Published in final edited form as:

Hastings Cent Rep. 2016 ; 46(1): 36–45. doi:10.1002/hast.523.

Balancing Benefits and Risks of Immortal Data: Participants' Views of Open Consent in the Personal Genome Project

Oscar A. Zarate [doctoral student in statistics],

Northwestern University.

Julia Green Brody [senior scientist and the executive director],

Silent Spring Institute, a nonprofit research group focused on environmental chemicals and health.

Phil Brown [teaches sociology and health sciences and directs],

Social Science Environmental Health Research Institute at Northeastern University.

Mónica D. Ramírez-Andreotta [assistant professor of soil, water, and environmental science],

Mel and Enid Zuckerman College of Public Health at the University of Arizona.

Laura Perovich [doctoral student], and

Massachusetts Institute of Technology Media Lab and a former research assistant of the Silent Spring Institute.

Jacob Matz [doctoral student]

Social Science Environmental Health Research Institute and the Department of Sociology and Anthropology at Northeastern University.

Abstract

The NIH Genomic Data Sharing Policy, effective in January 2015, encourages researchers to obtain broad consent to share data for unspecified biomedical research. The ethics of extensive data sharing depend in part on study participants' understanding of the risks and benefits. Interviews with participants in the Personal Genome Project show that study participants can readily discuss the risks, including loss of privacy, and are willing to accept risks because they value the opportunity to contribute to health science. They have expansive views of the benefits for science, medicine, and their own health and curiosity. With justice in mind, further exploration is needed to evaluate consent for data sharing among more diverse and vulnerable populations.

Background

In the digital era, an individual's health, genetic, or environmental exposure data, placed in an online repository, creates a valuable shared resource that can accelerate biomedical research and even open opportunities for citizen science to crowd-source discovery. But these data become "immortalized" in ways that may create lasting risk as well as benefit,

and they hold the potential to violate privacy and informed consent. Once shared on the Internet, data are difficult or impossible to redact, and identities may be revealed by matching online data sets to each other, a process called data linkage. Re-identification (re-ID), the process of associating an individual's name with data that were considered de-identified, poses risks such as insurance or employment discrimination, social stigma, and breach of the promises often made in informed consent documents. At the same time, re-ID also poses risks to researchers and the future of science if re-ID undermines trust and participation.

The ethical challenges of online data sharing are heightened as “big data” becomes an increasingly important research tool and driver of new research structures. “Big data” is shifting research to include large numbers of researchers and institutions as well as large numbers of participants providing diverse types of data, so the participants' consent relationship is no longer with a person or even a research institution. In addition, consent is further transformed, because “big data” analysis often begins with descriptive inquiry and hypothesis-generation, and the research questions cannot be clearly defined at the outset and may be unforeseeable over the long-term.

In this article, we consider how expanded data sharing poses new challenges, illustrated by genomics and the transition to new models of consent. We draw on the experiences of participants in an open data platform – the Personal Genome Project (PGP) -- to allow study participants to contribute their voices to inform ethical consent practices and protocol reviews for big data research.

Evolution of Data Sharing Consent in Genomics and other Fields

Genomics research offers an example of these issues, because the data are uniquely personal, but the scientific need for large and expensive datasets, aggregated across multiple studies, motivates the creation of open online archives. As early as 1996, the international Human Genome Project codified its commitment to public data sharing in the Bermuda Principles, and the National Institutes of Health later required sharing from genome-wide association studies (GWAS) and created a repository now known as the database for Genotypes and Phenotypes (dbGaP), which includes approximately 300 studies accessed by 2,200 investigators.¹ These early practices assumed that genomic data alone would be difficult to link to an individual. However, since then, researchers have demonstrated techniques that can identify individuals with known genomes in aggregated datasets and re-ID certain participants in the 1000 Genomes Project using genomic and demographic data.²

In response to these and other discoveries, NIH issued a new Genomic Data Sharing (GDS) policy in January 2015 that encourages researchers to explicitly obtain informed consent for broad data sharing for biomedical research.³ The new policy acknowledges that removing overt identifiers, including the 18 types of data specified by the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, is inadequate to prevent re-ID. It allows an option for some data to remain under controlled access under special circumstances, prohibits researchers from using data for re-ID, and states that breaches of researcher responsibilities will be penalized. It calls for participants to share data with unnamed researchers for unspecified purposes over the indefinite future. This broad consent approach

differs ethically from typical consent practices in the past,⁴ and NIH has provided guidance on the language and models that are needed to implement the new policy,⁵ but these proposed practices have yet to be widely adopted and tested.

