yet do not disproportionately restrict the efficiency of research.

However, integrating informed consent into Big Data driven research also touches upon the question whether or not the concept as such leads to an adequate protection of research participants. Especially with regard to other contexts, e.g. establishing so-called bio-banks, it was and still is discussed, if informed consent in its classic understanding is sufficient to protect the rights and interests of research participants in a sufficient manner (D'Abramo 2015; Hofmann 2009).

Due to the complexity and high dynamics of modern biomedical research, it is stressed that research participants may not realise the full implications of giving their consent. While agreeing to the use of their samples or test results for 'the purpose of research', participants may have a lack of understanding what exactly that vaguely phrased expression means (Cordasco 2013). Therefore, in a legal context, the restriction of the range of the informed consent has been consistently demanded. A restriction may be imposed with regard to a certain time frame (e.g. informed consent is given for a time period of 5 or 10 years, in combination with the obligation to newly clarify the purpose of the use of the elevated personal data in order to attain a renewed consent of the participants) or regarding a factual aspect which would restrict the given informed consent to a particular project or to the research of a specific disease pattern.

However, the approaches of restricting either the temporal or the factual context of informed consent fail to work even with regard to small-scale projects: a renewed informed consent can usually not be obtained due to research participants moving away or dying in the meantime. Furthermore, the maintenance of an address register would not only go beyond the scope of time effort, but also be a great financial burden to any project and may actually generate new privacy risks due to problems in securing this information from unauthorized access.

A restriction of the informed consent to a certain factual context is problematic as well, because Big Data research aims for a cooperation and combination of different projects, and not for individual projects. In addition, undertaking a follow-up project would be made impossible for the researcher who got the informed consent in the first place. Lastly, the restriction to only one specific disease pattern is problematic as well due to the difficulty of insufficient clarity and changing definitions and understandings of a certain disease – as we have outlined in detail in Sect. 3.

Regarding the reasons mentioned above, jurisprudence represents a general permissibility of a 'broad consent' which is of unlimited time and enables largely unrestricted factual research. As far as some legal systems assume an inadmissibility of a 'general consent', this concept deals with the consent given by a third person to carry out any kind of legal action and cannot be compared with the approach and content of a 'broad consent' (see for the case of biobanks: Serepkaite et al. 2014). The latter does not mean that contributors of genetic material or data do not obtain any rights. Personal rights and data protection laws as well as privacy issues obviously have to be respected. Therefore, the current legal understanding of the problem of Big Data driven research focuses on demanding technological solutions that ensure that privacy and data protection are respected, mainly through

aggregation and anonymization techniques or – more generally – privacy-by-design approaches. This intermediate conclusion from a legal point of view leaves open two questions: First, are these technologies actually able to protect the privacy of the research participant? And second, what are the deeper moral reasons and possible effects of ethical significance when such a new 'broad consent' regime is implemented? We will now focus on the second question.

## 5 Ethical Issues of Changing the 'Information Framework'

When assessing the problem of informed consent in a Big Data context, a historical perspective is helpful. The notion of informed consent has been put in the centre of bioethical considerations after one of the darkest episodes in the history of medical research – the horrific experiments carried out by doctors on concentration camp victims in Nazi Germany. In the Nuremberg trials of 1947, the requirement that "The voluntary consent of the human subject [to medical research] is absolutely essential" has been formulated for protecting the participant from harm. These requirements strongly influenced the Declaration of Helsinki, that later underwent several revisions, in particular related to the notion of informed consent (Carlson et al. 2004). Despite these changes, the 'moral core' of informed consent in the bioethical common-sense-understanding is protecting the individual from involuntarily incurred harm. From that perspective, the ethical question is, whether Big Data driven research backed by 'broad consent' could create additional harm for the subject – i.e. harm not directly related to the research intervention itself (e.g., the risks of some imaging techniques, which certainly are part in the information procedure when obtaining informed consent), but to long-term outcomes of the research. As the current legal discussion described in Sect. 5 demonstrates: the focus of the discussion is almost exclusively on privacy breaches as the main harm that could result. For example, one wants to avoid that the genetic data of a person with Huntington's disease made available for research can lead to a re-identification of this person, thereby harming this (still healthy) person in her social setting, e.g., by provoking a dismissal from her job.

