

**Comparing Genomic Data Sharing Policies from the National Institutes of
Health, Global Alliance, and Reg4All:
Common Ground and Future Directions**

Katherine L. Kwong

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Public Health

University of Washington

2014

Committee:

Helene Starks

Anna Mastroianni

Program Authorized to Offer Degree:

Epidemiology – School of Public Health

© Copyright 2014
Katherine L. Kwong

University of Washington

Abstract

Comparing Genomic Data Sharing Policies from the National Institutes of Health, Global Alliance, and Reg4All: Common Ground and Future Directions

Katherine L. Kwong

Chair of the Supervisory Committee:
Helene Starks, Associate Professor
Department of Bioethics and Humanities

Genomic data sharing has become increasingly important with “big data” genomics. Successful genomic data sharing requires multiple stakeholders cooperating with one another. Using discourse analysis, I compared three proposed genomic data sharing policies created by the National Institutes of Health, Reg4All and the Global Alliance for Genomics and Health. Data producers, data users, funders, participants, and end users were differently involved in the policy development process leading to policies that prioritize different needs and interests in genomic data sharing. The NIH policy satisfies the interests of data users and funders; Reg4All’s policy focuses on participants, end users, and data users, and the Global Alliance policy represents a compromise leaving all stakeholders somewhat satisfied. This analysis highlights how the policy options benefit the different stakeholders and suggests ways to create a system that more evenly addresses the concerns and interests of all stakeholders, allowing for more equitable genomic data sharing.

Acknowledgements

I would like to acknowledge Wylie Burke, Stephanie Malia Fullerton, Kelly Edwards, Ron Whitener, Rose James, Joon-Ho Yu, Erika Blacksher, and the members of the Center for Genomics and Healthcare Equality seminar for their insightful guidance and feedback during the development and writing of my thesis. This work was supported in part by a grant from the National Human Genome Research Institute to the Center for Genomics and Healthcare Equality, Wylie Burke, MD PhD, Principal Investigator (Grant # P50 HG 003374). It is based in part on earlier work by Kathleen West, Ron Whitener, Lorelei Walker, Stephanie Malia Fullerton, Wylie Burke, and Helene Starks.

I would also like to thank my cohort members, friends, and family for their support and encouragement. Particular thanks to the IPHG Qualitative Thesis Group, which provided feedback and empathy at several critical portions of the writing process.

Last but not least, immense thanks are due to my committee members, Anna Mastroianni and Helene Starks. Thank you both for your teaching and mentorship. Thank you for your patience, your critiques, your suggestions, and your faith in my project and me. Words truly cannot express my appreciation for your support.

Introduction and Background

Genomic data has become increasingly important as researchers search for answers to critical health problems by studying genetic factors. In order to maximize the benefits of genomic research, many researchers and organizations seek to work with larger data sets as a way to increase sample sizes and associated statistical power and improve the likelihood of finding significant genetic effects.¹ When researchers are able to access genomic data from other studies, they can create combined samples with more data than they themselves have created. In addition to creating larger sample sizes, pooling data from multiple studies can improve the racial and geographic diversity of the samples used, which can aid researchers who might otherwise only be able to use samples from more homogenous groups.²

But as genomic data has become increasingly important, so too have the policies that guide the process of sharing genomic data with others. The growing importance of these policies has led to increased scrutiny and questions about why the policies are written the way they are. Many stakeholder groups have expressed dissatisfaction with the state of genomic data sharing, arguing that the current dominant systems and policy models do not meet their needs in one or more ways.³

A number of characteristics present in current genomic data sharing policies are legacies of the long development process, influenced by past historical events in ways that may not be obvious. This history is critical for understanding the context in which current policies were written and which stakeholders' views were represented in the policy development process. Some of the earliest genomic data sharing policies were created by the U.S. National Institutes of Health (NIH), which has historically been the largest funder of genomic research in the world.⁴ In 1988, NIH and the U.S. Department of Energy created an agreement to jointly fund and

“coordinate research and technical activities related to the human genome.”⁵ This early involvement by NIH led to the 1989 founding of the Office of Human Genome Research (later the National Center for Human Genome Research and eventually the National Human Genome Research Institute).⁵ This specialized division of NIH became responsible for the Human Genome Project, which was conceived of as “international research effort to sequence and map all of the genes...of members of our species.”⁶ The Human Genome Project had an explicit goal of promoting “advanced methods for widely disseminating the information generated by the HGP to scientists, physicians and others.”⁷ In conjunction with other funders and research institutions, such as the Wellcome Trust in the United Kingdom, NIH co-sponsored several meetings and invited key stakeholders to come together to develop principles designed to ensure that genomic data would be shared among researchers to encourage further research development in the field of genomics.⁸ The agreements produced at these meetings (held in Bermuda) are called the “Bermuda Principles,” and represented a shift in genomic sequencing research from a paradigm of competition to one of cooperation and coordination.⁹ This represents NIH’s foundational concept of genomic data sharing – that genomic data should be shared publicly, to further the shared scientific enterprise of discovering information about genomes.⁸ The views and interests of the stakeholders involved in these early stages – primarily funders and researchers who were working toward the original collective goal of mapping the human genome – remain the focus of NIH’s genomic data sharing policy today.

The NIH’s genomic data sharing policy was developed based on earlier NIH policies related to data sharing. The NIH has had long-standing policies unrelated to genomics since 1984 that require that data from NIH-funded research be made available at a minimum to other researchers and ideally to the public.¹⁰ These general core directives provided a platform for NIH

to create specific data sharing policies applicable to genomics. Researchers conducting genomic research using more than \$500,000 of NIH funding are required to submit plans for how they will make the genomic data from their studies available through a “widely used data repository.”¹⁰ The original 2007 genomic data sharing policy creating this requirement limited it to certain types of genomic research; the draft 2013 policy is controversial in part because it greatly expands this requirement, making it apply to a much greater proportion of NIH-funded genomic research.¹⁰ The original participants in this process were primarily organizations and individuals who were interested in sharing genomic data for the purposes of advancing research into the structure and sequence of the human genome. In 2007, clinical utility was still mostly a goal rather than an operational capability, and the primary use of data was to help guide additional data production.⁹

Table 1 includes a summary of key meetings and documents that represent important points in the development of NIH’s current genomic data sharing policy. This history represents the context in which NIH’s current policy was developed, and provides insight into how the perspectives of funders and researchers who wanted to use genomic data became so integral to NIH’s policy. Each of these documents represents another building block of NIH’s current policy – and each document was developed primarily by NIH, other funders, and researchers who shared NIH’s goals and missions.

