



# The growth and gaps of genetic data sharing policies in the United States

Jalayne J. Arias<sup>1,\*</sup>, Genevieve Pham-Kanter<sup>2,‡</sup>  
and Eric G. Campbell<sup>3,\*\*</sup>

1. Cleveland Clinic, Cleveland, OH 44195, USA

2. Drexel University School of Public Health, Philadelphia, PA 19104, USA and Harvard University, Cambridge, MA 02138, USA

3. Massachusetts General Hospital, Boston, MA 02114, USA and Harvard Medical School, Boston, MA 02115, USA

\*Corresponding author. E-mail: [ariasj@ccf.org](mailto:ariasj@ccf.org)

## ABSTRACT

The 1996 Bermuda Principles launched a new era in data sharing, reflecting a growing belief that the rapid public dissemination of research data was crucial to scientific progress in genetics. A historical review of data sharing policies in the field of genetics and genomics reflects changing scientific norms and evolving views of genomic data, particularly related to human subjects' protections and privacy concerns. The 2013 NIH Draft Genomic Data Sharing (GDS) Policy incorporates the most significant protections and guidelines to date. The GDS Policy, however, will face difficult challenges ahead as geneticists seek to balance the very real concerns of research participants and the scientific norms that propel research forward. This article provides a novel evaluation of genetic and GDS policies' treatment of human subjects' protections. The article examines not only the policies, but also some of the most pertinent scientific, legal, and regulatory developments that occurred alongside data sharing policies. This historical

† Ms Arias is the associate director of the NeuroEthics Program and associate professional staff in the Department of Bioethics at the Cleveland Clinic. Her work incorporates empirical and conceptual projects addressing legal and ethical issues inherent in medicine and medical research.

‡ Dr Pham-Kanter is an assistant professor in the Department of Health Management and Policy in the Drexel University School of Public Health and holds a research fellow appointment at Harvard University. Her research focuses on policy questions related to physician-industry relationships and conflicts of interest in medicine.

\*\* Dr Campbell conducts research relating to physician conflict of interest and professionalism in medicine. He is the director of research at Mongan Institute for Health Policy and a professor at Harvard Medical School.

perspective highlights the challenges that future data sharing policies, including the recently disseminated NIH GDS Draft Policy, will encounter.

**KEYWORDS:** data sharing, genetic, genomic, research, policy, scientific norms

In 1996, the Bermuda Principles launched a new era in data sharing in genetics by codifying expectations that researchers would share data with other researchers and the public.<sup>1</sup> The Bermuda Principles reflected a growing belief that data sharing would maximize the benefits of the data, advance research, and prevent wasting scientific resources. Policy makers have continued to echo this sentiment in policies developed since 1996, including the most recent National Institutes of Health (NIH) Genetic Data Sharing Policy issued in August 2014.<sup>2</sup> An example of the public goods benefit of data sharing is reflected in Paltoo and colleagues' findings of the scientific developments that resulted from data sharing of genome-wide association study (GWAS) studies.<sup>3</sup> They report that over 900 secondary analyses of GWAS data had been published since the enactment of the GWAS Data Sharing Policy in 2007. Additionally, they credit secondary analysis of deposited GWAS data for identifying associations between the human leukocyte antigen and Parkinson's disease, as well as other previously unknown associations. There are thus recognizable public benefits and scientific advancement that stem from data sharing policies. Moreover, because this research is financed by a publicly funded entity, one might argue that the public at large is entitled to data and products generated from NIH-financed projects. Yet, data sharing policies are not without challenges inherently linked to scientific and research cultural norms.

Data sharing policies run counter to scientific cultural norms, which rely on publication of research as a means to garner recognition and professional success. By requiring data generators to share primary data, policies may reduce the potential professional gain researchers have at stake by producing valuable data. Additionally, secondary use of data may lead to potential consequences for research participants who consent to research or otherwise provide tissue for research. Yet since 1996, a number of policies have been issued that recommend, mandate or otherwise encourage data generators to make their primary data available to other researchers and the public. Most recently in August of 2014, the NIH published the final Genomic Data Sharing Policy (GDS Policy), following commentary on the draft policy by stakeholders, including the scientific and research community. The progression and trends of policies relevant to United States genetic and genomic researchers developed between the Bermuda Principles and the GDS Policy reflect a changing view on scientific norms and how researchers, funding agencies, and the public view genomic data. Accordingly, data sharing policies have differed in their treatment of (1) timing requirements for the submission and/or release of data, (2) protections for data generators in light of the priority the research and scientific communities place on publication, and (3) human subjects' protections, specifically issues of privacy and confidentiality. The 2014 NIH GDS Policy reflects these

---

<sup>1</sup> Eliot Marshall, *Bermuda Rules: Community Spirit with Teeth*, 291 SCIENCE 1192 (2001).