We certainly do not dispute that this kind of harm is of relevance in Big Data driven research – and the main ethical question here is whether the technological solutions for preventing such harm actually will do their job. This aspect will be further discussed in Sect. 7. But we suggest two further issues that need ethical consideration: First, the necessity of broad consent for Big Data driven research may pose additional problems that have harm-implications. Second, broad consent is associated with other (positive) ethical values than harm-prevention that may help to make Big Data research more ethical. We will now discuss these two issues in more detail.

The first issue relates to the point that providing informed consent requires informing the patient on the intervention that will generate the data. On the one hand this concerns information on the direct risks and consequences of the intervention

itself. For example, in case of a MRI scan of the brain, issues like technical risks (e.g., implants) or incidental findings have to be discussed. This part of the informed consent procedure is not affected by a subsequent use of the data in a Big Data context. On the other hand, the patient has to be informed at least to some degree on the potential use of the data. If the informed consent is broad, this degree will be quite unspecific, but still needs some framing. Patients are unlikely to accept an explicit formulation like 'You agree that your personal data will be used for any kind of application'. Thus, a framing in two respects will be necessary: First, one has to induce trust in the patient that harm through privacy breaches will effectively be prevented. Second, some factual framing will be needed. Probably the broadest kind of factual framing is that the data will be used for *research* purposes (and, e.g., not sent to a wellness company such that they can tailor new commercial offers for patients with similar diseases). More likely is, however, a (at least implicit) framing that the data will be used for research related to the medical condition of the patient. But why is such a framing necessary?

The reason for this is – as we suggest – that information frameworks play a decisive role for giving moral meaning to the world we live in. This insight can be partly attributed to the idea of *spheres of justice*, introduced in 1983 by the philosopher Michael Walzer, which proposes that societies consist of different social spheres (e.g., medical, political, market, family and educational) each defined by a different type of good that is central to that particular sphere. These different types of goods (e.g., medical treatment in the medical sphere, political responsibility and public office in the political sphere) and the meaning and significance they have in each of these spheres, have their own associated criteria, principles and mechanisms concerning their distribution and allocation. In order to prevent mixing up distributional criteria and goods from different spheres (and prevent, e.g., allocating seats in parliament on the basis of financial assets, family relationships or health condition, or making one's ranking on a waiting list in health care dependent on family relationships or college degrees) these spheres have to be kept separated. This idea implies amongst other things that the distribution of access to particular goods tracks the sphere's specific normative considerations (e.g., 'need' in the medical sphere, 'democratic election' in the political sphere). Goods have to be distributed along the mechanisms of the corresponding sphere and goods from different spheres ought not to influence each other in terms of distribution. Put differently, this means that the exchange of goods between spheres has to be "blocked"; Walzer talks about "blocked exchanges" and the "art of separation".

Walzer's work has been applied to the realm of information systems by Nissenbaum (2004) and Van den Hoven (2008). Nissenbaum coined the term *contextual integrity* of social spheres, whereas spheres are defined through the expectations and behaviour of actors that differ per sphere. In order for contextual integrity and sphere separation to be achieved, the type of information that is revealed and the flows between different parties have to be appropriate for the context.

Within the broader privacy debate, the challenge of Big Data is that information produced within these spheres (health, politics, criminal justice, market) travels much faster and is more difficult to control than in the traditional offline world.

So we face a set of phenomena that threaten the integrity of social spheres and the cultural and social meanings expressed in them, including our values. Of course the boundaries between spheres are to a certain extent relative to time and culture, and not carved in stone forever, but it is important to note that every age, society and culture does in fact draw and treat these boundaries – construed as sets of constraints on the flow of information – as of high normative relevance.

