Published in final edited form as:
Annu Rev Genomics Hum Genet. 2012 ; 13: 415–431. doi:10.1146/annurev-genom-082410-101454.

The Tension Between Data Sharing and the Protection of Privacy in Genomics Research

Jane Kaye
HeLEX, Department of Public Health, University of Oxford, Oxford OX3 7LF, United Kingdom; jane.kaye@law.ox.ac.uk

Abstract

Next-generation sequencing and global data sharing challenge many of the governance mechanisms currently in place to protect the privacy of research participants. These challenges will make it more difficult to guarantee anonymity for participants, provide information to satisfy the requirements of informed consent, and ensure complete withdrawal from research when requested. To move forward, we need to improve the current governance systems for research so that they are responsive to individual privacy concerns but can also be effective at a global level. We need to develop a system of e-governance that can complement existing governance systems but that places greater reliance on the use of technology to ensure compliance with ethical and legal requirements. These new governance structures must be able to address the concerns of research participants while at the same time ensuring effective data sharing that promotes public trust in genomics research.

Keywords

governance; research participants; e-governance; research ethics committees; global; public trust

INTRODUCTION

With the costs of sequencing technology falling rapidly, we are moving to a position where whole-genome scanning of individual DNA samples will start to become routine in medical research and clinical medicine. This is also a critical point in time for the building of infrastructure and the linkage of existing biobanks and bioclinical projects. These plans are starting to be operationalized to enable the sharing of data and samples in a systematic way on a large scale. However, the meta-level governance mechanisms that are needed to support this are still in development. The move to global data sharing has been facilitated by funding bodies on both sides of the Atlantic, which have supported large international collaborative projects and developed open access policies to encourage wide-scale data sharing.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

In combination, these trends challenge some of the basic principles of protection of research participants and the current governance frameworks for research. One of the key challenges is determining how to protect the privacy of participants while enabling the sharing of data and samples through global research networks. To provide some understanding of the concerns raised by data sharing, this review outlines the issues involved in privacy protection as well as the current trends that have transformed genomics research practice and facilitated data sharing. It describes how data sharing tests current ethical principles and oversight mechanisms for medical research. In conclusion, it discusses ways forward and some of the new initiatives being developed to facilitate data sharing and enable sustainable genomics research.

THE NATURE OF PRIVACY

The protection of individual privacy is enshrined in legal instruments of all liberal democracies and is a benchmark of civil society. Although privacy is not an absolute right, interference must be justified in the public interest and/or according to law. An example of how the courts in the United Kingdom regard privacy is from Lord Justice Laws of the Court of Appeal:

> Subject to [certain] qualifications…an individual's personal autonomy makes him -- should make him -- master of all those facts about his own identity, such as his name, health, sexuality, ethnicity, his own image…and also of the "zone of interaction"…between himself and others. He is the presumed owner of these aspects of his own self; his control of them can only be loosened, abrogated, if the State shows an objective justification for doing so. (77a at §21)

Individual expectations of privacy are context specific, and so can vary depending upon the individual and the circumstances. Within research, the expectations and norms associated with different kinds of research can lead to variation in the practices that apply. Privacy consists of four interrelated dimensions, which come into play in different ways depending upon the context: physical privacy, informational privacy, decisional privacy, and proprietary privacy (40). In the case of genomics research, any or all of these dimensions may be activated depending upon the context.

Within genomics research, some of the privacy risks have been identified "as analysis efforts aimed at exposing individual research participants' information, including revealing disease status, predicted future likelihood or past presence of other traits, or attempts to link another DNA result with a participant, for example, to determine presence or absence in a research cohort, ancestry, and relatedness (e.g., paternity/nonpaternity)" (30). To safeguard against such harms, protections must be established to prevent discrimination against participants and ensure that their medical and personal information is not disclosed to third parties---such as their family or community members, employers, or insurance companies---against their wishes (19). This is because the character of DNA means that sequence information has implications for other biologically related family members, and "the fact that children carry half the genetic information of their parents implies that a decision to reveal one's genetic information today has repercussions for generations to come" (30). These concerns have led to considerable debate within the genomics community as to how

best to protect participants' and their biological relatives' privacy while still allowing research to proceed.

TRENDS WITHIN SCIENCE

Over the past 10 years, there have been significant changes in the way that genomics research is carried out that have implications for privacy protection. These changes are part of a longer-term evolution in scientific practice that has been occurring over a number of decades. Genomics research is now increasingly dependent upon the sharing of data and samples through global collaborative research networks. This widespread data sharing and the building of global research networks are possible only because of technological advances, considerable investment in infrastructure and international consortia, and the implementation of open access policies by funding bodies. Achieving research goals and priorities at an international level would not have occurred at the same scale and speed without the advances in bioinformatics and computing technology, which in turn have led to changes in scientific practice and the way that research is carried out. The relatively recent introduction of next-generation whole-genome sequencing technology adds another layer of complexity to this situation.

NEW MODELS

The way that genomics research is carried out today, based on the principles of open access and sharing, has its origins in the Human Genome Project, which commenced in 1990 and was completed in 2001. This project marked the beginning of a new way of doing genomics research, as it relied on the collaboration of many scientists, institutions, and funders from around the world (13). It marked a transition from a "cottage industry" approach based on bespoke laboratories to high-throughput sequencing involving teams of multidisciplinary experts (73). The possibility of the human genome being patented by a private company, Celera, helped to confirm and develop the principle that such knowledge should be freely available to all (26a, 48).

Using the Human Genome Project approach, a number of data-generating projects have been initiated through joint efforts by national funders, including the Encyclopedia of DNA Elements (ENCODE; http://www.genome.gov/10005107), the Human Epigenome Project (http://www.epigenome.org), the International HapMap Project (http://www.hapmap.org), and, more recently, the 1000 Genomes Project (http://www.1000genomes.org). These have provided unrestricted access to sequence reference libraries via the Internet. Such resources allow new types of scientific questions to be asked, as "vast numbers of polymorphisms can be studied simultaneously, rather than focusing attention on a small number of genes," and "very many more individuals can be genotyped in a single study" (15). Such data sets have been presented as the "drivers of progress in biomedical research," and therefore open access policies have stated that they should be "made immediately available for free and unrestricted use by the scientific community to engage in the full range of opportunities for creative science" (48). The role of such projects in advancing science has been seen as testimony to the success of open access policies. However, there have been concerns

Europe PMC Funders Author Manuscripts

expressed about the privacy risks that placing individuals' sequence data on the Web may have for research participants (77a at §21).