<sup>2</sup> Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51,345 (Aug. 28, 2014).

<sup>3</sup> Dina N. Paltoo et al., *Data Use Under the NIH GWAS Data Sharing Policy and Future Directions*, 46 NAT. GENET. 9, 934–938 (2014).

trends by remodeling and expanding the policy to incorporate ongoing changes relevant to public concerns of control over their samples.

To understand the evolution of policies over time, we conducted an analysis of all policies published between 1996 and 2013 (Table 1) that would apply to genetic and genomic researchers in the United States, with a focus on NIH policies. Additional policies, including international policies (ie the Wellcome Trust) and journal policies, may be relevant to researchers. However, our focus was to determine how a major funding institution's policies have developed within the context of social and legal changes. The development and transformation of these policies is reflective of policy development within other agencies, including the National Science Foundation, as well as other countries.<sup>4</sup> An initial review of the policies identified core themes or trends within each policy, including treatment of the themes identified above. We coded each policy and summarized their treatment of themes. Using the summary, we analyzed policies' treatment of core themes within the context of scientific development, laws and regulations, and changing scientific norms relevant to data sharing. Using the analysis, we evaluated the potential challenges for current and future policies. This article evaluates the progression of policies within a historical context and evaluates the potential gaps remaining in the 2014 GDS Policy.

#### HISTORICAL DEVELOPMENT OF POLICIES AND BROADENING THE APPLICATION

Data sharing policies emerged following the Bermuda Principles developed in 1996 during the First International Strategy Meeting on Human Genome Sequencing in Bermuda.<sup>5</sup> Under these principles, endorsed by multiple international funding agencies, human genome sequence information '... should be freely available and in the public domain in order to encourage research and development and to maximize its benefit to society'.<sup>6</sup> Following the 1996 meeting in Bermuda, the Human Genome Project demonstrated the value of pre-publication release of data, which led to discovery of information on 30 disease genes.<sup>7</sup> In 1997, a second meeting reaffirmed the Bermuda Principles and addressed issues relevant to data quality. Also in 1997, the National Human Genome Research Institute (NHGRI) adopted its policy echoing the language from the 1996 and 1997 Bermuda Principles. This policy applied to all researchers conducting large-scale research funded by NHGRI.<sup>8</sup> Between its original policy in 1997 and 2003, the NHGRI extended and amended its data sharing policy three times to apply to broader types of data and to incorporate more specific details regarding the sharing of and access to data. The NIH adopted similar guidelines and policies, beginning in 1999 with guidelines that applied to all NIH-funded researchers.<sup>9</sup> In 2003, the

<sup>4</sup> National Science Foundation, *Award and Administration Guide: Other Post Award Requirements and Considerations*, <http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag/6.jsp#VID4> (accessed Oct. 20, 2014).

<sup>5</sup> Human Genome Project Information Archive, *Policies on Release of Human Genomic Sequence Data Bermuda—Quality Sequence*, [http://www.ornl.gov/sci/techresources/Human\\_Genome/research/bermuda.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml) (accessed July 20, 2014).

<sup>6</sup> *Id.*

<sup>7</sup> Toronto International Data Release Workshop Authors, *Prepublication Data Sharing*, 461 NATURE 168 (2009)

<sup>8</sup> National Human Genome Research Institute, *Current NHGRI Policy for Release and Database Deposition of Sequence Data 1997*, <http://www.genome.gov/10000910> (accessed July 20, 2014).

<sup>9</sup> Sharing Biomedical Research Resources: Principles and Guidelines for Recipients of NIH Research Grants and Contracts, 64 Fed. Reg. 246 (Dec. 23, 1999).

Table 1. Genetic data sharing policies: key provisions, language and treatment of research subjects protections.