Table 1: Summary of NIH Genomic Data Sharing Policy Development History

Year	Document	Description
1984	NIH Policy on Distribution of Materials ¹¹	Established NIH's default practice to mandate data sharing
1991	NIH-Department of Energy joint guidelines on data sharing ¹²	Established 6-month timeline for release of research-generated data
1996 and 1997	Bermuda Principles ⁸	Produced in conjunction with other funders and worldwide research community; codified assumption that genomic sequences were public resource that should be shared by all as soon as possible, and created shared standards for when and how genomic data sharing occurred. Revised and expanded in 1997.
2003	Fort Lauderdale Principles ¹³	Reaffirmed and elaborated upon the Bermuda Principles extending and refining the responsibilities of funders, data producers and data users.
2003	NHGRI Rapid Data Release Policies ¹⁴	Described large-scale sequencing as a community resource and implemented policies for pre-publication release of data
2007	NIH Genome-Wide Association Studies (GWAS) policy ¹⁵	Created standards requiring the submission of genomic data from NIH-funded studies to dbGaP (Database of Genotypes and Phenotypes, NIH repository for genomic data from GWAS, created 2006); established standards for data quality and maintenance
2009	Papers summarizing discussions at Toronto and Rome meetings ¹⁶	Resulted in NIH announcing its intention to expand its GWAS policy to apply to more types of genomic data
2013	Draft NIH Genomic Data Sharing Policy ¹⁰	Proposes greatly expanding NIH's requirements for genomic data sharing. Announced 9/27/2013; public comment period ended 11/20/2013. Response to public comments and final policy not yet announced as of June 2014.

The long history of sharing research data and the context in which NIH became involved in genomic research helps explain why NIH's lens on genomic data sharing primarily focuses on

the needs of research funders (such as NIH) and researchers who want to use existing genomic data for secondary analyses. Both of these stakeholders are critically important to the process of producing and using genomic data. The researchers who are generating genomic data (data producers) are also an important stakeholder group. As the technology of genomic research has changed, so have NIH's requirements for the type, quantity, and quality of genomic data that must be made publicly available. Data producers and users have also become two distinct groups – although they continue to overlap in many cases, not all data producers seek to use genomic information from NIH, and not all data users also produce genomic data of their own.

In addition to these three stakeholder groups which have been part of the data sharing policy development process from the start, there are a number of stakeholders who did not participate in the earlier process and have concerns about the policy and the potential harms of genomic data.³ Research participants represent one such important stakeholder group that has been historically absent. Many advocacy groups have concerns about potential harms to participants due to NIH's genomic data sharing policy. Specifically, some population subgroups have raised concerns about the potential for group harms and stigmatization that may arise out of linking genomic data with racial, ethnic, or other demographic information.¹⁷ Individuals with sensitive genetic conditions have also raised concerns about how their data will be used.¹⁸ Because NIH's policy does not directly address these concerns, the mandatory sharing of genomic data may hinder recruitment for future genomic studies.

Finally, NIH has stated goals to ensure that its research data are translated into clinically actionable information.¹⁹ Yet because the NIH policy was created within a research context, clinical end users were not included in the data sharing policy development process, raising

concerns about whether this stakeholder group's preferences are currently being adequately considered.

The differing perspectives of these stakeholder groups reveals a tension inherent within the field of genomic data sharing: although the potential benefits of genomic data sharing are great, so are potential risks to participants (discussed below). The principle of beneficence requires that NIH maximize benefit and minimize harm to research participants. The ethical conduct of research thus requires resolution of this tension between benefit and risk in genomic data sharing by NIH.

The risks of genomic data sharing for participants include worries about privacy concerns²⁰ and stigmatization and discrimination based on the genomic information. These concerns are warranted: for example, a study by Gymrek et al demonstrated it was possible to identify participants from their genomic data,²¹ raising concerns about stigmatization and discrimination. These concerns are particularly acute for minority subpopulations,²² which already tend to be underrepresented²³ in genomic research and genomic medicine. NIH's mandatory genomic data sharing policy may exacerbate the effects of these existing inequalities by fostering rules that do not address these concerns, thus contributing to low rates of minority participation in genomic research.

The benefits of genomic data sharing can be demonstrated through NIH's experience. NIH's genomic data sharing policy has successfully made substantial amounts of genomic data available to researchers and the public by placing the data in accessible databases. As of August 2010, more than eight hundred investigators had received approval to access data from NIH's primary genomic data sharing resource, the database for Genotypes and Phenotypes (dbGaP), with more than 2700 requests approved in dbGaP's first three and a half years of operation.²⁴

The secondary analyses conducted using dbGaP data sets would likely not have been possible without the mechanism for genomic data sharing provided by NIH policy.²⁴

Two other policy models for genomic data sharing have emerged recently; one is sponsored by the Genetic Alliance with its Reg4All as a data sharing service, and the other is sponsored by the Global Alliance for Genomics and Health (Global Alliance). Reg4All is a patient-centered service for people with genetic disorders who wish to connect with researchers interested in doing research about their conditions.²⁵ It was established in 2013 by Genetic Alliance, a patient advocacy group that was founded in 1986 to establish collaboration and cooperation among patients with genetic diseases.²⁶ Reg4All is designed to support participants' control of their data sharing by explicitly stating that they may decide when, how, and with whom their data are shared.²⁵ Reg4All allows patients to submit clinical and genetic data, and then use fine-grain consent processes to allow patients to closely control who is able to see their data and specify how their data are used. For example, patients can indicate the types of research and/or disorders to which they would authorize use.²⁵ Reg4All remains in a "beta" state as of June 2014, indicating that it is not considered a final product platform at this time.²⁵ No data are currently available on the number of participants who have successfully signed up for the service.

The Global Alliance represents a coalition of over 190 nonprofit organizations, advocacy groups, research institutions, private sector companies, and health care organizations in 40 countries around the world, including both the Genetic Alliance and NIH.²⁷ Created in 2013, the Global Alliance seeks to design a system that also facilitates the use of shared genomic data and analyses for application in clinical and other settings,²⁸ but uses a more dispersed model of control and distribution than NIH's model of centralized control. The Global Alliance issued a

white paper in 2013 which proposed a new structure for sharing and curating genomic data.²⁸ Its proposal describes technological and governance solutions to solve problems which remain unaddressed by NIH's policy, and provides a contrasting view on the ethical, technological, and privacy issues inherent to genomic data sharing. The Global Alliance is still developing and adding new member organizations, and its leadership continues to issue additional working papers with proposals and analysis on a variety of issues related to genomic data sharing.²⁹ The Global Alliance held its first plenary meeting in April 2014; at the conclusion of that meeting, working group members announced their timelines for release of additional scoping and framework documents, with delivery expected between the end of 2014 and the end of 2015.³⁰

By analyzing the language used in the policies of NIH, Reg4All, and Global Alliance, this research aimed to identify the ways in which the competing preferences of different stakeholder groups are addressed in each of these policies.

Methods

I utilized discourse analysis to illuminate the values and perspectives of different stakeholders for data sharing policies proposed by the NIH, Genetic Alliance, and the Global Alliance for Genomics and Health. Discourse analysis is a method that focuses on the use of language as a social practice and as “the medium for interaction.”³¹ The language being used both reflects and informs social structures, creating dominant discourses that reflect default or taken-for-granted positions based on powerful stakeholders and/or prevalent views.³¹ Policy documents play an important role in shaping, maintaining, and challenging the dominant discourse about the value of genomic data sharing and the ways in which data sharing practices advance the goal of improving the efficiency and reach of genome science. Discourse analysis

techniques help to reveal the underlying social constructions behind the use of words, such as the motivating values and organizing principles and assumptions behind the language in data sharing policies.³²

Data Sources

This analysis drew on a number of publicly-available texts (listed in the Appendix) that represent both the dominant and alternative discourses. The NIH policy represents the dominant discourse in the United States on how genomic data sharing should be done. The NIH assumes this position due to its role as the preeminent funder of genomic research in the U.S., and the explicit linkage between its genomic data sharing policy and funding availability.¹⁰ In November 2013, NIH shared a draft of proposed revisions to its 2007 genomic data sharing policy and invited public comments during a 60-day period.¹⁰ The comments are available on the Genomic Data Sharing page of the NIH website.³ The comments, as well as the introduction and appendix of the proposed policy, were included as sources used to characterize the general themes in critique or support of the proposed changes, as well as to identify key stakeholder groups for inclusion in my analysis. Other NIH documents that informed the development of the 2013 policy were also analyzed.