In addition to sequence reference libraries, repositories have been established to centrally organize the storage and sharing of data derived from genome-wide association studies (GWAS) (16). These studies compare the genomes of healthy controls with those of people who exhibit a disease or a specific trait in order to identify the genetic variants associated with that disease or trait (36). To obtain the sample sizes needed to do this, researchers have developed new models of collaboration and data sharing. Examples of these projects are the Wellcome Trust Case Control Consortium (73c) in the United Kingdom, the European Genome-Phenome Archive (http://www.ebi.ac.uk/ega), and the National Institutes of Health's Database of Genotypes and Phenotypes (dbGaP; http://www.ncbi.nlm.nih.gov/gap) in the United States. The aim of these platforms is to maximize the public benefits that can be realized from data sharing (17), and new methodologies and approaches have had to be developed to handle the vast amounts of data created (61). Data must also be deposited within a specific period of time and must meet certain standards of quality. This requirement also includes statements about the nature of the study, which are intended to standardize models for performing studies and reporting results (41). Such pooling of data ensures that the validity of results can be confirmed in replication studies before they are relied upon, providing a further reason for sharing data (29). Unlike access to sequence reference libraries, access to these data sets is provided through a managed access system that requires researchers to establish their credentials and then be approved by a data access committee. There is concern that this managed access model is not as effective for sharing data as Web-based sequence reference data sets, which receive many more hits.

INFRASTRUCTURE DEVELOPMENT

In the field of biobanking there has been considerable investment in new population biobanks and cohort studies. One of the research rationales behind the establishment of these resources is to develop information that can help elucidate the fine associations between the genotypes and phenotypes that influence the etiology of common diseases. The need for diverse, well-characterized, large sample groups, both for investigative purposes and for use as controls (7), has led to an increased emphasis on cooperation at both the national and international levels (23). A number of groups have been funded to develop the tools to standardize and harmonize collection and management procedures in order to facilitate wide-scale data and sample sharing, including the Public Population Project in Genomics (P^3G; http://www.p3gconsortium.org) and the International Society for Biological and Environmental Repositories (ISBER; http://www.isber.org). Over the past two years there has also been investment in infrastructure to facilitate the linkage and greater use of existing clinical collections of samples, including the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) (72a) and Biobank Standardisation and Harmonisation for Research Excellence (BioSHaRE; http://www.p3g.org/bioshare) projects in Europe and the Electronic Medical Records and Genomics (eMERGE) Network in the United States (49). The aim of this investment is to provide resources that networks of interdisciplinary teams and consortia located around the globe can use to answer a number of research questions. Large international consortia within Europe and the United States have been funded, as have

Europe PMC Funders Author Manuscripts

international collaborations on a grand scale such as the International Cancer Genome Consortium (http://www.icgc.org).

This new emphasis on the linkage of existing biobanks through a common infrastructure requires macro-level, international governance structures and processes to allow the secondary research use of existing information and samples. This raises significant questions about the oversight of global research activity and the best ways to safeguard researcher access to information while protecting the privacy of individuals. Many of the secondary research purposes such infrastructure will make possible were not anticipated at the time when consent was obtained for the collection of the data or samples. The arguments for using existing research collections for secondary research purposes are twofold: First, recruitment to large studies is expensive and time-consuming, and second, larger sample sizes are likely to accelerate research results. Reusing, integrating, and comparing collections will result in an efficient and effective use of funding. However, appropriate governance systems and procedures to link and network new and existing collections at this macro level are still being developed.

OPEN ACCESS POLICIES

The arguments for the efficient use of resources funded by the public purse also underpin many of the recent open access policies that have been developed by the leading funders of genomics research within the United States, the United Kingdom, and Canada. These policies started with the sequence data [the Bermuda Principles in 1996 (26b) and the Fort Lauderdale Agreement in 2003 (73a)] and have now been applied to other forms of data [the Toronto International Data Release Workshop recommendations in 2009 (5), the Amsterdam Principles in 2008 (63), and the Wellcome Trust joint statement by funders of health research (73b)]. In addition, there are a number of policies for data sharing by the Organisation for Economic Co-operation and Development (OECD) (57). All of these policies have statements requiring the protection of individual privacy and in some cases the dignity of communities while at the same time encouraging wide-scale data sharing for public benefit.

Although these policies are still in their infancy, we are starting to see their impact on the planning, execution, and oversight of genomics research and on the way results are disseminated. The question now is how to share data rather than whether data should be shared at all (36). These policies have created a climate in which data sharing is becoming more the norm---not just for large sequencing projects but for many different types of studies. However, there is still evidence to suggest that researchers are reluctant to share data (60).

Open access principles have also come into conflict with privacy concerns. In 2008, aggregate genetic data placed on the Web by researchers for GWAS use had to be withdrawn once it was realized that individual participants could be distinguished from the openly shared data (26). These problems of identifiability and disclosure risks are likely to become more frequent as increasingly diverse sources of data are linked (24, 57a).

Europe PMC Funders Author Manuscripts

TECHNOLOGICAL ADVANCES

Advances in information technology and genome sequencing technology have enabled significant changes in the ways that science is carried out and have provided a means to share data on a wide scale. Digital information can be deposited on the Web or in a cloud and then shared with colleagues and other third parties. Once DNA is sequenced from a sample and transformed into a digital form of AGTC base pairs, it can be used for many different purposes and analyzed by different researchers using different methodologies and approaches. The current challenges include issues of data storage, the quality of sequencing data, and the accuracy of genome assembly (9) as well as how best to manage and interpret large data sets of sequence information (46). The advances in next-generation sequencing technology have resulted in far richer and more detailed sequence information at a lower cost. Whereas it is estimated that the Human Genome Project cost US$2.7 billion (53a), in 2009 the company Complete Genomics announced that it could sequence an individual genome for US$5,000 (1). It is anticipated that these costs will continue to fall and that sequencing will no longer be a bespoke activity but will become a routine part of clinical care. As sequencing becomes cheaper, the use of whole-genome sequencing will become the norm in medical research and bring with it a number of new issues.