Manuscripts submitted to Journal of Law and the Biosciences										
Year	1996	1997	1997	1997	1998	1998	1999	2000	2000	2010
Name	Bermuda	Bermuda-Quality Sequence	NHGRI Policy for Release and Database Deposition of Sequence Data	NHGRI Policy for Release and Data Deposition of Sequence Data	Fort Lauderdale	NIH Data Sharing Policy	GWAS Policy	Toronto Statement	Toronto Statement	Wellcome Trust Policy on Data Management and Sharing
US/mt1	International Centers funded for large-scale human sequencing	International Same as 1996 Bermuda Policy	US Researchers conducting large scale research funded by NHGRI	US Researchers conducting large scale research funded by NHGRI	International Sequence producers, users and funding agencies.	US Sequence producers and users (implicit: funded by NHGRI)	US Investigators applying for investigator-initiated funding for more than \$500,000 or if studies, including proposals for contracts and intramural research studies	International Builds upon Bermuda rules: "best practices" for producers, funding agencies, users, and publishers/editors.	International All Wellcome Trust funded researchers	International All Wellcome Trust funded researchers
What does it apply to	All human genomic sequence information	Reaffirms 1996 Bermuda Policies	Extends the NHGRI Policy from 1997	Extends the NHGRI Policy from 1997	Reaffirms the 1996 policy and lays out recommendations for funders, users, and funders. Extend the policy to raw and assembled sequence data and "community resource projects."	Adopts Fort Lauderdale's 2000 NHGRI policy to large-scale whole genome shotgun projects.	All "final" research data	Large reference data sets in biology and medicine (beyond genomics and proteomics)	"Databases that have clear scope for wider research use and hold significant long-term value." (Adopting Fort Lauderdale Principles and the Toronto Statement)	
Mechanism/Aiming for sharing	Sequence assemblies should be released as soon as possible; assemblies of greater than 100 would be publicly available; finished annotated sequences should be submitted immediately to the public database.	Encourages annotation of data	Assemblies greater than 2kb should be released within 24 hours	Adopts Cold Harbor Int'l Policy (2000); updating accessions within 24 hours of a significant change (i.e., achievement of 100% coverage and public availability) should be publicly available prior to their information into sequence assemblies.	Funders: Should require data sharing as a condition of funding. Producers: should publish Project Description, create data accessibility and restriction. Users: should appropriately cite the source	Sequence producers to share project description and users have responsibilities consistent with scientific norms.	Requires investigators to include a data sharing plan in their research proposals or indicate why a plan is not appropriate; data sharing is required for all studies for which the NIH GWAS data repository.	Requires proposals for prepublication. Recommends that funding agencies require that data-sharing plans be presented as part of grant applications. Reviewer should provide a "clearable" statement about the initial intention for data analysis.	Data should be available for prepublication. Recommends that funding agencies require that data-sharing plans be presented as part of grant applications. Reviewer should provide a "clearable" statement about the initial intention for data analysis.	Requires proposals for research funded by the Wellcome Trust to include a data sharing plan.
Privacy	Not Specifically indicated	Not Specifically indicated	Not Specifically indicated	Not Specifically indicated	Not Specifically indicated	Not Specifically indicated	Data is stripped of identifiers; informed consent procedures, privacy, but encourages data producers to inform participants that their data will be shared with other scientists	Does not specifically address privacy, but encourages data producers to inform participants that their data will be shared with other scientists	Does not specifically address privacy, but encourages data producers to inform participants that their data will be shared with other scientists	Does not specifically address privacy, but encourages data producers to inform participants that their data will be shared with other scientists

NIH released a formal data sharing policy requiring all researchers requesting funding for \$500,000 or more in direct funds in a single year to include a data-sharing plan in their application.<sup>10</sup>

The NIH genome-wide association study (GWAS) Data Sharing Policy emerged in 2007 as the predominant policy for genetic and genomic researchers in the United States. In comparison to other policies, the GWAS Data Sharing Policy provides in-depth recommendations and requirements for data generators and data users. The GWAS Data Sharing Policy applies to all ‘investigators who receive NIH support to conduct genome-wide analysis of genetic variation’, without limiting the scope of requirements to researchers who seek funding above a given amount. In 2009, the NIH announced intentions to extend the GWAS Policy to a broader audience. In 2013, the NIH published the draft GDS Policy that applies to a wider range of genomic research, including research that collects human and non-human data. The policy remained open for public comments through November 2013 and the finalized policy followed in August 2014.

However, policy activities were not limited to the United States. International agreements frequently correlated with changes or adoption of new policies in the United States (see Table 1). An international meeting would often promulgate new standards for data sharing that US policy makers would subsequently adopt. For example, the 1997 NHGRI Policy followed and adopted language from the Bermuda meeting in 1996 and 1997. Similarly, after the Fort Lauderdale meeting (2003), the NHGRI reaffirmed their data sharing policy and adopted language from the Fort Lauderdale Agreement, including the extension to additional types of data. This correlation reflects the relevance of these policies and the context within which they are developed. Additionally, due to the number of collaborations that cross borders, international policies are increasingly more relevant.

### TIMING AND MECHANISM FOR SHARING

Policies have consistently provided some guidance regarding the mechanisms and timing for sharing data with the public and other researchers. However, the details and specificity of timing and mechanism for sharing have transformed between 1996 and 2014. Policies changed as the capabilities of researchers progressed and as researchers began to analyze larger sequences. The first wave of policies encouraged the rapid release of assembled sequence data. Attendees to the 1996 Bermuda Principles agreed to release data ‘as soon as possible’ and to submit assemblies greater than 1000 units within 24 hours. Under these Principles, researchers would submit data to a public database known as GenBank. This was a significant departure from the prior standard of making data available within six months.<sup>11</sup> The 1997 NHGRI Policy adopted the Bermuda Principles, but changed the timing to indicate that investigators should release assemblies of 2000 units within 24 hours of their generation.