Data sources for the Genetic Alliance's Reg4All policy included publicly-available interviews by its founders, its recruitment materials, its privacy policy, and other documentation available on its website. The Global Alliance's policy was analyzed based on news reports, press releases, the 2013 white paper, and papers by the working groups that were available as of June 2014.

All documents which were used as sources are listed in the Appendix. For the purposes of this analysis, “document” refers to an individual document listed as a row in the appendix, and “policy” refers to all of the documents listed for an organization in the Appendix. The “NIH policy” referred to in this analysis consists of a number of documents ranging from the Bermuda Principles to NIH’s genomic data sharing website; similarly, the “Reg4All policy” and the “Global Alliance policy” refer to the collection of documents listed in the Appendix. All of the documents are available to the public online.

These three organizations were selected because they represent different mixtures of stakeholders involved in policy development and thus offer different perspectives on genomic data sharing, which provides rich ground for comparison. All of these policies are currently under continuing development. Genetic Alliance’s Reg4All remains in “beta,” indicating that it is considered incomplete; Global Alliance is in the process of specifying and clarifying the concepts laid out in its whitepaper; and the NIH policy is still a draft policy, not yet formally finalized. This ongoing development makes all three policies interesting because of the opportunity for change.

Analysis

My analysis involved an iterative, three-step process that began with a close read of the selected texts to identify key stakeholders and the topics of primary interest to them that were addressed in the policies. For the first step, I organized all of the documents by organization, then put them in chronological order and read them closely multiple times using the following structured questions to guide my initial analysis.³³

Structured Questions:

1. What are the requirements of this document?
 - a. To whom is this data sharing document applicable?
 - b. What types of data must be shared?
 - c. What flexibility is provided in these requirements?
2. What control is given to different groups under this document?
 - a. Participants
 - b. Researchers
 - c. Oversight organizations
 - d. Broader societal groups (cultural groups, other groups with a specific condition, etc)
3. What control mechanisms are built into the document to ensure enforcement?
 - a. What preventative measures are included to prevent violations?
 - b. What enforcement mechanisms are there to react to breaches of the document?
4. How are risks addressed in this document?
 - a. Discussed
 - i. Accepted
 - ii. Dismissed
 - b. Ignored
5. What protections are included in this document?
 - a. For individuals
 - b. For groups
 - c. For organizations
6. What rights are included in this document?
 - a. For individuals
 - b. For groups
 - c. For organizations
7. What assumptions does the text make regarding the appropriate balance between protection and ease of access to the data?
 - a. Stated values
 - b. Implied values
 - c. Values that are not included
8. Which groups/values are given priority/privileged by this document?
9. How does the text respond to the concerns and values of other discourses?
 - a. Explicitly
 - b. Implicitly
 - c. Critiques that are ignored/not addressed
10. Who is writing/influencing this document?
 - a. Explicit role
 - b. Indirect role
 - c. Who is excluded from the document creation process?
11. What is the stated goal in regards to access to and inclusion in genomic research?

In the second step, I conceptualized primary stakeholder groups in five ways: data producers, data users, funders, participants (as data contributors) and end-users. I developed these roles based on descriptions of stakeholders within the three policies, as well as the stated positions of the public comments on the NIH policy. *Data producers* represent researchers and scientists who are generating new genomic data. *Data users* represent individuals or organizations who seek to reanalyze or utilize genomic data which has been shared. *Participants* represent the individuals who are the subjects of research, whose genomic data is being analyzed and shared. *Funders* represent organizations or individuals who are providing financial support for genomic research, generally government and non-profit funding agencies. *End users* represent individuals such as patients and clinicians who seek to utilize clinically- or personally-actionable information derived from genomic data, or other end products of research.

Individuals or organizations may have more than one role; however, their needs and preferences will probably vary based on their role. For example, a researcher could represent two roles: producing new data and using data from other research studies. Researchers' preferences for data sharing will likely vary depending on which role they are considering – their preferences as a data user may differ from their preferences as a data producer. These roles are broad generalizations representing specific viewpoints, and are not intended to represent every possible nuance of perspective; rather, they are designed to provide a useful framework for policy analysis from multiple perspectives.

In this part of the analysis, I also summarized the dominant preferences for data sharing activities for each of these stakeholder groups across three content domains that encompassed the majority of the policy concerns. The content domains included (1) consent and privacy preferences; (2) preferences for technical considerations of contributing to and receiving data

from data repositories; and (3) specifying controls over data access and use. *Consent and privacy* preferences focus on the appropriate balance between making the consent process for research and genomic data sharing streamlined and protecting the privacy and autonomy of research participants. The views on consent range from broader types of consent, in which participants are asked at the beginning of research to consent to the general reuse of their data to enable the most expansive distribution and use of genomic data; narrow consent, which limits use to specific types of research; to flexible consent, which gives participants control over their data and the opportunity to stay involved in the process and update their consent for new data use requests.

Technical considerations preferences focus on who bears responsibility for cleaning, standardizing, uploading, downloading, and distributing genomic data. The process of cleaning, annotating, and standardizing data sets can be expensive and time-consuming but makes the data easier to reuse and combine with other data sets. The technical requirements for systems that distribute data sets and make them available to others are also important, because poorly-designed or maintained systems can cause frustration and inability to access and use data. Yet overseeing and maintaining data sets and distribution systems are a burden and expense; at stake is which entity should have and maintain this responsibility.

Control of data preferences relate to who maintains control of the genomic data once it is shared, as well as mechanisms for oversight. This includes preferences about tracking data usage, who gives permission to use data sets, and how the use of data is acknowledged and controlled. Table 2 provides an overview of the preferences for each of the stakeholder groups across these three domains.