New informed consent models are similarly needed for other types of health and social research data, since a range of information is vulnerable to data linkage, leading to re-ID. These vulnerabilities blur the definition of personally identifiable data, so investigators across the spectrum of human research need to rethink the distinction between identified data, which are protected by the Common Rule, and de-identified data, which are not considered human subjects research. Participants often do not understand the definition and implications of de-identification and may be unaware that de-identified data may be widely shared, so better explanations are especially needed now that re-ID risks have increased.⁶

In addition to the GDS policy on broad consent, recent models of consent that begin to address these issues range from open consent, in which participants agree to online sharing of data with no protections for privacy, to dynamic consent in which participants may interact repeatedly with researchers, releasing specific data for specific analyses.⁷ Proponents of dynamic consent argue that participants should not be required to anticipate all future uses of their data and should have an opportunity to opt out as new uses arise. Other options include disaggregated checklists of choices in an initial consent document. Portable Legal Consent is a form of broad consent in which participants control their data and donate it to a repository.⁸ Another novel model is openPDS, which creates a personal data store that allows an individual to share only the data needed to answer a specific query.⁹

The ethics of these options depends in part on how well participants understand them. To meet the Common Rule standard of “autonomy,” participants must be able to understand the procedures and evaluate potential risks and benefits. To satisfy “justice,” new models of informed consent must not raise barriers to the participation of diverse, representative populations in research, a potential problem if informed consent language is too complex.

To guide new forms of consent, researchers and institutional review boards (IRBs) can benefit from understanding the experiences of people who volunteer for studies with extensive data sharing and novel forms of consent. To investigate how study participants view choices about data sharing and privacy, we interviewed participants in the Personal Genome Project, a research setting where individuals consent to have their genome and self-posted health and social data in an open access database with the explicit statement that privacy is not assured. Our key point is that study participants’ perspectives can contribute to empirically-informed decisions about ethical and effective practices by revealing what participants care about, why they decide to share data, and what their expectations are regarding scientific impact, other benefits, and risk. PGP is a unique setting to investigate these questions, because participants go through a formal training process about privacy risks and then consent to unrestricted data sharing.

The Personal Genome Project

The Personal Genome Project is an international large-scale genomic sequencing and biobanking project founded at Harvard Medical School in 2005. It aims to sequence the

genomes of 100,000 volunteers and make their genetic and trait information available as a resource for scientific discovery and public use. Online profiles include genomic data, information on physical traits, and disease history, sometimes including medical records, psychiatric information, and behavioral information. Participants are not required to include their names, but they have the option to do so, and the first 10 participants prominently displayed their names and photographs. Biological samples are stored in biobanks and can be used for unspecified purposes. At the end of 2013, over 3,000 participants were enrolled in the PGP and 163 had their genomes sequenced through the project.¹⁰

PGP researchers developed open consent, a relatively new model of informed consent, because they believe that the identifiability of genomic and other health data means that traditional promises of research privacy do not meet the principle of “veracity – telling the truth.”¹¹ Open consent makes no assurances of privacy or anonymity. Rather, it informs participants about the risks and benefits of unrestricted access to data and specimens, including the possibility that risks cannot be fully anticipated. The consent process includes an eligibility questionnaire, an online training program, and a multiple choice entrance test of about 20 questions that must be completed with a perfect score. Candidates are encouraged to discuss their participation with family members, since genetic results have implications for them. Initially, the PGP restricted the study to people with graduate degrees or equivalent experience in a further effort to ensure that participants understand the risks. As an emblem of the open research process, one early participant put the informed consent session on YouTube.¹²

Although the PGP entrance test and extensive discussion of risks are innovative, the permissions requested -- to share data indefinitely with unspecified researchers for unspecified purposes -- are fundamentally the same as envisioned by the GDS Policy, so PGP experiences can directly inform implementation of the new policy. As background for interpreting our interviews with PGP participants, we briefly review here the risks and benefits described by PGP in its consent and online guide.