A second wave of policies began to recognize growing capabilities to produce larger sequences and that nuances specific to different types of data sets may justify

<sup>10</sup> Final NIH Statement on Sharing Research Data, NOT-OD-03-032 (Feb. 26, 2003), <https://www.genome.gov/EdKit/pdfs/1992b.pdf> (accessed Dec. 4, 2014).

<sup>11</sup> DOE-NIH Joint Subcommittee on the Human Genome, *NIH, DOE Guidelines Encourage Sharing of Data, Resources*, 4 Human Genome News (1993).

different timing for release. The 2000 NHGRI Policy extended the 1997 NHGRI Policy by requiring deposition of sequences following any ‘significant change’ within 24 hours. The policy also states, ‘[s]equence data, and all ancillary information specified in a standard format provided by the databases, should be released weekly’. The 2000 NHGRI Policy sought to encourage researchers to make sequencing data available before the sequence assembly was prepared.

A third wave of policies, following the Fort Lauderdale Agreement in 2003, began to request that investigators share descriptive information of their projects when submitting data. The 2003 NHGRI Policy indicates an expectation that investigators deposit sequence assemblies of 2000 or greater within 24 hours and that whole genome shotgun projects are to be deposited in a trace archive within a week. The 2003 NHGRI Policy also encouraged investigators to share project descriptions as did the GWAS Data Sharing Policy (2007). Descriptive information includes the protocol, questionnaires, study manuals, variables measured, and other supporting documentation. Under the GWAS Policy, investigators are encouraged to submit data to the database of genotypes and phenotypes (dbGaP).

The NIH’s adoption of a formal policy in 2003 was broad-sweeping and is an outlier of the progression of policies tailored to genetic and genomic research. The NIH Policy, which applied to all researchers, requires investigators to include a data-sharing plan within their grant application. While the 2003 NIH Policy does not mandate when or how investigators are to share data, it states that sharing ‘should be timely and no later than the acceptance for publication of the main findings from the final data set’.

The most recent 2014 NIH GDS Policy segregates the mechanism and timing for sharing data according to the type of data. The policy specifically distinguishes between human and non-human data. The policy provides an algorithm of differentiated ‘levels’ of data according to the type of data. For example, data after an initial round of analysis is a ‘level two’, including DNA sequence alignments. For this level, the data submission expectation depends on the project, but is generally expected within three months after data generation. Comparatively, level 4 data applies to the final analysis of data that relates the genomic data to phenotype or other biological states. Investigators are expected to submit level 4 data as the analyses are completed. Similar to the 2003 NIH Policy, investigators are required to submit a data-sharing plan that meets sharing expectations with their application for funding.

### PROTECTIONS FOR DATA GENERATORS

In the broader context of scientific culture, data sharing seems to contradict an accepted norm that prioritizes publication. Publication of research findings is the method of garnering ‘intellectual credit’ relevant to job prospects, future funding, and tenure consideration.<sup>12</sup> In fact, publication is classically described as the coin of the realm in science and the basis upon which the rewards of science (such as promotion, funding, fame, and ultimately a place in the history of science) are distributed. In this context, requirements to share data may expose researchers to being ‘scooped’ or in the least allow scholars and researchers to benefit from others’ work without incurring the initial cost of

<sup>12</sup> COMMITTEE ON RESPONSIBILITIES OF AUTHORSHIP IN THE BIOLOGICAL SCIENCES, NATIONAL RESEARCH COUNCIL, SHARING PUBLICATION-RELATED DATA AND MATERIALS: RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES 3 (The National Academies 2003).

research. Counter to this norm, the first wave of data sharing policies did not protect data generators' interest in publishing the data. Additionally, policies largely discouraged protecting an investigator's intellectual property interests to patent any type of data.<sup>13</sup>

The 2000 NHGRI Policy contained the first protections for data generators. The policy permitted data users to use data 'for all types of analyses'. However, under this policy users were to acknowledge data generators in subsequent publications and could not use the data to publish the initial publication of the whole genome sequence assembly or other large-scale analysis. Acknowledgement of data generators was carried forward in future policies. For example, the 2003 Fort Lauderdale Agreement recognized researchers' publication interests by encouraging secondary users to recognize the producers of the large-scale data.<sup>14</sup>

The NIH GWAS Data Sharing Policy promulgates two principles and methods for protecting primary researchers. First, the policy establishes boundaries for access to data through the repository. Under the policy, Data Access Committees (DACs) play a critical role in protecting primary investigators' intellectual interests. The GWAS Policy requires that researchers requesting access must include a Data Use Certification that includes restrictions relevant to data generators' interests. Second, the GWAS Policy establishes that the primary investigators will 'retain the exclusive right to publish analyses of the dataset for a defined period of time'. The period of exclusivity runs for a period of 12 months from the date the data is made available through the repository. During this time, secondary researchers may access and analyze the data, but cannot publish a secondary analysis.