Table 2: Stakeholder Preferences

	Domains		
Stakeholders	<i>Consent and Privacy</i>	<i>Technical Considerations</i>	<i>Control of Data</i>
<i>Data Producers</i>	Prefer simplified consent and privacy procedures to remove barriers and burdens during data collection process	Prefer someone else be responsible for data cleaning, annotation, or submission process	Prefer to maintain control over data that is produced, particularly for attribution purposes
<i>Data Users</i>	Prefer simplified consent and privacy procedures to enable easy combining and reanalysis of data sets	Prefer cleaned, annotated, standardized data that is easy to download	Prefer simplified and centralized access to data, with streamlines or no approval process
<i>Participants</i>	Prefer fine-grained privacy controls and the ability to selectively consent depending on personal preferences for data consent	Prefer to have little to no interaction with technical considerations	Prefer to maintain personal control over data
<i>Funders</i>	Prefer simplified consent processes that enables standardized policy and easy reuse of data	Prefer to receive data that has been cleaned, annotated, and standardized	Prefer complete funder control over data produced by funded studies
<i>End Users</i>	Prefer consent and privacy controls that enable use by clinical users as well as researchers	Prefer data that is accessible and usable by clinical end users	Prefer centralized control of data to enable simplified access by end users

In the third step, I evaluated how well each of the policies satisfied the stakeholders’ preferences in each domain on a five-point scale: mostly satisfied, somewhat satisfied, somewhat dissatisfied, mostly dissatisfied, and not ascertainable (N/A). A score of “mostly satisfied” indicated that preferences were completely or nearly completely satisfied. “Somewhat satisfied” indicated that preferences were substantially fulfilled in a domain, but some important aspects of their preferences were not satisfied. “Somewhat dissatisfied” indicated that a policy met some preferences, but failed to meet many important aspects. “Mostly dissatisfied” indicated that all or nearly all preferences were not met by a policy in a given domain. Policies were rated as N/A in a domain if a policy either did not address a given domain or if the wording was such that I could

not determine its impact upon a role. Each domain was rated for each stakeholder role, and assigned one of these five categories.

Results

Policy Differences Across Domains

Table 3 presents selected quotes from each of the three policies that are representative of the way the policies address each of the three domains. In the domain of consent and privacy preferences, the NIH policy favors one-time broad consent, while the Reg4All and Global Alliance policies emphasize the need to allow participants to control their data on an ongoing basis in real time. In the domain of technical considerations, the NIH policy places the burden of data cleaning and submission solely on data producers. Reg4All accepts data straight from participants and clinicians. Global Alliance creates standards for genomic data, and distributes the technical and financial burdens across multiple groups. In the domain of control of data, NIH gives all control over the data in their databases to NIH-run committees. Reg4All and Global Alliance both advocate for participant control of data.

Table 3: Domain Mapping by Policy

	NIH Policy	Genetic Alliance (Reg4All)	Global Alliance
Consent and Privacy	"NIH expects the informed consent process and documents to state that a participant's genomic and phenotypic data may be shared broadly for future research purposes" unless an exception is approved at the time of application" ³⁴	"Participants can change their choices at any time. The decision to allow (or not allow) a requested use of, or access to, a participant's data is <u>always</u> based on the privacy preferences in effect at the time such use or access is attempted." ³⁵	"By enabling on-going communication, individuals can give consent to research, specify levels of personal privacy and become partners in the research process. By providing control over personal information and the potential to give on-going consent in real time, these initiatives meet international legal standards for the protection of privacy." ³⁶
Technical Considerations	Requires submission by data producers of "Data after an initial round of analysis or computation to clean the data and assess basic quality measures" and "Analysis to identify genetic variants, gene expression patterns, or other features of the dataset" on a project-specific timeline, "generally within 3 months after data generation" ³⁷	"All information collected and stored in the Reg4ALL database is provided directly by participants, through their responses to questions, or is health information participants have asked their healthcare provider to send to us." ³⁸	"Operating entities will need to be well- managed and resourced enterprises, and offer all or some of the following: (a) technical capabilities for storage, processing, analysis and/or controlled sharing of information, (b) datasets provided by their partners and users; (c) management of data access and compliance with local informed consent and data use provisions; (d) an interface between users, cloud providers and external software developers; (e) hosting of portals that allow users (researchers, clinicians, patients) to access data and results, (f) management of withdrawals of data (as requested)." ³⁹
Control of Data	"Applicable studies with human genomic data should be registered in the database of Genotypes and Phenotypes (dbGaP) no later than the time that data cleaning and quality control measures begin. Investigators should submit human data to the relevant NIH-designated data repository (e.g., dbGaP, GEO, SRA, the Cancer Genomics Hub)." Once in dbGaP, "Requests for controlled- access data are reviewed by NIH Data Access Committees (DACs)." ⁴⁰	"Participants cannot make genuinely informed decisions when sharing and decisions about secondary use of their data is beyond their reach and control." ⁴¹	"Advance the idea that patients have a right to share genomic and clinical information to advance human health, as well as to privacy and to transfer data as they choose" along with "punishment for mishandling of data." ⁴²

Stakeholder Groups Privileged by Each Policy

NIH's policy was primarily designed to operationalize the values expressed by NIH as a whole, and the Bermuda Principles specifically, with the goal of creating clear expectations for how and when genomic data will be shared. From NIH's perspective, the purpose of the policy is to make it simple for NIH to administer the data sharing process, and to streamline the process of getting data back out of the system once it is submitted. As such, its policy primarily prioritizes the perspectives of funders and data users. Little to no priority is given to the concerns of data producers, participants, or end users. The NIH's priority is not on simplifying the process for data producers; they have already received NIH funding, and dbGaP submission is just a corresponding requirement of getting this funding. Given this context, it makes sense that the emphasis within the NIH policy is on streamlining the process for funders and data users. The NIH's policymakers do not see its role in genetic data sharing as one of needing to protect human subjects; that responsibility is delegated to the institutional review boards (IRBs or research review committees) of the data producers, not NIH or dbGaP.

The Genetic Alliance developed Reg4All in part due to the lack of participant-controlled resources for genomic research. The Genetic Alliance's history as a patient advocacy organization further influenced the development of Reg4All.²⁶ Accordingly, the policies of Reg4All and the Genetic Alliance place a strong emphasis on the perspectives of participants and the end users of genomic data. Some consideration of the data user perspective (namely, researchers who hope to use Reg4All's data) is represented. The interests of data producers are not represented in Reg4All's policy because they receive data directly from participants and clinicians. Similarly, the interests of funders (such as large government or nonprofit entities) are

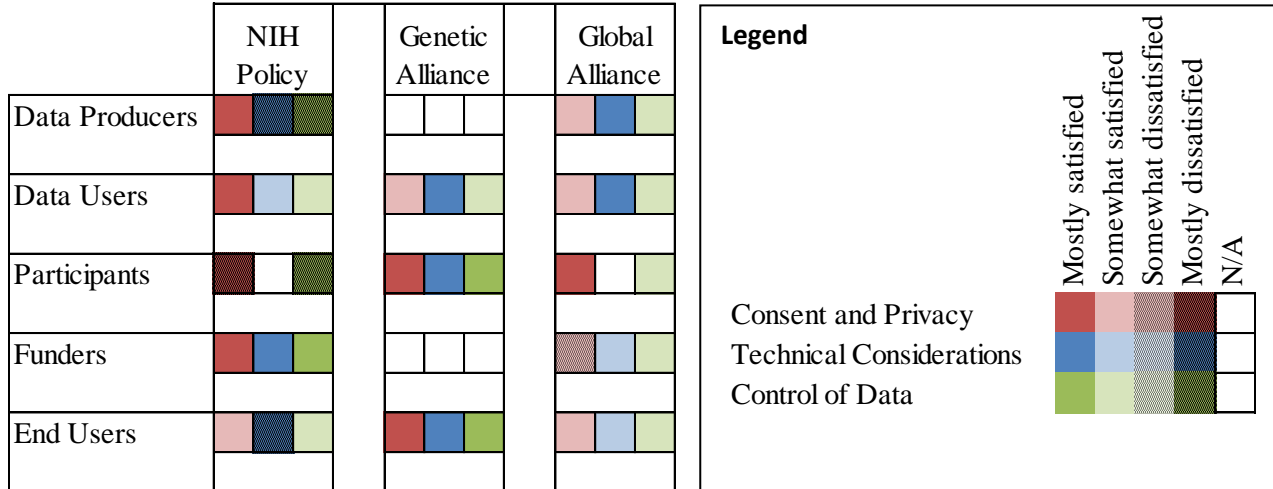
not addressed in Reg4All's current policy; Reg4All is currently funded by a start-up grant from a pharmaceutical company, some operational funding from Genetic Alliance, and donations.⁴³