The PGP highlights a range of risks, including downstream consequences of re-ID as well as potential drawbacks of receiving personal results. Participants are specifically alerted that their data may become publicly associated with their name. Other risks include the possibility that genomic findings may be unsettling -- for example learning about disease risks or future uses of samples, such as for cloning. Possibilities for social complications, such as learning about paternity or racial history, are discussed. PGP clarifies that results are not intended for clinical or medical use.¹³ Risks identified in the informed consent are summarized in Table 1.

PGP documents also discuss benefits for society and for personal education. Open-access is framed as “critical to scientific progress,” and the PGP website uses scientific imagery expressing the idea of a “frontier” in many ways, for example:

Like other areas in human history, personal genomics will likely benefit greatly from ‘early adopters’ who are willing and able to endure the difficulties and uncertainties that go along with exploring relatively uncharted territories.¹⁴

The website describes PGP as “a public resource” freely available for all: “We believe sharing is good for science and society. The PGP is dedicated to creating public resources that everyone can access.”¹⁵ It anticipates contributions to human health and medical advances for future generations. Though the focus is on societal benefits, PGP adds that participants have the opportunity to learn about their own genomes and about scientific research. Participants are offered a sense of community and pride, and PGP writes of the relationships of reciprocity and mutual engagement that they wish to cultivate between researchers and participants through conferences and online forums.

Our interviews with PGP participants allow us to compare their experiences, attitudes, and values with the stated values of the study and to better understand the decision to enroll. This study represents the first systematic investigation, as far as we are aware, of participants’ views on open consent and is the first to explore experiences of participants within the PGP. Our data can contribute to a fuller understanding of the processes of informed consent and the risks and benefits associated with large genomic health studies overall.

Methods

Participant Recruitment

We recruited 34 PGP participants for in-person interviews at the 2013 Genomes Environments Traits (GET) Conference in Boston, MA, an annual conference hosted by PersonalGenomes.org, a nonprofit organization that supports the PGP. Approximately 150 PGP participants attended.¹⁶ The conference included expert presentations, networking breaks, and “labs.” Labs allowed approved researchers to recruit attendees online and at booths set up at the conference; our research project was one of these labs. Approval was obtained from the Northeastern University IRB. Our team also hosted a related ‘lab’ led by Latanya Sweeney and the Harvard Data Privacy Lab that allowed attendees to estimate the likelihood that combinations of their birthdate, ZIP code, and gender could be used to re-identify them from voter lists.¹⁷

Interviews

Trained staff conducted interviews during the two-day conference. The semi-structured interview included questions and prompts about experiences with the consent process, views on sharing of research data, discussions with family members, risks and benefits of participation, and re-ID specifically. Questions addressed motivations for participation and expectations about outcomes. Interviews were digitally recorded, with the exception of one participant who requested written notes instead. Interviews were typically about 20 minutes long and ranged from 6 to 42 minutes.

Analysis

Audio recordings were transcribed and imported into NVivo 10, a qualitative data analysis program. Names and overt identifiers were removed. Initial codes were developed based on our research goals and the interview questions. Four members of the research team reviewed portions of interview transcripts to build a more detailed list of codes. We discussed key themes and developed definitions for when to apply the codes. Two members of the team

then coded portions of the transcripts separately, met to compare coding, further refined the codebook, and worked toward agreement about definitions of codes. Once they were in general agreement with each other, the rest of the interview transcripts were coded and analyzed. All uncited quotes come from the interviews.

Results and Discussion

Interviews provide insights into participants' beliefs about the benefits and risks of participating in an open data project that shares information about their genome, health, and social history. Interviews also cover perspectives on the future of genetic science, personalized medicine, and research ethics. We focus on whether participants understand risks, which is crucial to evaluating open consent as an ethical practice.

Participant Characteristics

Among 34 PGP participants we interviewed, 20 (59%) were male. All but two identified themselves as white; the others were Asian or multiracial. Age ranged from twenties to sixties; the largest group was in their thirties. About half were from the Northeast U.S. Participants were well educated and many had advanced degrees, consistent with PGP eligibility and recruiting practices. Sixteen already had their genomes sequenced; others were waiting due to cost and staffing limits. A limitation of our study is that participants were recruited at a PGP conference, so they may be more engaged and better informed than others in the study. Because participants attend the GET Conference in order to contribute biological samples, our interviews may over-represent more recent participants who hadn't already contributed samples in earlier years. In addition, we had access only to people who ultimately joined PGP and not to those who began the consent process and then declined to share their data. PGP has reported that 10% of people who pass the entrance exam do not enter the study.¹⁸

Motivations and Perceived Benefits for Joining PGP

PGP participants expressed an expansive sense of excitement about participating in a study that they believe will accelerate biomedical science and medicine by creating a large, open data resource to link genetics and health. When asked why they joined the PGP, participants generally focused first on social benefits by advancing science and medicine and, secondarily, on learning about themselves, often with a hope of learning about personal or family disease. They also mentioned scientific curiosity, hopes for social change, and the rewards of being part of a pioneering community.