Going back to our example, this also means that broad consent should respect these boundaries. The point is that providing broad consent for using data can transgress these boundaries in ways that generate indirect harm for the person who provides the data *even in cases when privacy is fully respected*. For example, researchers emerging from fields completely unrelated to the disease condition of the patient may use the (aggregated and anonymized) data to check for connections between health conditions and credit rating; resulting finally in a policy that prevents the patient in future to obtain certain bank credits. This would be considered a breach of a boundary between two social spheres with quite different moral regimes: the health sphere on the one hand and the economic sphere on the other hand. Other researchers may use the data in a way that finally results in a genetic test that allows testing foetuses – and in this way offer the option of abortion to the future parents. Such a development may be against core-values of the patient when she reads about this type of research in the newspapers, as she realises that her data may have played a role in this research. In this case, personal boundaries between acceptable and unacceptable applications of scientific research are breached. Yet other researchers may – based on research that includes the anonymised data of the patient – come to the conclusion that some sub-form of a neurological disorder (actually the condition from which the patient suffers) is associated with another disease that has a much stronger social stigma – and the patient is finally confronted with social exclusion resulting from the public dissemination of this reconceptualised disease space.

The underlying problem of these still hypothetical cases is that through broad consent, the consenting person risks that his data finally leads to research result that transgress important moral boundaries of this person or of society in general. The person contributes to a "new world" which he personally rejects. Thus, the question emerges how broad consent can be made compatible with respecting these boundaries.

Answering this question involves the insight that requiring consent is not merely an act to protect a person from unwanted harm – the classic understanding of informed consent. But it also involves requiring an explicit agreement to contribute to something that the person considers to be a valuable goal. Consenting is an act of autonomy that has a positive motive (e.g., compassion) and is backed up by some understanding of fairness (e.g. that the resulting research is not leading to unjustified discriminations). Understanding consent as such an active act entails the notion of responsibility in two ways: First, the consenting person trusts that the researcher will deal responsibly with this data – both with respect to preventing privacy breaches as well as with respect to the goal of the study. Second, the consenting person may to some degree be set in a position to control the usage of the data. Although it will probably be an exception that the person herself would like

to track the usage of her data, one may consider a model of "data stewardship", i.e. an institutional setting that (as a representative of the data provider) allows tracking data usage and regularly report on how the personal data of people has contributed to research. Both ensuring trust and responsibility will have to be "materialised" through technological solutions that function and that can be understood by both the users of Big Data technologies as well as those who provide the (Big) data. Whether these technologies are available is the topic of the next section.

## 6 Technological Ways of Securing Open Consent

A major technological problem related to this aim of enhancing trust and responsibility is that current anonymisation practice does not take the informational self-determination of the data subject into account. Since in most cases the data releaser is held legally responsible for the anonymisation (for example, this happens in official statistics), the releaser favours global anonymisation methods, where he can make all choices (methods, parameters, privacy and utility levels, etc.).

When asked to provide data and consent, the subjects must hope there will be a data protector who will adequately protect their privacy in case of release. Whereas this hope may be reasonable for data collected by the public health care system or more generally by (democratic) administrations, it may be less founded for private surveys (data collected by pharmaceutical companies or by any other private company). Indeed, a lot of privately collected data sets end up in the hands of data brokers (U.S. Federal Trade Commission 2014), who trade with them with little or no anonymisation. Hence, there is a fundamental mismatch between the kind of subject privacy (if any) offered by data releasers/protectors and privacy understood as informational self-determination: usually, the subject is not given control on how her data is protected.

To empower the data subject, Domingo-Ferrer and Muralidhar (2015) proposed a permutation-based paradigm of data anonymisation. They showed that any anonymisation method is functionally equivalent to permutation plus (perhaps) a small amount of noise. In a nutshell, if one compares the ranks of the values of each original and each anonymised attribute, one finds that the effect of any anonymisation method is to change the ranks to some extent, which can be viewed as a permutation (see Domingo-Ferrer and Muralidhar (2015) for more details and a running example). Based on this, they defined a new privacy model, called ($\boldsymbol{d}$, $\boldsymbol{v}$, $\boldsymbol{f}$)-permuted privacy that is verifiable by the subject. When given the anonymised data set by the data protector, each subject can check how much the values in her record have been permuted and whether this permutation is sufficiently protective.