The Global Alliance is made up of an array of stakeholders from around the world, so it is logical that their policy represents a consensus proposal that seeks to satisfy multiple stakeholder perspectives. Satisfying multiple perspectives means a policy compromise in which no stakeholder group is either fully satisfied or dissatisfied with the Global Alliance proposal.

Policy Alignments with Stakeholder Preferences

Each of the three policies satisfies some stakeholder groups. In general, NIH policy tends to create strong “winners” and “losers” – that is, stakeholders' preferences tend to be very satisfied or very dissatisfied. The Reg4All/Genetic Alliance policy satisfies the preferences of the stakeholder groups it addresses, but does not address two of the stakeholder groups. The Global Alliance generally evenly distributes satisfaction across stakeholder groups, resulting in most groups' preferences being somewhat satisfied rather than an extreme level of dis/satisfaction. The Figure below depicts the satisfaction ratings given to each organization's policy for each domain.

Figure: Stakeholder Satisfaction with Policy Positions



Data Producers

Data producers’ concerns in the three domains were satisfied to varying extents across the three policies. The NIH policy mostly satisfies data producers’ preferences in the area of consent and privacy by encouraging broad consent, but places heavy technical burdens on data producers in terms of preparing their data to comply with dbGaP standards. NIH also maintains control of the data once in their hands, leading to dissatisfaction.

Reg4All’s policy does not address the interests of this stakeholder group separately from participants, because under Reg4All’s policy model, participants and clinicians provide the data, not researchers.

Data producers’ consent and privacy preferences are somewhat satisfied by the Global Alliance policy, because the Global Alliance supports fine-grained participant control of data, but does so through the creation of standardized systems to avoid burdening data producers with the consent process. Global Alliance also intends to provide standards and infrastructure support for cleaning, annotating, and submitting data, relieving data producers of much of this burden

and satisfying the technical considerations preferences. Although the Global Alliance proposal does not allow data producers to have complete control over how their data are used, the Global Alliance's policy does include provisions for attribution of data producers.

Data Users

Data users' preferences in the three domains are reasonably well served by all three policies. The NIH policy's encouragement of broad consent allows data users to reuse data sets more easily than if fine-grained participant control were allowed. For technical considerations, NIH only somewhat satisfies data users; while NIH does require that data producers clean and annotate data sets, the lack of standardized processes for this means that data users cannot always easily combine data sets from different sources. In the domain of control of data, NIH maintains control over much of the genomic data through the centralized dbGaP, which somewhat satisfies data users. However, the NIH policy requires users to register and ask permission for use from Data Access Committees.

Reg4All and Global Alliance share a pattern of how well they satisfy data users' preferences in each of the three domains. Both advocate for fine-grained consent controls, which prevent them from satisfying some of data users' preferences; however, since both organizations also advocate for standardized systems to make fine-grained consent less burdensome for data users, they do partially satisfy data users' preferences in consent and privacy. Both Reg4All and Global Alliance satisfy data users' technical preferences by envisioning standardized technical platforms and simplified protocols. Both organizations allow some control of data by participants, and in so doing, only partially satisfy data users' preferences; however, standardized protocols for data access offset this and thus further contribute to satisfying these preferences.

Participants

The degree to which participants' preferences are satisfied by the three policies varies a lot. Participants' preferences for consent and control of data are not well satisfied by the NIH policy because broad consent and lack of participant representation on the Data Access Committees leaves participants with few options about how they can remain involved and have a say about how their data are used in future research. Similarly, participants do not have much of a role in the domain of technical considerations.

Reg4All's policy satisfies participants' consent and privacy preferences by supporting fine-grained, ongoing consent processes. The Reg4All policy satisfies participant technical preferences by envisioning a simple-to-use interface for participants. Reg4All gives participants real-time control over their data, and so satisfies their preferences in this domain.

The Global Alliance's policy also satisfies participants' preferences in the consent domain through fine-grained consent procedures. The Global Alliance does not envision participants having a role in the technical considerations domain and does not address them. Control of data is distributed across both participants and Global Alliance organizations, so while participants' preferences are partially satisfied in this domain, there remains some dissatisfaction with control of data.

Funders

The three organizations' policies satisfy funders' preferences to different extents. The NIH's policy encourages broad consent, which fulfills funders' consent preferences and allows them to ensure they maximize the amount of research that can be done with a single data set generated with their publically-funded research dollars. By requiring data producers to take responsibility for the cleaning, annotation, and standardization of data, the NIH's policy satisfies

funders' preferences in the technical considerations domain. The NIH's policy also satisfies funders in the area of control of data, by granting sole control of the data to the funding entity rather than allowing data producers or participants to have partial or complete control over the data.

Reg4All's policy envisions the creation of a centralized repository of patient- and clinician-provided data, which researchers can then access by request. Their proposed research model does not utilize traditional grant-style funding from government or nonprofit entities to produce the data being entered, and so their policy does not address the funders' group directly.

The Global Alliance somewhat fails to satisfy funders by encouraging fine-grained consent, which may mean that funders cannot promote reuse of the data they have paid to produce. Global Alliance partially satisfies the technical preferences of funders by creating standards for cleaning and annotating data, but does not completely satisfy funders' preferences because they share responsibility for this with data repositories. Funders' data control preferences are partially satisfied by Global Alliance's proposal to centralize systems for controlling data access; however, some control is distributed to participants, rather than allowing funders exclusive control over data.

End users

End users' satisfaction with the policies across the three domains also varies among the three policies. The consent provisions of the NIH policy somewhat satisfy end user's needs because it encourages a range of uses; broad consent implies consent for clinical use as well. The NIH's policy was designed with a research context in mind and lacks systems to enable easy access to data by clinicians. For example, there is no mechanism for clinicians to receive access to data on a short timeline; they must submit data requests like researchers do, even if their

intended usage is far more limited in nature – so end users are dissatisfied with NIH in the technical considerations domain. Data control preferences are partially satisfied by the NIH policy, because the centralized data control makes it easier to gain access to data; however, the NIH's control over the data means that end users must request access instead of being automatically granted it.

End users are most satisfied by Reg4All's policy, which was designed with clinical integration in mind. In the consent domain, they are able to see any data to which participants have granted them access; one of Reg4All's stated goals is to encourage the sharing of data with clinicians. In the technical domain, Reg4All is designed to allow easy access by end users, who are the primary audience of their data sharing. For data control, Reg4All provides a central location from which end users can request data, as well as streamlined processes for end users to get access.