The challenges that this presents led Mardis & Lunshof (47) to write that “the established framework of ethical, legal and social issues (ELSI) in genomics has been shaken to its foundations by something as simple as the emergence of personal genomes.” Tabor et al. (68) note that

> whereas conventional technological approaches might generate data on hundreds of thousands, or even millions of polymorphisms, the overwhelming majority of these variants are located in noncoding regions and likely not of functional significance themselves. In contrast, both ES [exome sequencing] and WGS [whole-genome sequencing] provide information on virtually all functional, protein-coding variants in the genome for each individual participant. This includes most variants known to influence risk of human diseases and traits.

These technologies increase the possibility of identifying serious treatable conditions and generating other incidental findings (77) and have created a heated debate as to whether there is an obligation to report research findings to participants and, if so, how this should be done. This reporting raises a number of ethical issues, such as how to develop management pathways and privacy safeguards, and questions of whether secondary and tertiary researchers also have an obligation to report back findings. New models of reciprocal participation in research that also provide individual-level information have been developed by companies such as 23andMe, where participants are treated as customers rather than “health information altruists” and are given access to genomic information (39). Further research is needed to establish whether such new models of participation are truly reciprocal, whether they could have wider application, and how management pathways for feedback could be developed (72).

Europe PMC Funders Author Manuscripts

THE EFFECT ON SCIENTIFIC PRACTICE

In combination, these trends have had a marked effect on the scientific agenda and the conduct of genomics research. Research is now carried out by interdisciplinary teams of specialists brought together in flexible research collaborations that can process and analyze large amounts of information and large numbers of samples. The collection of information and samples is still carried out by individual researchers, but the model of large interdisciplinary collaborations means that existing collections can be brought together and reused for new purposes. This is possible only because technological advances make it easy to share and distribute data through global networks. Open access policies are changing the way that data are generated and distributed and are enabling new ways to mine data. Increasingly, there is now a distinction between data generators and data users. Data sets are no longer the sole creation or in the control of one individual or institution, but must be made available to the whole research community. For GWAS, this has been achieved through a new managed access model with formal application processes and access determined by data access committees in consultation with collectors, rather than decided by the principal investigator alone.

These changes in the way that science is conducted mean that the “secondary users of the data are far removed from the researchers who carried out the collection of the samples and data, as well as from the research participants” (36). Data sharing has the potential to sever the ties between the researcher responsible for participant enrollment and the individual participants in an original study. The onward sharing of data raises questions about who is accountable not only to research ethics committees approving new research but also to the research participants for the secondary uses of data in other studies. These advances also challenge our legal and ethical frameworks as data-sharing practices give a new twist to the old questions of informed consent, protection of privacy, and governance of medical research. These trends have had a significant effect on the principles that underpin research and the basis of research participation.

PROTECTIONS FOR RESEARCH PARTICIPANTS

The main purpose of the current research governance system is to protect participants’ interests and ensure that research is carried out ethically. It does not have a mandate to consider the broader ethical issues associated with data sharing, such as equitable access to biorepositories for researchers. A number of procedures, practices, and oversight bodies have been established that are designed to protect research participants. Common to all jurisdictions are the requirements that consent must be obtained before the research commences (although there are a number of exceptions to this basic principle), that an individual has a right of withdrawal, and that there must be some review of the research proposal by an appropriate committee, such as an institutional review board (in the United States) or a research ethics committee (within Europe and elsewhere). These protections derive from the Nuremberg Trial principles (69, pp. 181--82), which were intended to protect individual research participants from physical harm rather than informational harm. They were not designed for use in global networks where information and samples flow through international research collaborations; rather, they were developed for a time when

research was oriented toward one principal investigator, leading one research project, based within one country, located at one point in time---the “one researcher, one project, one jurisdiction” model (33). As a result, they are focused at the beginning of the research process, and oversight is largely reliant on expert committees.

The nature of whole-genome sequence data and the potential for global data sharing also brings into question the social contract that underpins research participation and the governance mechanisms that have been built around it. The basis for medical research participation has traditionally been an appeal to altruism (22), solidarity (38), and/or the gift model (8, 71), depending upon the nature of the study (39). The degree of participant involvement in the research process has varied depending upon the type of study---for example, whether it is clinically based with direct patient contact or epidemiological and concerned with population trends. In some cases, participants have had a passive role as providers of samples, information, and interesting case examples of disease. In other types of research, such as research on HIV/AIDS, participants have been more actively involved in defining the research agenda (31). In all cases, good practice has required that in return for being altruistic, participants’ personal, identifiable information should remain confidential and, if possible, be rendered anonymous. This has also been the basis for not having to obtain explicit consent for new research in cases where this may be difficult and when the risks to individuals are perceived to be low. Research procedures and practices have been established on the basis of this implicit social contract. The traditional workhorses of medical research governance---informed consent, withdrawal, anonymization, and oversight mechanisms---are tested by the new developments in genomics research practice caused by wide-scale data and sample sharing.

INFORMED CONSENT

Informed consent has been used to respect individuals and to enable research participants to exercise their autonomy in medical research and make decisions about privacy risks. The requirements for informed consent have been enshrined in a number of ethical documents, one example being the Declaration of Helsinki (78). With wide-scale data sharing, it is impossible to fulfill the conditions of traditional informed consent as outlined in many ethical and legal documents (6). Participants cannot be informed of all future uses of their information and samples over many years at the time of collection, nor can they be given an assessment of all the potential privacy risks of participation in the research (3). Broad consent has become a practical solution to this problem for biobanks, but this is still contentious within the bioethics literature (11) as research participants are giving a broad consent at the beginning of the research process for the use of their sequence and data for many years. There is some doubt as to whether this enables individuals to fully exercise their autonomy, as they cannot choose whether to be involved in specific research projects using different biorepositories, determine what kind of research they participate in, or properly assess the privacy risks of involvement (6). The focus on individuals and informed consent can also eclipse legitimate family and group privacy concerns, which may differ from those of individuals.