Just allowing the subject to verify protection may not be enough or even worse than not allowing verification if the subject is left unsatisfied with the level of protection provided. An unsatisfied subject may refuse to answer and/or to give consent the next time the data collector approaches her. A more constructive alternative would be to allow the subject to take care of the anonymisation of

her own data record (*local anonymisation*, e.g. Song and Tingjian Ge 2014). In the context of the HBP, in some cases it may be viable for patient subjects to use their personal devices (e.g. smartphones) to conduct local anonymisation. For example, if a patient is being continuously monitored through sensors connected to her smartphone while at home, clearly all data being collected can be locally anonymised by her smartphone.

Beyond assuming a well-informed subject with some basic knowledge of the implications of anonymisation, a problem of local anonymisation is that the subject must anonymise her record without seeing the records of the other subjects. Hence, the subject cannot know whether the anonymisation she is applying will permute the values of her record enough with the values of the other subjects. To play it safe, each subject is likely to add a lot of noise to her values, which results in an anonymised data set with too poor utility.

In Soria-Comas and Domingo-Ferrer (2015) *collaborative anonymisation* has been proposed as a synthesis alternative that seeks to empower the subjects while preserving data set utility as in the case of centralised anonymisation for the same privacy level. The idea is that subjects generate the anonymised data set in a distributed and collaborative manner. Neither the data collector nor subjects gain more knowledge about the confidential information of a specific subject than disclosed by the anonymised data set.

Let us analyse the motivations of a rational subject to engage in collaborative anonymisation. Rationally, she will only contribute to form an anonymised data set if the benefits she obtains compensate her privacy loss:

- A subject without any interest in the research made possible by the data being collected is better off by declining to contribute (privacy prevails). Note, however, that subjects may have indirect interests, like expecting a potential benefit from the research conducted with the data (better healthcare, better life conditions, etc.) or simply satisfying a philanthropic inclination.
- A subject without privacy concerns can directly supply her data without any anonymisation requirements (potential benefit prevails).
- A subject who is interested in the research made possible by the data but has privacy concerns should prefer the collaborative approach to both the centralised and the local approaches because: (i) It outperforms centralised anonymisation by offering privacy with respect to the data collector; (ii) it outperforms local anonymisation because it yields less information (utility) loss and hence enables better research.

Collaborative anonymisation leverages the notion of co-utility (Domingo-Ferrer et al. 2015), which refers to protocols (interactions) designed in such a way that the best strategy for a rational selfish player to attain her goal is to help some other players to attain theirs. Co-utile protocols make mutual help self-enforcing. In anonymisation of individual data, the privacy protection obtained by a subject positively affects the protection that others get. In other words, when masking the identity of a subject within a group, none of the subjects in the group is interested in making any of the other subjects re-identifiable, because that makes her own data

more easily re-identifiable. In this sense, we can say collaborative anonymisation is co-utile. Specifically, Soria-Comas and Domingo-Ferrer (2015) give a co-utile protocol to achieve k-anonymity in a collaborative way. *k*-Anonymity is a privacy model in which each subject in indistinguishable within a group of *k* subjects when looking at the released data set.

While the above ***(d, v, f)***-permuted privacy model can allow a patient/subject to verify how well her data have been anonymised, and local/collaborative anonymisation can give the subject full control on the anonymisation process, privacy is not all a patient may need, as mentioned in Sect. 6. Being able to track the usage of her data is a complementary (and probably more ambitious) requirement. In fact, for some types of data used in HBP, anonymisation may be unfeasible because the data is inherently identifying and cannot be altered to make it less identifying (e.g. this is the case of genetic data or even human brain scans); for such data, all the patient could be promised by the researchers/collectors is to keep track of who accesses it and how it is used (the data stewardship mentioned in Sect. 6). Such tracking is addressed by the so-called provenance technologies. Provenance refers to the chain of successive custody (including sources and operations) of information (or even hardware equipment). The current practice of information provenance is rather rudimentary and still far from being dependable enough. The good side is that there are many sectors interested in improving provenance technologies: beyond healthcare research and HBP, the banking sector, the software industry and the cybersecurity sector are important fields where tracking information usage is very important. Hence, substantial research efforts are underway: technologies have been demonstrated for annotation in scientific computing, for provenance-aware data storage (automatically tracking accesses, downloads, etc.), for building tamper-resistant chains of custody, for pedigree management (tracking the source of data), etc. (See Chapter 9 of U.S. Homeland Security 2009 and references therein).