Global Alliance tries to satisfy end users' preferences, but compromises between research and clinical needs. In consent and privacy, Global Alliance does envision clinical utilization in its consent procedures, but its proposed fine-grained consent may not allow clinicians to use data if participants do not consent to non-research purposes. By standardizing technical considerations, Global Alliance partially satisfies clinical users by providing them a way to access data and integrate it with clinical information, but does not provide a simplified mechanism for them to access data. Finally, control of data is decentralized under the Global Alliance policy, potentially reducing the ease with which end users can gain access and permission to use data.

Discussion

For each of these policies, different stakeholders were involved in the creation and development process. These differences in perspective led to policies that prioritize different aspects of genomic data sharing, reflecting the differing emphases and needs of the various stakeholder groups. Understanding how each of these policies does or does not satisfy stakeholder preferences enables a deeper discussion of the implications and future directions for genomic data sharing policies.

For stakeholders whose preferences are not being satisfied by the existing NIH policy, these results provide insight into how policies could be created that more closely match their preferences. The NIH needs to be aware of these emerging alternatives. If NIH fails to adapt to the preferences of multiple groups, then those groups may choose to “vote with their feet” and seek to participate in genomic research through mechanisms that do not require compliance with NIH policy that does not meet their needs. Selective nonparticipation is of particular concern for participants, many of whom are currently dissatisfied with several aspects of the NIH’s policy. This raises the concern that NIH’s research will not reflect the genetic makeup of the populations it seeks to serve. Because NIH is charged with both foundational genomic research and translational work designed to improve the health of populations, this raises real concerns about disparities in health and health care for populations that are not participating in NIH’s research.⁴⁴

The strength of the Reg4All policy model is that it gives exclusive control over the data to participants/patients. The purpose of the Reg4All policy model is to allow participants to take control of their own data and connect with researchers to encourage patient-centered research. The challenge of this policy model is that it does not easily scale to the widespread sharing that the NIH policy is intended to foster. The NIH policy focuses on creating a system of

administrative efficiency, in which genomic data is widely shared and frequently reused.

Reg4All focuses on the patient end of the spectrum of research; NIH focuses on the full gamut of research, from non-human genomes to clinical translation. If the Reg4All policy model is to be useful in a broader sense, it must find a way to satisfy the data producer and funder perspectives.

The Global Alliance has designed a system that addresses multiple stakeholder views and attempts to find compromise solutions to the diverse interests and needs. In doing so, it has fulfilled its mission to create a system which brings together as many groups to genomic data sharing as possible. Its primary challenge now is to find a way to operationalize its ideals and create functioning systems that implement its plans.

For NIH, these results highlight a failure to satisfy several important groups. The NIH has codified ethical principles for human subjects research that require the minimization of harm and the maximization of benefit. The NIH's current policy fails on both counts. By leaving many potential participants dissatisfied, NIH forces them to either accept a policy that does not meet their preferences or be excluded from the genomic research NIH is funding. For participants, participating without adequate protection means risking stigmatization and discrimination.⁴⁵

Other groups may also be harmed if they are forced to comply with the NIH policy despite its failure to meet their preferences. For data producers, this includes the potential harms of the unfunded mandate that they process and clean their data to be compliant with the NIH standards, even if doing so takes time and resources away from other research. For end users, the NIH policy's failure to satisfy their preferences for clinical usage means a barrier to translational science.

By creating such a limited set of choices, NIH's policy pressures dissatisfied groups to choose between one bad choice and another: either they risk the downsides of complying with

the NIH policy, or they risk the downsides of not being part of NIH's research enterprise. This does not represent a minimization of harm.

The NIH's policy also fails to maximize potential benefit. The NIH's original policies were designed to benefit a limited range of stakeholders, primarily organizations which were funding research for the human genome project and researchers who were working towards a single shared goal. Since that time, more stakeholder groups have become important to the process of genomic research, yet NIH's policy has failed to adapt to benefit those groups as well. Modern genomic research requires the buy-in of many groups; it cannot succeed in achieving all of NIH's goals without the willing cooperation of data producers, data users, participants, and end users.

The NIH can – and should – work harder to balance the preferences of these different groups. The other policy models demonstrate that it is possible to create a system that creates a more even distribution of satisfaction. As a public entity, NIH has an ethical obligation to serve multiple groups that is even greater than the obligation of voluntary groups such as Reg4All and the Global Alliance, because NIH's funding base is drawn (through taxes) from members of all of the groups.

No single solution will address all of the issues with the current policies. As the shortcomings of the Reg4All and Global Alliance policies demonstrate, there is no perfect solution. But by explicitly working to incorporate the concerns and suggestions found in the public comments, NIH can move closer to maximizing benefit and minimizing harm for all groups. This will require more than just an incremental adjustment to existing policies; it will require that NIH reconceptualize how it can incorporate more of these stakeholder groups into its policy development and oversight process. Reg4All and Global Alliance provide examples of the

types of dramatic redesigns NIH could consider to better meet the preferences of the stakeholder groups that were originally absent from NIH's policy development process.

These three policies collectively represent a critical point in the genomic data sharing discourse. As more stakeholder groups become increasingly vocal about their preferences about genomic data sharing, genomic data sharing policies will need to adapt in order to receive support from a broad base of stakeholder groups. None of these policies is the final word on how any of these organizations will treat genomic data sharing. There is a crucial window of opportunity in which NIH, Global Alliance, and Genetic Alliance can learn from one another and potentially alter their policies. As this analysis has shown, all three organizations share common ground in some areas. What remains to be seen is whether they can use their common ground as a basis to move forward together, rather than in three different directions.

Limitations

There are a number of limitations of this analysis, the foremost being the differential development and scope of influence of each of the three policies. The NIH Genomic Data Sharing Policy has been in development for over three decades and is supported by a wealth of history, experience and institutional processes that have been tested and refined over time. In contrast, Reg4All and the Global Alliance were launched into the public space approximately one year ago. Neither Reg4All nor the Global Alliance policy proposals have weathered nearly the same level of challenges and changing circumstances (including several significant technological advances in genomic sequencing capacity and cost reductions) as the NIH policy, which has been developed based on a history stretching back decades. It may be that neither will successfully live up to their ideals in the long term. Nonetheless, these two alternatives represent

emerging voices in the genomic data sharing discourse; although they lack the weight of history, their very development sheds light on areas in which the NIH policy is currently lacking.

The NIH Genomic Data Sharing Policy also has the strong force of the NIH's funding process behind it. This gives the NIH policy an outsized role in the discourse surrounding genomic data sharing, and makes it the most obvious target for criticism and discontent. Some of the weaknesses of the Reg4All and Global Alliance have likely not been revealed, because any hypothetical opponents would have less of an incentive to highlight weaknesses. If a group does not wish to participate in the Reg4All or Global Alliance policy proposals, then there are currently no financial repercussions of not joining a voluntary coalition. Without any significant consequences of nonparticipation, there is little to no impetus to protest against these policies or offer public critiques. The NIH policy, however, has real monetary implications for data producers who choose not to participate, by making it significantly more difficult (if not impossible, given the outsized role of NIH in genomic research funding) to obtain funding for genomic research. The NIH's role in translating genomic science into health policy also means that individuals or groups who choose not to participate in NIH-funded studies due to concerns about NIH's data sharing policies may also be excluded from future translational work. Because NIH's policy currently has a larger impact on research funding and longer-term health and health care effects, there is a greater incentive for critics to be vocal about their displeasure and offer detailed critiques.