At the present time, consent forms are the only means by which the wishes of research participants can be obtained and recorded. This occurs at the beginning of the research process, when potential participants are presented with an agreement that they cannot negotiate (but can refuse to sign) and that, in the case of biobanks, has to hold for a considerable amount of time. Another limitation of the one-off informed consent form is that researchers must anticipate all eventualities to make the consent future-proof and avoid costly and time-consuming recontact processes. This means that if data-sharing plans are described---which quite often is not the case---it is usually done in very broad terms (3). This raises questions as to how informed participants actually are (59) and whether they are really in a position to assess the privacy risks of research involvement. Currently, efficient and cost-effective mechanisms by which to go back to individuals for new consent for secondary research are not commonplace. Effectively, broad consent is “consent for governance” by others, as judgments about appropriate uses of data and samples often fall to researchers, advisory boards, or research ethics committees, who must make decisions on behalf of research participants (32). In response to these shortcomings, other models have been proposed such as tiered consent (21, 76), authorization instead of informed consent (2), and “open consent” (44). New forms of governance models, such as patient interfaces that give individuals greater control over their information (J. Kaye, E. Whitely, S. Creese, D. Lund & K. Hughes, manuscript in preparation), are in development, and “adaptive governance” mechanisms that give voice to group concerns rather than just those of individuals have also been proposed to address some of the deficiencies of the individual consent model (56, 75).

WITHDRAWAL

The other foundational principle of medical research ethics is the right of withdrawal, which is the notion that a research participant can discontinue his or her participation in research at any time (16a). This also applies to the data and samples that an individual may have given consent to use in research, and is one of the key ways by which an individual can enact decisional privacy. However, in the case of international data sharing this is extremely difficult, if not impossible, to achieve when data and samples are shared widely. Computer data sets containing personal information must be continually archived, and it is difficult to claw back minute segments of sequence spread over a global network when they are used in multiple research projects (80). The various deidentification and aggregation methods that have been put in place to protect privacy may also make it difficult to trace and remove individual derived data (10, 54). As we move increasingly toward networks and infrastructure research, it becomes impossible to withdraw completely from research---instead, it is only possible to prohibit the entry of new information and samples into the system. Therefore, the mechanisms that have been put in place to protect the privacy of data may actually make it very difficult to allow participants to withdraw and be forgotten.

In addition to the practical difficulties, there are also economic and public-good arguments for disallowing absolute withdrawal. Arguments have been made in the biobanking field that complete withdrawal could lead to the wastage of resources invested in biorepositories and that, by withdrawing, “a person not only withdraws from one project, but from an indefinite

number of future projects, including possibly some (many) that the person might want to participate in" (25).

PROTECTING PRIVACY

Part of the social contract in research is keeping information confidential as a way of respecting individual privacy. Best practice has involved the use of information technology mechanisms such as firewalls, encryption, passwords, and security compliance as well as the restriction of access to raw data that could directly identify an individual. Current research practice requires the deidentification of samples through the removal of personal identifiers when they are shared with other researchers. However, the uniquely identifiable nature of human genetic information means that with increased data sharing, it is very difficult to guarantee absolute confidentiality for research participants or their genetically related family members (51), especially if this involves only the removal of personal identifiers. Exome sequencing reveals rare alleles, including those that are of clinical or personal utility, which increases the risk of identity disclosure as well as breaches of confidentiality or privacy (68). In 2008, Homer et al. (26) demonstrated that it is possible to reidentify individuals who had been genotyped and even those in pooled mixtures of DNA, provided a reference sample is available. Once a person is reidentified, there is the potential for further personal information to be revealed about the formerly anonymous source. The standard protections, such as encryption, that have been used to anonymize or deidentify sequence information are challenged with new sequencing technology that produces richer and more detailed information on individuals (20).

A key issue in current research practice for the sharing of genomic data is whether an individual can be "distinguished" from sequence data or whether it is possible to actually "identify" who the individual is. Malin et al. (45) have argued that "genomic sequence data, for instance, and possibly other laboratory and molecular expression data, are often highly distinguishing" but that this is "insufficient to claim that the corresponding individual's privacy will actually be compromised"; to breach privacy, as these sequence data must be matched with data of a named individual, there must be a mechanism to relate the deidentified and identified resources. This assumes that the individual's privacy interest is extinguished if information and data are rendered nonidentifiable or anonymous. However, individuals may still have a concern about how their data might be used in research, even if the data is rendered anonymous. The difficulty here is developing a workable model that allows individuals to control the use of their information but also enables research to proceed.

Current research practice rests on the premise that being able to identify an individual would constitute a breach of privacy because of the potential harms that could result, whereas just being able to distinguish someone from a number of sequences does not. As Tabor et al. (68) note, "inferring the identity of a study participant could facilitate linking them to other genetic or phenotypic data that might be stigmatizing or discriminatory, regardless of whether or not they have a disease or carry a disease-causing variant." Once sequence information is linked to other data sets that contain identifiable information, the potential for privacy breaches increases. The difficulty is that this can be contrary to the research

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

endeavor, which can require linkage for innovation. There is the danger that privacy protections may become so elaborate that they start to limit what can be done with the data in research terms (2a).

The increasing availability of publicly available data sets on the Internet and sequence data outside the controlled cocoon of research also raises a number of concerns about potential privacy risks. Whereas genomic information used to be obtainable only through research sequencing laboratories, people can now obtain access to their own genome through direct-to-consumer companies (43) as well as ancestor-tracing companies. Examples of the use of publicly available information to identify individuals are becoming more common, such as the case of the boy who tracked down his sperm donor father using an ancestor-tracing company (53) and Gitschier's (18) use of publicly available data sets on the Mormon population to identify individuals. This is possible because relatively small numbers of single-nucleotide polymorphisms (SNPs) can identify an individual. If someone has access to individual genetic data and performs matches to public SNP data, a small set of SNPs could lead to successful matching and identification of the individual (55). Therefore, the development of networks must include appropriate controls both to protect individual information when data are shared within the research community and to ensure that this information is protected from those outside. The traditional focus of privacy protection in research on consent and anonymization cannot address the concerns raised by data sharing and whole-genome sequences. Owing to the extensive nature of information available on the Internet, privacy risks must be assessed in this broader context, not only within the narrow confines of one project or the activities of the research community in isolation (24).