## 7 Conclusion

In conclusion, we suggest that a broader ethical focus would allow understanding the ethics of Big Data driven research not solely as an issue of upholding the privacy of the individual who consents to her data being used, but also as a matter of individuals that decide to contribute to a positive goal and thus would like to be put into a position such that they can trust that they are indeed making the world a better place. This requires generating an understanding on how Big Data research may affect the ontology upon which consent decisions are based (e.g., disease ontologies) as well as the underlying, morally significant boundaries. This also requires developing and integrating technologies in Big Data research that empowers the subject so that she really is in control of what happens to her data before it is released. This should enable her to give her data and her consent in conditions that are more compatible with informational self-determination.

## References


Carlson, Robert V., Kenneth M. Boyd, and David J. Webb. 2004. The revision of the declaration of Helsinki: Past, present and future. *British Journal of Clinical Pharmacology* 57(6): 695–713.

AQ2 Christen, Markus, Nikola Biller-Andorno, Berit Bringedal, Kevin Grimes, Julian Savulescu, and Henrik Walter. Ethical challenges of simulation-driven big neuroscience. *AJOB Neuroscience* (Under revision).

Cordasco, Kristina M. 2013. Obtaining informed consent from patients: Brief update review. In: Making health care safer II: An updated critical analysis of the evidence for patient safety practices. Rockville: Agency for Healthcare Research and Quality (US) (Evidence Reports/Technology Assessments, No. 211) Chapter 39. Available from: http://www.ncbi.nlm.nih.gov/books/NBK133402/

AQ3 D'Abramo, Flavio. 2015. Biobank research, informed consent and society. Towards a new alliance? *Journal of Epidemiology and Community Health* 69(11): 1125–1128. doi:10.1136/jech-2014-205215.

Domingo-Ferrer, Josep, and Krishnamurty Muralidhar. 2015. New directions in anonymization: permutation paradigm, verifiability by subjects and intruders, transparency to users. Technical report. Available at: http://arxiv.org/abs/1501.04186

Domingo-Ferrer, Josep, Jordi Soria-Comas, and Oana Ciobotaru. 2015. Co-utility: Self-enforcing protocols without coordination mechanisms. Proceedings of the 2015 international conference on Industrial Engineering-IEOM 2015, Dubai, 3–5 Mar, 1–7.

Ferguson, Adam R., Jessica L. Nielson, Melissa H. Cragin, Anita E. Bandrowski, and Maryann E. Martone. 2014. Big data from small data: Data-sharing in the 'long tail' of neuroscience. *Nature Neuroscience* 17(11): 1442–1447.

Frackowiak, Richard, and Henry Markram. 2015. The future of human cerebral cartography: A novel approach. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 370(1668) pii: 20140171. doi:10.1098/rstb.2014.0171.

Grillner, Sten. 2014. Megascience efforts and the brain. *Neuron* 82(6): 1209–1211.

HBP Report. 2012. The Human Brain Project. A report to the European Commission. Available at: https://www.humanbrainproject.eu/documents/10180/17648/TheHBPReport_LR.pdf/18e5747e-10af-4bec-9806-d03aead57655. Accessed 4 Sept 2015.

Hofmann, Bjørn. 2009. Broadening consent – And diluting ethics? *Journal of Medical Ethics* 35(2): 125–129.

Insel, Thomas. 2013. Transforming diagnosis. Available at: http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml. Accessed 3 Sept 2015.