Other limitations are related to the scope of this analysis, which only focused on policies relevant to U.S.-based stakeholders of genomic data sharing. Other countries have their own genomic data sharing models, which are not included in this analysis. In addition, by focusing on only the five stakeholder groups selected here (out of a myriad of possible perspectives), this

analysis limited exploration of the nuances inherent in this complex issue. This simplification process limits the extent to which the analysis can be extrapolated to complex stakeholders, relationships, and structures of real-world groups and organizations. Other possible stakeholder groups might include private companies, taxpayers, politicians, patients, ethnic and racial minority groups, universities, and others; these groups were not included in this analysis because their interests in each domain were not as clear-cut, and their roles in the genomic data sharing discourse are not as prominent. However, many of these groups are partially represented by one or more of the five stakeholder groups that were included.

Future Research

Further research should explore the ongoing development of NIH's 2013 Genomic Data Sharing Policy. NIH will be responding to public comments, as it has done with public comments on previous versions of the policy.⁴⁶ When the Global Alliance completes its process of soliciting and responding to additional comments on their proposed policy, it also would be worthwhile to examine how they change and respond to any criticisms or suggestions. The Global Alliance will need to codify and formalize their proposals into actionable policies that their member organizations can agree upon. The likely challenges of balancing the competing demands of stakeholders will be a rich area of further study, particularly as both the Genetic Alliance and NIH are currently partner organizations within the Global Alliance. It is unknown how the Genetic Alliance and NIH may influence the development of the Global Alliance's policy, nor how participation in the Global Alliance may affect NIH and Genetic Alliance. As Reg4All continues to grow and develop partnerships, further research into how their systems scale and how they interact with researchers and end users will reveal how sustainable and

practical their model truly is. Finally, future research should examine the development of international agreements and genomic data sharing policy models in other countries. Although outside the scope of this project, these agreements represent a rapidly-developing area within genomic data sharing, and will become even more important as genomic research is conducted in numerous countries and jurisdictions.

Appendix: List of Sources

	Year	Title	Description	Ref #
	1997	Summary of the Report of the Second International Strategy Meeting on Human Genome Sequencing	Bermuda Principles	8
	2002	NIH Announces Draft Statement on Sharing Research Data	Background on why NIH developed draft statement	47
	2003	NIH Data Sharing Policy and Implementation Guidance		
	2003	Final NIH Statement on Sharing Research Data	Minor changes to draft policy	46
	2003	Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects		14
NIH	2003	Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility	Fort Lauderdale Principles	13
	2007	Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)		15
	2009	Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data		16
	2013	Request for Information: Input on the Draft NIH Genomic Data Sharing Policy		10
	2013	Compiled Public Comments on Draft Genomic Data Sharing Policy		3

	Year	Title	Description	Ref #
Reg4All	2013	The haystack is made of needles	Article about Reg4All	48
	2013	Engaging Everyone in Research	Blog posting	49
	2013	New Disease Registry Gives Patients Some Privacy	MIT article	50
	2013	What is Reg4All? Reg4All and Researchers Reg4All and Individuals	Reg4All Webinars	51
	2014	Privacy Policy		52
	2014	More about us	Participant recruitment page	25
	Global Alliance	2013	Geneticists push for global data-sharing	Nature article
2013		Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data	Whitepaper	28
2013		Broad Institute	Announcement by member organization	54
2013		Global Alliance to Create Standards For Sharing Genomic Data	American Journal of Medical Genetics paper	55
2014		Priorities for the Regulatory and Ethics Working Group: 2014 and Beyond	Mission of the Regulatory and Ethics Working Group	56
2014		International Code of Conduct for Genomic and Health-Related Data Sharing,	May 5 th draft of Code of Conduct developed by Regulatory and Ethics Working Group	57

References

-
- ¹ Ohashi, J. “The power of genome-wide association studies of complex disease genes: statistical limitations of indirect approaches using SNP markers.” *J Hum Genet.* 2001;46(8):478-82.
- ² Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. “A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics.” *Am J Epidemiol.* 2000 Feb 15; 151(4):346-57.
- ³ National Institutes of Health. “Input on the Draft NIH Genomic Data Sharing Policy.” Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-018.html>. Accessed June 2014.
- ⁴ Pohlhaus J.R. Cook-Deegan R.M. Genomics research: world survey of public funding. *BMC Genomics.* 2008;9:472.
- ⁵ National Institutes of Health, National Human Genome Research Institute. “An Overview of the Human Genome Project: A Brief History of the Human Genome Project.” Available at <http://www.genome.gov/12011239>, accessed June 2014.
- ⁶ National Institutes of Health, National Human Genome Research Institute. “All About The Human Genome Project (HGP).” Available at <http://www.genome.gov/10001772>, accessed June 2014.
- ⁷ National Institutes of Health, National Human Genome Research Institute. “An Overview of the Human Genome Project: What was the Human Genome Project?” Available at <http://www.genome.gov/12011238>. Accessed June 2014.
- ⁸ Wellcome Trust. “Summary of the Report of the Second International Strategy Meeting on Human Genome Sequencing.” (“Bermuda Principles”) 1997. Available at <http://www.genome.gov/EdKit/pdfs/1997a.pdf>
- ⁹ McElheny, Victor K. *Drawing the map of life: inside the human genome project.* Basic Books, 2012.
- ¹⁰ National Institutes of Health. “Request for Information: Input on the Draft NIH Genomic Data Sharing Policy.” Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-13-119.html>. Accessed June 2014.
- ¹¹ National Institutes of Health. “NIH Policy Relating to Reporting and Distribution of Unique Biological Materials Produced with NIH Funding.” *NIH Guide for Grants and Contracts* Vol. 13, No. 4 March 30, 1984. Available at http://grants.nih.gov/grants/guide/historical/1984_03_30_Vol_13_No_04.pdf