The 2014 NIH GDS Policy does not institute an embargo for users to abide by, but incorporates an expected timeline for investigators to submit data to the repository. Timing for submission is determined according to the type of data to be submitted. In most circumstances, data generators are expected to submit data after the data has been cleaned, generally within three months of generation. While the timing of release from the repository will differ according to the type of data, generally data will be released at the time of publication or six months after data has been submitted. Once data is released to secondary researchers, there will not be any restrictions on users' publication or dissemination of their analysis. Under this structure, data generators are afforded nine months to analyze and publish their data before it will be released to other researchers for analysis. The policy maintains the expectation that users will acknowledge data generators in any subsequent publication or presentation of the data.

### HUMAN SUBJECTS' PROTECTIONS

The primary risk to research participants of genomic studies is the risk of loss of confidentiality and privacy concerns.<sup>15</sup> A significant challenge to the protections is making

<sup>13</sup> Policies varied on their treatment of patenting genomic data. Most policies discouraged patenting, but recognized the relevance of the Bayh-Dole Act. We do not address the treatment of patents in this article, due to the scope and purpose of this article. See Jorge L. Contreras, *Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons*, 12 MINN. J. L. SCI. TECH. 61 (2011).

<sup>14</sup> The Wellcome Trust, *Sharing Data from Large-Scale Biological Researcher Projects 1*, 4 (2003), <http://www.genome.gov/pages/research/wellcomereport0303.pdf> (accessed Dec. 4, 2014).

<sup>15</sup> William W. Lowrance & Francis S. Collins, *Identifiably in Genomic Research*, 3 SCIENCE 600 (2007).

data truly anonymous.<sup>16</sup> The process of coding data or presenting data in the aggregate may lessen the value of the data to a secondary user. Additionally, genetic data is, by nature, identifiable. Following findings by Homer in 2008,<sup>17</sup> the NIH removed relevant data from publicly accessible databases. Yet, early policies did not articulate explicit protections for research subjects, including privacy protections. The lack of consideration of privacy concerns in early policies is consistent with the perspectives of biomedical research in 1996, exemplified by the seminal case *Moore v. Regents of California* in 1990.<sup>18</sup> This case held that once removed, body tissue was no longer the property of the individual. Therefore, under this rationale, there would not be a ‘research participant’ after samples were removed from an individual. Additionally, prior to 1996 federal law did not protect medical information under the auspice of privacy.<sup>19</sup> Since then, with the passage of the Health Information Portability and Accountability Act (HIPAA) and other privacy protections for research participants, informed consent procedures and measures to protect an individuals’ privacy have become more relevant. Ultimately, there is a tension between maximizing benefits of sharing genetic data with increasing potential risks associated with the same activity.<sup>20</sup>

The 2003 NIH Policy is the first mention of human protections. Yet policies that directly applied to genomic research, including the NHGRI Policy and the Fort Lauderdale principles published in 2003, continued to lack any mention of human subjects’ protections. The NIH Policy articulates expectations that data ‘intended for broader use should be free of identifiers’. Additionally, the NIH Policy offers privacy as justification for researchers to refrain from including or executing a data-sharing plan.

In 2007, the NIH GWAS Data Sharing Policy was the first genome-specific policy to address privacy protections. The policy requires data generators to provide written certification that assures that research participants’ identities will not be disclosed to the repository and that the institutional review board or Privacy Board has reviewed and verified the submission of the data. The policy requires that submitting researchers and institutions fully de-identify all data. The policy also acknowledges that privacy concerns may be grounds for non-submission of data. DACs serve as gatekeepers for protecting the data from misuse by secondary users. The Preamble extends an additional level of protections for human subject by incorporating informed consent measures. The Preamble establishes an expectation that researchers will document discussions with research subjects that inform participants that researchers will share their genotype and phenotype, if GWAS is conceived as part of the study design at the time of consent. The Preamble goes on to indicate that if a participant withdraws his or her consent, it is the submitting institution’s responsibility to request removal from the repository. Data previously distributed for approved research cannot be retrieved from the user.

---

<sup>16</sup> Jane Kaye et al., *Data Sharing in Genomics— Re-Shaping Scientific Practice*, 10 NAT. REV. GENET. 331 (2009).

<sup>17</sup> Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 PLOS GENET. e1000167 (2008).

<sup>18</sup> *Moore v. Regents of California*, 793 P.2d 479 (1990).