Lunshof, Jeantine E., Ruth Chadwick, Daniel B. Vorhaus, and George M. Church. 2008. From genetic privacy to open consent. *Nature Reviews Genetics* 9(5): 406–411.

Monroe, Scott M., F. Samantha, and S.F. Anderson. 2015. Depression: The shroud of heterogeneity. *Current Directions in Psychological Science* 24(3): 227–231.

Morris, Sarah E., and Bruce N. Cuthbert. 2012. Research domain criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience* 14: 29–37.

Nissenbaum, Helen. 2004. Privacy as contextual integrity. *Washington Law Review* 79: 119–157.

Owen, Michael J. 2014. New approaches to psychiatric diagnostic classification. *Neuron* 84(3): 564–571.

Poldrack, Russell A., and Krzysztof J. Gorgolewski. 2014. Making big data open: Data sharing in neuroimaging. *Nature Neuroscience* 17(11): 1510–1517.

Serepkaite, Jurate, Zivile Valuckiene, and Eugenijus Gefenas. 2014. 'Mirroring' the ethics of biobanking: What should we learn from the analysis of consent documents. *Science and Engineering Ethics* 20(4): 1079–1093.

Shin, Jaehoon, Guo-li Ming, and Hongjun Song. 2014. Decoding neural transcriptomes and epigenomes via high-throughput sequencing. *Nature Neuroscience* 17(11): 1463–1475.

AQ4 Song, Chunyao, and Tingjian Ge. 2014. Aroma: A new data protection method with differential privacy and accurate query answering. Proceedings of the 23rd ACM international conference on Conference on Information and Knowledge Management-CIKM '14, 1569–1578. Shanghai: ACM.

Soria-Comas, Jordi, and Josep Domingo-Ferrer. 2015. Co-utile collaborative anonymization of microdata. In *Modeling decisions for artificial intelligence-MDAI 2015*, LNCS 9321. Springer, to appear.

AQ5 Stonnington, Cynthia M., Tan Geoffrey, Klöppel Stefan, Chu Carlton, Draganski Bogdan, Clifford R. Jack Jr, Chen Kewei, Ashburner John, and Frackowiak Richard SJ. 2008. Interpreting scan data acquired from multiple scanners: A study with Alzheimer's disease. *Neuroimage* 39(3): 1180–1185.

Thagard, Paul. 2008. Mental illness from the perspective of theoretical neuroscience. *Perspectives in Biology and Medicine* 51(3): 335–352.

Thompson, Paul M., et al. 2014. The ENIGMA consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior* 8(2): 153–182.

U.S. Federal Trade Commission. 2014. Data brokers: A call for transparency and accountability. Available at: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf. Accessed 6 Sept 2015.

U.S. Homeland Security Department. 2009. A roadmap for cybersecurity research. Available at: http://www.dhs.gov/sites/default/files/publications/CSD-DHS-Cybersecurity-Roadmap.pdf. Accessed 6 Sept 2015.

Van den Hoven, Jeroen. 2008. Information technology, privacy, and the protection of personal data. In *Information technology and moral philosophy*, ed. Jeroen Van den Hoven and John Weckert, 301–321. Cambridge: Cambridge University Press.

Walter, Henrik. 2013. The third wave of biological psychiatry. *Frontiers in Psychology* 4: 582.

AQ6 Walzer, Michael. 1982. *Spheres of justice: A defense of pluralism and equality*. New York: Basic Books.

Wang, Xiao-Jing, and John H. Krystal. 2008. Computational psychiatry. *Neuron* 84(3): 638–654.

AUTHOR QUERIES

AQ1. Please confirm the corresponding author and affiliation details.
AQ2. Please update complete details for Christen et al. (Under revision).
AQ3. Please confirm the inserted volume number, issue number and page range for D'Abramo (2015).
AQ4. Please confirm the inserted conference location for Song and Ge (2014).
AQ5. Please confirm the inserted year for Stonnington (2008).
AQ6. Please provide citation for Walzer (1982).