-
- ¹² National Institutes of Health and Department of Energy. “NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources.” 1991. Available at <http://www.genome.gov/10000925> Accessed June 2014.
- ¹³ Wellcome Trust. “Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility.” (“Fort Lauderdale Principles”) 2003. Available at <http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf> Accessed June 2014
- ¹⁴ National Institutes of Health, National Human Genome Research Institute. “Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects.” February 2003. Available at <https://www.genome.gov/10506537> Accessed June 2014
- ¹⁵ National Institutes of Health, National Human Genome Research Institute. “Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).” 28 August 2007. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html> Accessed June 2014
- ¹⁶ National Institutes of Health, National Human Genome Research Institute. “Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data.” 19 October 2009. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-HG-10-006.html> Accessed June 2014
- ¹⁷ Race and ethnicity in the genome era: The complexity of the constructs. Bonham, Vence L.; Warshauer-Baker, Esther; Collins, Francis S. *American Psychologist*, Vol 60(1), Jan 2005, 9-15. doi: [10.1037/0003-066X.60.1.9](https://doi.org/10.1037/0003-066X.60.1.9)
- ¹⁸ Tekola, Fasil, et al. "Impact of social stigma on the process of obtaining informed consent for genetic research on podoconiosis: a qualitative study." *BMC medical ethics* 10.1 (2009): 13.
- ¹⁹ Zerhouni, Elias A. "Translational and clinical science—time for a new vision." *New England Journal of Medicine* 353.15 (2005): 1621-1623.
- ²⁰ Rodriguez, L. L., et al. (2013). "Research ethics. The complexities of genomic identifiability." *Science*, 339(6117): 275-276.
- ²¹ Gymrek, M., et al. (2013). "Identifying personal genomes by surname inference." *Science*, 339(6117): 321-324.
- ²² Foster, Morris W., Deborah Bernsten, and Thomas H. Carter. "A model agreement for genetic research in socially identifiable populations." *The American Journal of Human Genetics* 63.3 (1998): 696-702.
- ²³ Hensley Alford, S., et al. "Participation in genetic testing research varies by social group." *Public Health Genomics* 14.2 (2010): 85-93.

-
- ²⁴ Walker, Lorelei, et al. "dbGaP data access requests: A call for greater transparency." *Science translational medicine* 3.113 (2011): 113cm34-113cm34.
- ²⁵ Reg4All Website. "More About Us." <https://www.reg4all.org/more.php> Accessed June 2014
- ²⁶ Genetic Alliance. "About Us." <http://www.geneticalliance.org/about> Accessed June 2014
- ²⁷ Global Alliance for Genomics and Health. "Partners." Available at <http://genomicsandhealth.org/partners> Updated June 12 2014
- ²⁸ Global Alliance for Genomics and Health. "Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data." June 2013 Available at <https://www.ebi.ac.uk/about/news/press-releases/Global-Alliance>
- ²⁹ Global Alliance for Genomics and Health. "History." Available at <http://genomicsandhealth.org/about-global-alliance/history>
- ³⁰ Global Alliance for Genomics and Health. "Working Groups." Available at <http://genomicsandhealth.org/our-work/working-groups>
- ³¹ Wood and Kroger, "Doing Discourse Analysis: Methods for Studying Action in Talk and Text." SAGE 2000. P4
- ³² McCloskey, Rose. "A guide to discourse analysis." *Nurse Researcher* 2008, 16, 1
- ³³ LeGreco M, Tracy SJ. Discourse tracing as qualitative practice. *Qualitative Inquiry*. 2009;15(9):1516-1543.
- ³⁴ National Institutes of Health. "Request for Information: Input on the Draft NIH Genomic Data Sharing Policy." Section IV C 4. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-13-119.html>. Accessed June 2014.
- ³⁵ Reg4All. "Privacy Policy – Individual Choice." Update 4 April 2013. Available at <https://www.reg4all.org/privacypolicy.php>
- ³⁶ Global Alliance for Genomics and Health. "Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data - Protection of Privacy and Participant-Centered Initiatives." Page 13. June 2013 Available at <https://www.ebi.ac.uk/about/news/press-releases/Global-Alliance>
- ³⁷ National Institutes of Health. "Request for Information: Input on the Draft NIH Genomic Data Sharing Policy." Appendix A. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-13-119.html>. Accessed June 2014.
- ³⁸ Reg4All. "Privacy Policy – Collection and Use." Update 4 April 2013. Available at <https://www.reg4all.org/privacypolicy.php>

-
- ³⁹ Global Alliance for Genomics and Health. “Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data – Relationships with operating entities.” Page 25 June 2013 Available at <https://www.ebi.ac.uk/about/news/press-releases/Global-Alliance>
- ⁴⁰ National Institutes of Health. “Request for Information: Input on the Draft NIH Genomic Data Sharing Policy.” Section V A. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-13-119.html>. Accessed June 2014.
- ⁴¹ Genetic Alliance. “Input on the Draft NIH Genomic Data Sharing Policy – Genetic Alliance.” Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-018.html>. Accessed June 2014.
- ⁴² Global Alliance for Genomics and Health. “Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data – A path forward.” Page 10. June 2013 Available at <https://www.ebi.ac.uk/about/news/press-releases/Global-Alliance>
- ⁴³ Sharon Terry. “Changemakers: Reg4All.” Available at <http://www.changemakers.com/healthbiz/entries/reg4all> Accessed June 2014
- ⁴⁴ Haga, Susanne B. "Impact of limited population diversity of genome-wide association studies." *Genetics in Medicine* 12.2 (2010): 81-84.
- ⁴⁵ Davis, Dena S. "Groups, communities, and contested identities in genetic research." *Hastings Center Report* 30.6 (2000): 38-45.
- ⁴⁶ National Institutes of Health. “Final NIH Statement on Sharing Research Data.” 26 February 2003. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- ⁴⁷ National Institutes of Health. “NIH Announces Draft Statement on Sharing Research Data.” 1 March 2002. Available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html>
- ⁴⁸ Terry, Sharon F., et al. "The haystack is made of needles." *Genetic testing and molecular biomarkers* 17.3 (2013): 175-177.
- ⁴⁹ Genetic Alliance. “Engaging Everyone in Research.” 15 August 2013. Available at <http://blog.reg4all.org/engaging-everyone-in-research.html>
- ⁵⁰ Humphries, Courtney. “New Disease Registry Gives Patients Some Privacy.” *MIT Technology Review*. 14 March 2013. Available at <http://www.technologyreview.com/news/512456/new-disease-registry-gives-patients-some-privacy/>
- ⁵¹ Genetic Alliance. Webinars. Available at <http://www.geneticalliance.org/events/webinars/reg4allseries> Accessed June 2014
- ⁵² Reg4All. “Privacy Policy.” Update 4 April 2013. Available at <https://www.reg4all.org/privacypolicy.php>

-
- ⁵³ Hayden, Erika Check. "Geneticists push for global data-sharing." *Nature* 498.7452 (2013): 16.
- ⁵⁴ Broad Institute. "International partners describe global alliance to enable secure sharing of genomic and clinical data." 4 June 2013. Available at <https://www.broadinstitute.org/news/globalalliance> Accessed June 2014
- ⁵⁵ Altshuler, David. (2013), Global Alliance to Create Standards For Sharing Genomic Data. *Am. J. Med. Genet.*, 161: xi. doi: 10.1002/ajmg.a.36168
- ⁵⁶ Global Alliance Regulatory and Ethics Working Group. "Priorities for the Regulatory and Ethics Working Group: 2014 and Beyond." Available at <https://genomicsandhealth.org/files/public/REWG%20Priorities%202014%2006%2012%20FINAL%20FOR%20POSTING.pdf> Accessed June 2014
- ⁵⁷ Global Alliance Regulatory and Ethics Working Group. "International Code of Conduct for Genomic and Health-Related Data Sharing – Draft #6." Available at <http://genomicsandhealth.org/our-work/work-products/international-code-conduct-genomic-and-health-related-data-sharing-draft-6> Accessed June 2014