<sup>19</sup> Institute of Medicine (US) Committee, *HIPAA, The Privacy Rule, and Its Application to Health Research*, in HEALTH RESEARCH AND THE PRIVACY OF HEALTH INFORMATION: THE HIPAA PRIVACY RULE (Sharyl J. Nass, Laura A. Levit & Lawrence O. Gostin eds., 2009).

<sup>20</sup> P3G Consortium et al., *Public Access to Genome-Wide Data: Five Views on Balancing Research with Privacy and Protection*, 5 PLOS GENET. e1000665 (2009).



The introduction of privacy and other human subjects' protections in 2007 is consistent with the environment in which the GWAS Policy was written. Several key developments were ongoing and beginning to impact genomics research. Genomic research data was advancing and beginning to provide information about individuals' traits. Societal and researchers' perceptions on risks associated with secondary analysis were also influenced by key events. In 2008, the Genetic Information Non-Discrimination Act (GINA) became law, protecting individuals from certain discriminations based on genetic information.<sup>21</sup> The passage of GINA reflects a societal appreciation for the potential consequences of disclosing genetic information and the importance of privacy in this arena. In 2004, the Havasupai Tribe filed a legal case against Arizona State University (ASU) for conducting a secondary analysis on collected samples that neither the tribe nor the participants had consented to.<sup>22</sup> Tribe members had consented to donating samples for research on diabetes, a public health concern relevant to the tribe. ASU researchers used these samples to conduct research on mental illness, inbreeding, and the tribe's origin story, all of which were perceived to be linked with stigma. While this case did not establish legal precedent because it settled in 2010 without adjudicating the legal issues, it placed additional scrutiny on investigators and secondary analysis of data. More recently, the family of Henrietta Lacks came to an agreement over the use of the HeLa cell and release of genomic information associated with the cell line. Per this agreement, the Lacks family must provide approval for the line to be used in future research.<sup>23</sup> Lastly, the Department of Health and Human Services released Advanced Notice of Proposed Rulemaking (ANPRM) to revise the Common Rule in 2011.<sup>24</sup> Proposed changes include alterations to research participant consent for research on biospecimens, even those that have been stripped of identifiers. The ANPRM for the Common Rule includes expectations that researchers will seek out consent from participants before secondary analysis of their samples may be conducted.

The 2014 NIH GDS Policy adopts many of the provisions in the GWAS Policy relevant to research subject protections and privacy. First, the policy permits primary researchers to name limitations of use for secondary researchers, which may also be limited by a participant during the consent process. However, unlike the 2007 policy which allows researchers to use privacy as grounds for non-submission, the 2014 policy allows for investigators to request an exception to submission of relevant data to the repository. However, investigators must provide an alternate data-sharing plan. Presumably, this may include sharing specific data directly with another investigator upon request through a data-sharing agreement between institutions. Additionally, the policy expands on protections by making informed consent a key component of policy language. The policy explicitly states that, in studies initiated after the effective date of the policy, NIH expects 'to obtain participants' consent for their genomic and phenotypic data to be used for research purposes and to be shared broadly'. The policy goes on to provide

---

<sup>21</sup> National Human Genome Research Institute, *Genetic Information Nondiscrimination Act (GINA) of 2008*, <http://www.genome.gov/24519851> (accessed July 21, 2014).

<sup>22</sup> Michelle M. Mello & Leslie E. Wolf, *The Havasupai Indian Tribe Case—Lessons for Research Involving Stored Biologic Samples*, 363 NEJM 204 (2010).

<sup>23</sup> National Institutes of Health, *Lacks Family Reach Understanding to Share Genomic Data of HeLa Cells* (Aug. 7, 2013), <http://www.nih.gov/news/health/aug2013/nih-07.htm> (accessed July 21, 2014).

<sup>24</sup> Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44512 (July 26, 2011).

additional specificity as to consent expectations for data collected in studies initiated prior to the effective date of the policy. For studies initiated before the policy becomes effective, researchers are to seek consent for any data from cell lines or clinical specimens collected after the effective date. However, for data from cell lines or specimens collected before the effective date varied approaches may be used. In these circumstances, IRBs or similar are expected to review consent documents to assess whether submission would be inconsistent with participants' consent.

### THE FUTURE FOR THE GDS POLICY

Genetic data sharing policy development and implementation will likely continue to be influenced by societal and legal progress. The recent public comments for the NIH Draft GDS Policy (published in November 2013) reflect societal and cultural norms through questions, concerns, and praise for the policy.<sup>25</sup> Among the 107 comments submitted, two prominent themes emerged: commenters raised concerns about (1) the proposed timetable structure for submission and release of shared data, and (2) informed consent in the context of balancing protections for human subjects while retaining feasibility for research. The supplemental materials for the 2014 NIH GDS Policy address these questions and comments. The prevalence of these themes within the comments is not surprising given the historical development of these issues within policies since 1996. Additionally, in future policies these themes are likely to be influenced by a changing legal and social environment.