OVERSIGHT OF RESEARCH

A number of interrelated mechanisms are involved in the governance of research, which can vary depending upon the jurisdiction (35). However, in most countries the key gatekeepers are the institutional review boards or research ethics committees. These bodies have the authority to determine whether research will proceed, yet they have limited enforcement powers, and decision making can vary between regions and countries. In the current governance system, these committees, largely run by medical professional peers, are the bodies that are increasingly being asked to stand in the shoes of research participants as part of "consent for governance." The authority of these bodies is national, yet in the context of increasingly global research, such bodies adjudicate on the complex issues associated with international data sharing and privacy. National oversight bodies do not always have the authority, scope, or expertise to assess the privacy risks associated with global data sharing or to ensure compliance with their decisions.

For research using stored samples and/or data, this system is largely dependent upon researchers going back to oversight bodies for additional approvals for new research and to ensure that all ethical requirements are met. Research participants themselves are not always asked to make a decision on privacy risks associated with new research using stored samples and/or data, as such deliberations are usually made by researchers or by oversight bodies. When data are shared publicly on the Web, they become freely available, and no additional oversight is provided for new research using these shared data. This has often been regarded

as legitimate because the new research falls within the scope of the original consent or because there have been mechanisms that effectively render data deidentifiable or anonymous. As sequencing technology improves and data are shared more widely, whether this current form of governance is desirable and whether these conditions for approval of new research can be fulfilled are brought into question.

Concerns about the ethical sharing of GWAS data and the implications for privacy have led to the development of managed access mechanisms to protect sequence data generated through research. These require researchers to establish their credentials and sign a number of contracts that outline their obligations and the approved use of the data. Special data access committees have been established for GWAS projects in addition to the normal research ethics committees. Although these committees provide accountable access oversight for specific projects, they do not provide a simple solution for researchers who may want to access multiple data sets, as they must obtain a new approval from each project they wish to access. If at some point there were a proposal to integrate the different GWAS projects, then this policy may need rethinking. These bodies replicate the model of expert committees and the use of paper forms that characterize institutional review boards and research ethics committees.

These special data access committees develop a new tier of oversight in addition to research ethics committee approval, but there is some doubt as to whether they are the right model to enable effective supranational data sharing. For example, such bodies do not have effective mechanisms to monitor data use by secondary or tertiary researchers once information has been obtained from a managed access repository. Regarding cases where results are then reposted on the Web, Johnson et al. (30) opine, “Once results are posted publicly, they cannot be deemed safe even if posted results are eventually retracted, since backups may have been created…. [C]ontrolled access models have decreased the overall risk for results misuse across studies, but the examples of reposting uncovered illustrate that controlled access is not fool-proof.” They go on to say that there needs to be clearer guidance regarding appropriate disclosure of derived data when data are accessed through managed access repositories.

These factors in combination raise questions as to whether the current protection for the privacy of research participants is suitable in the case of genomics and data sharing. The scientific agenda requires the long-term participation and commitment of research participants. But, at the same time, it is difficult to guarantee anonymity for participants, provide information to satisfy the requirements of informed consent, and ensure complete withdrawal from research when requested. Research participants are being asked, on the basis of altruism and solidarity, to involve themselves in research projects that could potentially expose them and their families to privacy risks (74). Current mechanisms to protect privacy are increasingly tested by the comprehensiveness of the data that can be assembled from different sources for research purposes. At the same time, oversight systems are not equipped for global data sharing and the possible breaches of privacy that may arise, and yet these are the bodies that are being asked to stand in the shoes of research participants. However, it is widely accepted that there can be research benefits if sequence as well as phenotypic data are shared, as this is considered a good use of public resources.

Europe PMC Funders Author Manuscripts

For the future of genomics research, we need to develop a sustainable and coherent governance framework that addresses these concerns.

WAYS FORWARD

To move forward, we need to improve the current governance systems for research so that they are responsive to individual privacy concerns but can also be effective at a global level. These systems must be anchored and informed by individual and local contexts but also able to be enacted at a global level. The current system of governance based on paper forms and expert committees is derived from a "one researcher, one project, one jurisdiction" model that is not effective for global networks. We need to move to a system of e-governance that can complement these existing governance systems, with a greater reliance on the use of technology to ensure compliance with ethical and legal requirements. This requires the development of e-governance systems that are designed according to ethical, legal, and social norms---or "ELSI by design"---to ensure that participants' privacy is protected and public trust is maintained. Sustainability can be achieved only by building partnerships between all stakeholders in research, developing appropriate global governance structures, and enhancing translational research approaches.

BUILDING PARTNERSHIPS WITH RESEARCH PARTICIPANTS

Sustainable data sharing requires long-term commitment and wide-scale support from the public to maintain the high participation levels necessary for infrastructure initiatives as well as to guarantee the heavy investment of resources by publicly funded bodies. The recent Havasupai case (14) and the HeLa case (66) demonstrate that a lack of transparency and respect for individual wishes can have a detrimental effect on trust (28); to ensure the sustainability of genomics research, we need to move to a situation where participants, rather than being passive providers of information, have the opportunity to become more active partners in the research process. This will require a change in practice and attitude for some and the development of new forms of governance to enable such participation. These new governance structures must address the concerns of participants while at the same time ensuring effective data sharing that allows the translation of new knowledge into the clinic and promotes public trust in genomics research.

Recent research on the views of data-sharing participants from the United States suggests that plans for data sharing must be clearly explained to participants and that the main concern of participants is that they wish to be asked for their consent for involvement in data sharing. One study found that "90% of survey respondents, all of whom had consented to sharing with dbGaP, reported that it was important that researchers had asked for their consent"; alternatives to consent, or informing people after data had been deposited for sharing, were viewed as unacceptable by many respondents (42). This was confirmed in another study that found that participants would feel deceived or angry if they found out that their data were shared without their knowledge or consent---a finding that "is noteworthy because it demonstrates the high value research participants place simultaneously on the benefits of health research and on personal autonomy" (52). Research with the eMERGE participants, however, "suggests that, although most are willing to share their data, there is a

strong desire for some data-use limitations" (50). As noted by Tabor et al. (68), "these studies suggest that in any genetic studies that require broad genetic data sharing in dbGaP that researchers should consider the trust relationship with participants and mechanisms for providing transparency about how genetic data are being shared, and for what purposes they are being used."