Nearly all participants (n=29) spoke of their motivation to help others by contributing to science and medicine. For example, one participant explained, "A blood donation generally helps one person. The PGP project is helping an entire body of knowledge." Many expected to advance personalized medicine: "I think personalized medicine based [on the] genome will come here sooner or later. I'd rather see it sooner, so that's why I joined, so I can help." Others had hopes to assist in research on specific issues, for example: "there's a lot of mental disease in my family...maybe I'm helping...maybe next time I'll have a sister who's just a normal happy kid." Consistent with the overall themes of altruism, some described

participation as a gift: “I consider that I’ve donated my genome to science instead of my body, and I don’t even have to wait until I’m dead.”

Many expected to learn something relevant to their own health or personal attributes, including migraines, blood pressure, learning disabilities, sexuality, and unexplained deaths in their families (n = 16). Because individual results are online, participants can see them, and some joined PGP specifically to gain free access to their genome.

While perceived benefits of contributing to medicine and personal health are found in other genetic or medical studies, the PGP participants have a more independent, proactive, and futuristic spirit. They often planned to interpret and act on their data themselves, independently of the PGP research team: “The stuff [genetic results] that I’ll get back that I’ll be able to look at through GET-Evidence¹⁹ and tools like that will inform my health decisions.” They expect the personal utility of results to increase as science progresses. Many expressed a deep curiosity about themselves, for example, “We take all kinds of different steps in life as people to know ourselves, but this is a whole different way of being able to know yourself, perhaps on more of a cellular level.” For most participants (n = 23), learning new science was a primary reason to join PGP.

The interviews reflect a rich sense of participating in an active community of people who are generating knowledge and also social change. Participants “feel like they got to be among the first group of astronauts going to another planet.” They felt a strong sense of community with one another and with other aspects of citizen science and quantitative self-tracking.²⁰ For example, participants said:

“It feels really cool to kind of be connected to this community of a lot of do-it-yourself researchers, but also a larger community of people who are just willing to give their data...”

“It’s been fun and rewarding, and I feel like I found my peeps.”

Open consent contributed to the sense of pioneering contribution. Participants mentioned the value of creating a large dataset that could be used repeatedly by multiple researchers around the world as a way to encourage discovery and lower the cost. Some had hopes that by making their own genetic data and information about stigmatized conditions public, they would raise awareness in ways that would normalize and de-stigmatize people like themselves in the future.

Our findings shed new light on altruism in research, since participants are intentionally pioneering data sharing as a novel opportunity that many believe will contribute significantly to scientific discovery. These perspectives include elements of health information altruism²¹ and the extended concept of “research altruism.” Research altruism aims not only to contribute generally to research but also to provide family members and communities with vital information and empowerment. This expectation has been observed, for example, in environmental biomonitoring studies in which participants hope to learn about local pollution, so it can be remedied.²² Research altruism encompasses the willingness of people to participate in research when they see themselves as part of a collectivity, joining in a shared venture of participants partnering with a trusted research organization to benefit a

group that is important to them. This concept is an example of “generalized exchange,” in which the beneficiaries are numerous and benefits may be in the future.²³

Our results encourage ethics reviewers to consider how research protocols match up with participants’ desire to create benefits from data sharing. In open consent and broad consent, the participant contributes data as a gift that may be irreversible, and others, including the IRBs, have ongoing responsibility to ensure that the initial promises, whether explicit or implicit, can be sustained. The overarching questions about prospects for scientific breakthrough may be un-answerable and beyond the scope of the IRB, but short-term practices can influence the utility of data. For example, if researchers promise accelerated scientific discovery from data sharing, they have responsibility at a practical level to ensure that data are promptly cleaned, machine readable, and documented to facilitate analysis by others, as described, for example, in the White House open data project²⁴ and G8 Open Data Charter.²⁵ Meeting these standards will require new infrastructure for studies like the PGP in which participants and laboratories are uploading data from diverse sources that may not be uniform. Over the longer term, researchers will need to hold themselves accountable for translating and disseminating results into health and social benefits, or the trust of participants will be lost. Ethics review can encourage researchers to think about these down-the-road issues and allocate resources to communicate with participants and the public about what has been achieved.