Limitations on data user access or publication of shared data to protect data generators' interests continue to entail a delicate balance of interests and objectives. Overly restrictive embargos or limitations that prevent data users from accessing or publishing analysis of shared data may reduce the effectiveness of data sharing policies to advance science. Conversely, overly liberal policies that minimize restrictions expose data generators to risks of being 'scooped' and limit their ability to garner recognition from their investment. This risk could also reduce the incentive to pursue or develop primary research. A majority of public comments to the NIH draft GDS addressing data submission and release report concerns that timing provided to protect data generators' interests are not sufficient. Commenters expressed concern that the three-month time frame to submit data was too quick to assure quality of data, the six months or publication timeline for release did not allocate researchers sufficient time to analyze and publish data, or that the combination of both the submission and release time frame was not sufficient. Policy writers addressed these comments by first clarifying that the six-month deferral does not begin until the data have been cleaned and also indicating that the timeline for submission and release will be project specific. It is unclear from the policy what factors investigators or others might use to determine the timeline for a specific project. Additionally, the alteration of the timeline may have adverse consequences for data generators due to the time required to produce, analyze, and publish data. Recent research has shown that, although 77 per cent<sup>26</sup> of researchers who submitted their GWAS data to a repository were able to submit their first publication

<sup>25</sup> COMPILED PUBLIC COMMENTS ON THE DRAFT NIH GENOMIC DATA SHARING POLICY (Sept. 20, 2013–Nov. 20, 2013).

<sup>26</sup> Genevieve Pham-Kanter, Darren Zinner & Eric G. Campbell, *Codifying Collegiality: Recent Developments in Data Sharing Policy in the Life Sciences*, 9 PLOS One e108451 (2014).

using this data within 12 months, only 44 per cent were able to submit their first publication within six months.<sup>27</sup> The GWAS embargo differs in that it allowed secondary researchers to access and begin analyzing data before data generators had published their results. Per the NIH GDS Policy supplementary information, this was revised due to logistical challenges of tracking when secondary researchers could publish their findings. Under the NIH GDS Policy, secondary researchers would not have access to data before the release date. Given the different structure of protections for generators, it is unclear whether this change will unfairly disadvantage researchers who generate primary data or if six months is sufficient lead time ahead of secondary researchers.

The shift in policies between 1996 and 2014 demonstrates that policy developers are unclear as to how to best approach this issue. Unlike the development of human subjects' protections that have become increasingly more stringent, policies have vacillated on the issue of generator protections. The GWAS Policy established a specific embargo, an increase in protection. The 2014 policy uses timing and mechanism for sharing as a means of extending protections, an implicit and diminished protection for generators. The finalized GDS Policy indicates another attempt at balancing data generators' protections. In light of a more difficult and tenuous funding environment, researchers may be more inclined to accept funding with limited protections and appreciate the extension of value the collected data may offer by being shared at an earlier date.

Informed consent is an emerging challenge for finalizing the GDS Policy and future policies. As is reflected in the public comments, the GDS Policy is inconsistent with other regulations, including the Common Rule, in its application of informed consent. Currently, the Common Rule exempts research that involves 'existing data [...] if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects'.<sup>28</sup> This is further supported in the Common Rule's definition of 'Human Subjects' as a living individual who the investigator either interacts with or collects identifiable private information.<sup>29</sup> However, recent movement to revise the Common Rule would require informed consent for use of biospecimens, even when samples have been de-identified.<sup>30</sup> Given the coincident timing of finalizing ANPRM for the Common Rule and the GDS, it is possible that neither policy will be able to incorporate fully the other's language and requirements to guarantee consistency. If the policies do not incorporate consistent approaches, there is a significant risk of causing very practical complications for researchers. Researchers will be challenged to develop consent documents and processes that meet both set of requirements. For example, researchers will need to determine whether a general consent process conducted at the collection of data is sufficient for secondary data analysis. The 2014 GDS Policy supplementary materials directly address the potential that the policy and the ANPRM (once finalized) will be inconsistent by stating that the NIH will address inconsistencies once the Common Rule revisions are finalized. This, however, does not actually resolve likely

---

<sup>27</sup> Personal Communication with Dr Pham-Kanter on June 25, 2014 (Dr Pham-Kanter conducted a statistical analysis on existing data on first publication submission within six months).

<sup>28</sup> 45 C.F.R. § 46.101(b)(4).

<sup>29</sup> 45 C.F.R. § 46.102(f).

<sup>30</sup> Department of Health and Human Services, *Regulatory Changes in ANPRM*, <http://www.hhs.gov/ohrp/humansubjects/anprmanchangeable.html> (accessed Feb. 11, 2014).

challenges for researchers or places the burden on policy makers working on the Common Rule revisions to assure that the Common rule is consistent with the GDS Policy.