As a way of addressing these concerns, a number of participant-centric interfaces (PCIs) have been developed that place patients and research participants at the center of decision making by enabling them to give consent for the use of their samples and personal information for research over time and to be active partners in the research process (34). Examples of PCIs in research are 23andWe (United States; http://www.23andme.com/research), Cooperative Health Research in South Tyrol (CHRIS) (Italy; http://www.eurac.edu/en/research/institutes/geneticmedicine/chrisstudy) CuraRata/String of Pearls Initiative (Netherlands; http://www.curarata.nl/uk/3/patients), Ensuring Consent and Revocation (EnCoRe)/Oxford Radcliffe Biobank (United Kingdom; http://www.encore-project.info), Genomera (United States; http://www.genomera.com/about), Genomes Unzipped (United Kingdom; http://www.genomesunzipped.org), Indivo personally controlled health records (United States; http://www.indivohealth.org/research), PatientsLikeMe (United States; http://www.patientslikeme.com), Private Access (United States; http://www.privateaccess.info), and TuAnalyze (United States; http://www.tudiabetes.org/forum/topics/tuanalyze-is-here).

An example of a PCI approach to consent that supports individual privacy preferences is the EnCoRe IT interface developed for the Oxford Radcliffe Biobank in the United Kingdom (J. Kaye, E. Whitely, S. Creese, D. Lund & K. Hughes, manuscript in preparation). It has a "dynamic" rather than "informed" consent model, which enables consent to be obtained when needed, in real time, as part of a bidirectional, ongoing, interactive process between patients and researchers as well as other health care professionals. This interface enables individuals to change their mind and preferences over time, to have their choices revoked where appropriate, to track and audit any changes made, and to choose when and how they are contacted. The use of "sticky policies" enables individual preferences to be tracked as information and samples move through localities over time.

Using PCIs ensures that research is compliant with the legal requirements for data and privacy protection, as consent for the use of identifiable information is a general requirement in all jurisdictions. By acting with consent, researchers can code data out of respect for individuals, not because it is difficult to obtain their consent. This reduces the resources and time that have to be put into anonymization strategies that are not necessarily beneficial for research. In addition, such interfaces expedite the research process by allowing easy recontact with participants for involvement in further studies, cutting down on referrals to research ethics committees. This technology has the potential to encourage new forms of engagement with the public by fostering new kinds of partnerships with research participants. PCIs have been used to develop new ways of carrying out research (e.g., 23 and We, Patients Like Me, and TuAnalyze) and to provide the basis for personalized medicine. A PCI approach means that other bodies no longer stand in the shoes of participants and that the local context can influence how data are continually used by researchers. Through these

Europe PMC Funders Author Manuscripts

mechanisms, the importance of research priorities can be explained so that the balance between individual decision making and the research agenda can be more nuanced.

PCI approaches can also complement other governance mechanisms that support participant involvement, such as representation on key oversight bodies. Trinidad et al. (70) suggest that innovation is needed in the consent and notification procedures for data repositories and other data-sharing resources, but also propose that there need to be "transparent, accountable oversight processes that include community representation; and…opportunities for study participants to provide input on decisions concerning the stewardship of their data, e.g., dialogue between researchers and participants, ongoing community consultation, deliberative processes, or reconsenting a representative sample of participants." Adaptive governance models have been proposed for biobanking, as they put in place representative governance systems that can be responsive to changing conditions and allow for the consideration of community concerns by participant representatives (56).

IN CONCLUSION

Our current governance system for research does not allow participants to have any further control of their data once they have signed a consent form. Recent empirical studies focused on data sharing in research suggest that this is contrary to what research participants want. The development of IT interfaces for research participants has the potential to enable individuals to be more informed about the research uses of their data and to give consent for secondary uses to protect their privacy interests by exercising their autonomy and decision-making capacities. Such systems are designed to complement the existing governance structures of research ethics committees and the mechanisms used to safeguard the confidentiality of information. They must also be constructed as part of a broader system of new e-governance tools that incorporates biobank and researcher IDs as well as mechanisms to enable statistical analysis without compromising privacy, such as the Sage bioinformatics platform (http://www.sagebase.org). Moving to such participant-centric systems will enable individuals to know how their data are used and enhance public trust and knowledge of the research process. It could also increase the transparency and accountability of the research governance system as data sharing and the use of next-generation sequencing technology become more widespread. The choice is whether to continue to invest resources in attempts to anonymize information---which is impossible and so will always carry a risk of privacy breaches---or to consider new ways of engaging with research participants that could include e-governance mechanisms. To do so respects the dignity of participants and protects fundamental human rights, and is also a hallmark of civil society.

ACKNOWLEDGMENTS

J.K. is funded under Wellcome Trust Award 096599/2/11/Z. My thanks also go to Jim Duncan for editing the manuscript and to the anonymous reviewer for comments.

LITERATURE CITED

1. Aldhous P. Genome sequencing falls to $5000. New Sci. Feb 6.2009 http://www.newscientist.com/article/dn16552-genome-sequencing-falls-to-5000.html

2. Arnason V. Coding and consent: moral challenges of the database project in Iceland. Bioethics. 2004; 18:27–49. [PubMed: 15168697]

2a. Anderson, N.; Edwards, K. Building a chain of trust: using policy and practice to enhance trustworthy clinical data discovery and sharing; Proc. Workshop Gov. Technol. Inf. Policies; Austin, TX. New York: Assoc. Comput. Mach.. 2010; Dec 7. p. 15-20.

3. Beskow LM, Burke W, Merz JF, Barr PA, Terry S, et al. Informed consent for population-based research involving genetics. JAMA. 2001; 286:2315–21. [PubMed: 11710898]

5. Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, et al. Prepublication data sharing. Nature. 2009; 461:168–70. [PubMed: 19741685]

6. Boddington P, Curren L, Kaye J, Kanellopoulou N, Melham K, et al. Consent forms in genomics: the difference between law and practice. Eur. J. Health Law. 2011; 18:491–519. [PubMed: 22128519]

7. Burton PR, Hansell AL, Fortier I, Manolio TA. Size matters: Just how Big is BIG? Quantifying realistic sample size requirements for human genome epidemiology. Int. J. Epidemiol. 2009; 38:263–73. [PubMed: 18676414]

8. Busby, H. Blood donation for genetic research: what can we learn from donors' narratives?. In: Tutton, R.; Corrigan, O., editors. Genetic Databases: Socio-Ethical Issues in the Collection and Use of DNA. Routledge; London: 2004. p. 39-56.

9. Butler D. Human genome at ten: science after the sequence. Nature. 2010; 465:1000–1. [PubMed: 20577184]

10. Cambon-Thomsen A. The social and ethical issues of post-genomic biobanks. Nat. Rev. Genet. 2004; 5:866–73. [PubMed: 15520796]

11. Caulfield T, Kaye J. Broad consent in biobanking: reflections on seemingly insurmountable dilemmas. Med. Law Int. 2009; 10:85–100.

13. Colins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003; 300:286–90. [PubMed: 12690187]

14. Dalton R. When two tribes go to war. Nature. 2004; 430:500–2. [PubMed: 15282577]

15. Day N. Commentary: How small is small? Int. J. Epidemiol. 2009; 38:274–75. [PubMed: 18984600]

16. GAIN Collab. Res. Group. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. Nat. Genet. 2007; 39:1045–51. [PubMed: 17728769]

16a. Gertz R. Withdrawing from participating in a biobank---a comparative study. Eur. J. Health Law. 2008; 15:381–89. [PubMed: 19180978]

17. Gibbs R. Deeper into the genome. Nature. 2005; 437:1233–34. [PubMed: 16251930]

18. Gitschier J. Inferential genotyping of Y chromosomes in Latter-Day Saints founders and comparison to Utah samples in the HapMap project. Am. J. Hum. Genet. 2009; 84:251–58. [PubMed: 19215731]

19. Gitter DM. The challenges of achieving open-source sharing of biobank data. Biotechnol. Law Rep. 2011; 29:623–35.

20. Greenbaum D, Sboner A, Mu XJ, Gerstein M. Genomics and privacy: implications of the new reality of closed data for the field. PLoS Comput. Biol. 2011; 7:e1002278. [PubMed: 22144881]

21. Haga SB, Beskow LM. Ethical, legal, and social implications of biobanks for genetics research. Adv. Genet. 2008; 60:505–44. [PubMed: 18358331]

22. Hallowell N, Cooke S, Crawford G, Lucassen A, Parker M, Snowdon C. An investigation of patients' motivations for their participation in genetics-related research. J. Med. Ethics. 2010; 36:37–45. [PubMed: 20026692]

23. Hattersley AT, McCarthy MI. What makes a good genetic association study? Lancet. 2005; 366:1315–23. [PubMed: 16214603]

24. Heeney C, Hawkins N, de Vries J, Boddington P, Kaye J. Assessing the privacy risks of data sharing in genomics. Public Health Genomics. 2011; 14:17–25. [PubMed: 20339285]

25. Holm S. Withdrawing from research: a rethink in the context of research biobanks. Health Care Anal. 2011; 19:269–81. [PubMed: 21918887]

26. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008; 4:e1000167. [PubMed: 18769715]

26a. Human genomes, public and private. Nature. 2001; 409:745. [PubMed: 11236960]

26b. Hum. Genome Organ.. Summary of principles agreed at the first International Strategy Meeting on Human Genome Sequencing; Bermuda. 1996; Feb 25--28. http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1

28. Hudson KL. Genomics, health care, and society. N. Engl. J. Med. 2011; 365:1033–41. [PubMed: 21916641]

29. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. Int. J. Epidemiol. 2008; 37:120–32. [PubMed: 17898028]

30. Johnson AD, Leslie R, O'Donnell CJ. Temporal trends in results availability from genome-wide association studies. PLoS Genet. 2011; 7:e1002269. [PubMed: 21931563]

31. Kahn, JP.; Mastroianni, AC.; Sugarman, J., editors. Beyond Consent: Seeking Justice in Research. Oxford Univ. Press; Oxford, UK: 1998.

32. Kaye, J. Biobanking networks: What are the governance challenges?. In: Kaye, J.; Stranger, M., editors. Principles and Practice in Biobank Governance. Ashgate; Farnham, UK: 2009. p. 201-13.

33. Kaye J. From single biobanks to international networks: developing e-governance. Hum. Genet. 2011; 130:377–82. [PubMed: 21785980]

34. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, et al. User-centric initiatives in health and biomedical research. Nat. Rev. Genet. 2012 In press.

35. Kaye, J.; Gibbons, SMC.; Heeney, C.; Parker, M.; Smart, A. Governing Biobanks: Understanding the Interplay Between Law and Practice. Hart; Oxford, UK: 2012.

36. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. Data sharing in genomics---re-shaping scientific practice. Nat. Rev. Genet. 2009; 5:331–35. [PubMed: 19308065]

38. Knoppers BM, Joly Y. Our social genome? Trends Biotechnol. 2007; 25:284–88. [PubMed: 17467834]

39. Kohane IS, Altman RB. Health-information altruists---a potentially critical resource. N. Engl. J. Med. 2005; 353:2074–77. [PubMed: 16282184]

40. Laurie G, Mallia P, Frenkel DA, Krajewska A, Moniz H, et al. Managing access to biobanks: How can we reconcile individual privacy and public interests in genetic research? Med. Law Int. 2010; 10:315–37.

41. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med. 2009; 6:e22. [PubMed: 19192942]

42. Ludman EJ, Fullerton SM, Spangler L, Trinidad SB, Fujii MM, et al. Glad you asked: participants' opinions of re-consent for dbGap data submission. J. Empir. Res. Hum. Res. Ethics. 2010; 5:9–16. [PubMed: 20831417]

43. Lumley T, Rice K. Potential for revealing individual-level information in genome-wide association studies. JAMA. 2010; 303:659–60. [PubMed: 20159874]

44. Lunshof JE, Chadwick R, Vorhaus DB, Church GM. From genetic privacy to open consent. Nat. Rev. Genet. 2008; 9:406–11. [PubMed: 18379574]

45. Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J. Investig. Med. 2010; 58:11–18.

46. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011; 470:198–203. [PubMed: 21307932]

47. Mardis ER, Lunshof JE. A focus on personal genomics. Pers. Med. 2009; 6:603–6.

48. Marris E. Free genome databases finally defeat Celera. Nature. 2005; 435:6. [PubMed: 15874985]

49. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med. Genet. 2011; 4:13.

50. McGuire AL, Basford M, Dressler LG, Fullerton SM, Koenig BA, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. Genome Res. 2011; 21:1001–7. [PubMed: 21632745]
51. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. Nat. Rev. Genet. 2008; 9:152–56. [PubMed: 18087293]
52. McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: research participants' perspectives. Genet. Med. 2008; 10:46–53. [PubMed: 18197056]
53. Motluk A. Anonymous sperm donor traced on internet. New Sci. Nov 3.2005 :6. [PubMed: 16363078]
53a. Natl. Hum. Genome Res. Inst.. US Dep. Energy Off. Sci.. The Human Genome Project completion: frequently asked questions. 2010. http://www.genome.gov/11006943
54. Natl. Inst. Health (NIH). Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS). NIH; Bethesda, MD: 2007. Not. NOT-OD-07-088
55. Nyholt DR, Yu CE, Visscher PM. On Jim Watson's APOE status: genetic information is hard to hide. Eur. J. Hum. Genet. 2009; 17:147–49. [PubMed: 18941475]
56. O'Doherty KC, Burgess MM, Edwards K, Gallagher RP, Hawkins AK, et al. From consent to institutions: designing adaptive governance for genomic biobanks. Soc. Sci. Med. 2011; 73:367–74. [PubMed: 21726926]
57. Organ. Econ. Coop. Dev. (OECD). OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD; Paris: 2007.
57a. P^3G Consort. Church G, Heeney C, Hawkins N, de Vries J, et al. Public access to genome-wide data: five views on balancing research with privacy and protection. PLoS Genet. 2009; 5:e1000665. [PubMed: 19798440]
59. Pearce N, Smith AH. Data sharing: not as simple as it seems. Environ. Health. 2011; 10:107. [PubMed: 22188646]
60. Piwowar HA. Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS ONE. 2011; 7:e18657. [PubMed: 21765886]
61. Pop M, Salzberg SL. Bioinformatics challenges of sequencing technology. Trends Genet. 2008; 24:142–49. [PubMed: 18262676]
63. Rodriguez H, Snyder M, Uhlén M, Andrews P, Beavis R, et al. Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: the Amsterdam Principles. J. Proteome Res. 2009; 8:3689–92. [PubMed: 19344107]
66. Skloot, R. The Immortal Life of Henrietta Lacks. Macmillan; London: 2010.
68. Tabor HK, Berkman BE, Hull SC, Bamshad MJ. Genomics really gets personal: how exome and whole genome sequencing challenge the ethical framework of human genetics research. Am. J. Med. Genet. A. 2011; 155A:2916–24. [PubMed: 22038764]
69. Trials of War Criminals Before the Nuremberg Military Tribunals Under Control Council Law No. 10. US Gov. Print. Off.; Washington, DC: 1946-49.
70. Trinidad SB, Fullerton SM, Ludman EJ, Jarvik GP, Larson EB, et al. Research practice and participant preferences: the growing gulf. Science. 2011; 331:287–88. [PubMed: 21252333]
71. Tutton R. Gift relationships in genetics research. Sci. Cult. 2002; 11:523–42.
72. Van Ness B. Genomic research and incidental findings. J. Law. Med. Ethics. 2008; 36:292–97. [PubMed: 18547197]
72a. Viertler C, Zatloukal K. Biobanking and Biomolecular Resources Research Infrastructure (BBMRI). Implications for pathology. Pathologe. 2008; 29(Suppl. 2):210–13. [PubMed: 18810444]
73. Watson JD. The Human Genome Project: past, present, and future. Science. 1990; 248:44–49. [PubMed: 2181665]
73a. Wellcome Trust. Sharing data from large-scale biological research projects: a system of tripartite responsibility; Rep., Wellcome Trust Meet.; Fort Lauderdale, FL. 2003 Jan 14-15; http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf

73b. Wellcome Trust. Sharing research data to improve public health: full joint statement by funders of health research. 2010. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm

73c. Wellcome Trust Case Control Consort. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]

74. Winickoff DE. Partnership in UK biobank: a third way for genomic property? J. Law Med. Ethics. 2007; 35:440–56. [PubMed: 17714253]

75. Winickoff DE, Winickoff MD. The charitable trust as a model for genomic biobanks. N. Engl. J. Med. 2003; 349:1180–84. [PubMed: 13679534]

76. Wolf LE, Lo B. Untapped potential: IRB guidance for the ethical research use of stored biological materials. IRB. 2004; 26:1–8. [PubMed: 15449408]

77. Wolf SM, et al. Managing incidental findings in human subjects research: analysis and recommendations. J. Law Med. Ethics. 2008; 36:219–48. [PubMed: 18547191]

77a. Wood, v. Commissioner of Police for the Metropolis. 2009 EWCA Civ. 414.

78. World. Med. Assoc. (WMA). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects; Helsinki, Finland. 2008; Adopted by the 18th WMA General Assembly June 1964, amended in 1975, 1983, 1989, 1996, 2000, 2002, 2004, and 2008 http://www.wma.net/en/30publications/10policies/b3

80. Zika E, Schulte in den Bäumen T, Kaye J, Brand A, Ibarreta D. Sample and data use and protection in biobanking in Europe: legal issues. Pharmacogenomics. 2008; 9:773–81. [PubMed: 18518854]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

SUMMARY POINTS

1. Next-generation sequencing and wide-scale data sharing challenge some of the basic principles for the protection of research participants and the current governance frameworks for research.
2. Open access policies are challenging the way that data are generated and distributed, which has led to significant changes in practice.
3. It is difficult to promise anonymity, withdrawal, or robust oversight mechanisms with wide-scale data sharing.
4. The use of e-governance as a complement to existing governance structures can help to protect privacy while building partnerships with participants.

Europe PMC Funders Author Manuscripts

FUTURE ISSUES

1. How can we respect and protect research participants when data sharing will increase over time?
2. How can we build privacy protections into e-governance mechanisms?
3. How can we inform research participants about the privacy implications of the use of their data and samples?