Informed consent is further complicated by the ability of research subjects to withdraw their consent. The 2014 GDS Policy permits subjects to withdraw their consent and thus require the research institution to remove their sample from the data set for future research. While theoretically this may be feasible, practically speaking this would require researchers to consider their method of anonymizing the sample, to be able to re-identify the sample and remove it if necessary. But by maintaining the sample as something that can be re-identified, the researcher is not actually fully de-identifying the sample. This then exposes a subject to additional privacy risks, which are the primary risks of participating in this type of research.

A final consideration of the consent process is the implications of secondary research on vulnerable populations or communities. While the most significant individual risk associated with genetic research may be loss of privacy, there are community risks that can be associated with findings of research that may stigmatize or have adverse consequences for a group. The *Havasupai v. ASU* case provides a clear example of how conducting research that is inconsistent with the consent document or was not perceived by the consenting participant as a possible use of the data at the time of consent may lead to group harms. Additionally, family members of a subject, like the Lacks family, may have a stake in the secondary use of the data, given the genetic nature of research. Given this, there may be types of research that an individual may have moral objections to and would not consent to if they knew that it was a possible use of their sample for secondary research. Researchers will be undoubtedly challenged to incorporate future uses of data within their consent procedures, particularly if the use has not yet been identified. The GDS Policy does not specify how to address issues of consent when the use may have stigmatizing consequences for a group or vulnerable population.

Despite the potential risks relevant to privacy, at least one study demonstrates that research participants would consent to researchers sharing their data with others in either through restricted or open access.<sup>31</sup> Data finding that current and potential research participants would support data sharing to maximizing the benefits of research data also found that participants place an emphasis on trust in researchers to protect their data.<sup>32</sup> This may reflect research participants' willingness to contribute to research or a trust in the system. However, another study reported that although research participants felt that they had enough information to make consent to enrollment in research studies that included consent to data sharing, many participants could not correctly report whether they had or had not consented to sharing their data.<sup>33</sup> Such data raises the question of whether there is a lack of understanding regarding the details related to sharing data, including potential risks, and thus support the argument that meaningful informed consent is not attainable in genomics research. Additionally, groups or populations with special interests, including tribal interests and rare disease

---

<sup>31</sup> Amy McGuire et al., *To Share or Not Share: A Randomized Trial of Consent for Data Sharing in Genome Research*, 13 GENET. MED. 948 (2011).

<sup>32</sup> Susan B. Trinidad et al., *Genomic Research and Wide Data Sharing: Views of Prospective Participants*, 12 GENET. MED. 8, 486–495 (2010).

<sup>33</sup> Jill O. Robinson et al., *Participants' Recall and Understanding of Genomic Research and Large-Scale Data Sharing*, 8 J. EMPIR. RES. HUM. RES. ETHICS 4, 42–52 (2013).

populations, serve as a contrast to individuals who consent or report to support sharing data derived from their samples. James and colleagues suggest that tribal representation and interests be considered and included in multiple aspects of research, including tribal representation on DACs.<sup>34</sup> The same could be said for rare disease populations or families with a stake in the research, similar to the Lacks family. However, inclusion of every group with a special interest in the entire research process is not feasible. Thus researchers, institutions, and the NIH will be charged with operationalizing mechanisms to appropriately balance the privacy and interests of groups as well as individuals. These concerns are mitigated in the NIH GDS Policy by the ability of primary researchers to identify limitations of research. However, this assumes that data generators are able to predict future uses that may lead to stigma or harms to a particular group.

### CONCLUSION

The progression of data sharing from the 1996 Bermuda Principles reflects ongoing changes in the environment affecting genomic researchers. As capabilities to connect genomic information with phenotypes have progressed, policies have been challenged to adapt. This is matched by growing societal concerns relevant to a loss of privacy with the disclosure of genetic information. Scientific norms regarding publication and the reliance on intellectual credit for career development challenges policy makers to consider how to protect not only research participants, but also to encourage research innovation by protecting data generators. The NIH's finalized GDS Policy marks another point of data sharing history. However, it is unlikely that this will be the end of the story. Future policies will continue to be challenged by privacy, consent, data generator protections, and the logistical questions surrounding the production and sharing of genetic data.

### ACKNOWLEDGEMENTS

Research supporting this manuscript was funded by the National Human Genome Research Institute (SR01HG006281). Additionally the authors would like to acknowledge Dr. Darren Zinner, PhD, The Heller School for Social Policy and Management, Brandeis University.

---

<sup>34</sup> Rosalina James et al., *Exploring Pathways to Trust: A Tribal Perspective on Data Sharing*, 16 GENET. MED. 820, 826 (2014), <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201447a.html> (accessed Oct. 20, 2014).