Perception of Risks

While PGP participants were enthusiastic about benefits of data sharing, they also were aware of risks. We asked participants about risks to themselves, their family or their community that stem from health data disclosure. Responses frequently mentioned the risks described in the PGP informed consent and website. The concordance between participants’ responses and the informed consent is illustrated in Table 1.

The most salient risk for participants was the potential for discrimination in employment, health insurance, or other insurance resulting from information about a past illness or genetic predisposition (n = 29). In addition, almost 40% (n=13) mentioned the possibility of sensitive information -- for example about mental illness, drugs, or sexually transmitted disease -- reaching social contacts, causing other people to treat them differently. Participants were aware that getting their genome sequenced might reveal distressing information about health risks, for example a predisposition to Alzheimer’s disease or cancer, but all said they would rather know than avoid bad news. A few mentioned that open data made family members uncomfortable. Participants were readily able to describe futuristic risks, such as cloning or malicious use of DNA, but regarded them as unimportant. Some were concerned that by making their data and biological samples available for science to benefit society, they might also be used commercially for private gain (n = 4).

The sense of being a pioneer came up in discussions of risks as well as benefits, for example, in this participant’s expectation that the discrimination risks for early participants will be resolved: “[you might have] people getting fired for having a chance to develop cancer...but, I think after those first few cases, then there’ll be a lot of thinking about how to mitigate that, and then it won’t be so bad.” The Affordable Care Act provision to insure

people with pre-existing medical conditions is an example of social change consistent with this expectation, but this change had not been fully implemented at the time of our interviews. Some participants mentioned the Genetic Information Nondiscrimination Act of 2008 as a relatively new, but limited, protection.

Most often, PGP participants considered the risks as possible but not likely to be a problem for them personally. Specifically, they recognized the potential for re-ID but thought that others are “not that interested.” For example, one said, “I don’t think anybody looking at hiring me would go look to see.” Others said they felt less vulnerable to the risks because they were older, wealthy, or childless, or they were more able or willing to overcome possible risks through legal remedies. Comments like these indicate that participants are weighing their own circumstances and values, consistent with the principle of autonomy. At the same time, the thought process reveals how open data can pose barriers to justice, because some groups are more vulnerable to risks, such as employment discrimination, or less able to overcome them.

While most risks were discussed hypothetically, three participants described real situations that made them uncomfortable. Two participants were criticized by a relative or colleague for joining PGP: “when they found out I was part of the project, they said, ‘Well, that’s just stupid.’ So I’m, ‘Oh, really?’” One participant whose parents were also in the PGP described a period of psychological stress after learning his parents were heterozygous for a gene related to Alzheimer’s and before learning he did not carry the gene. In the most striking incident, PGP participants at the GET conference were approached by another attendee, a journalist, who interviewed them about views on re-ID and then asked for names. A participant we interviewed spoke with the reporter and then was surprised when asked for a name for attribution. This participant complied but later regretted the hasty response.

Data Sharing and Re-ID

The risks of discrimination or social stigma from open data are downstream consequences of re-identification, so we probed specifically about re-ID. All 34 participants said they understood that their identity could be associated with their online data, although opinions differed on the likelihood of re-ID. Many felt that re-ID will become easier and “eventually be trivial”:

I think we have protections for privacy, but we have no assurance of privacy and I think understanding that is important to be involved in this kind of research and certainly with the de-identification findings from the past year and then re-identification findings, I think any real expectation of genomic privacy is a bet, at best.

When asked if they would still participate in the PGP if they knew they would be re-identified, all participants said they would. However, some (n=6) told us that they decided to edit their profiles to reduce identifiability, for example, by deleting their name, which may have been inadvertently imported with data from external medical records, or by removing some digits of their postal zip code.

Our interviews show that this segment of self-selected Americans, at least, is willing to share data even when they acknowledge the risk of re-ID, and it would be helpful in implementing broad consent to know whether a wider public would agree. In contrast to the PGP consent process, which draws attention to re-ID, the model language suggested by NIH for broad consent under the GDS policy briefly characterizes the risk of re-ID as “unlikely.”²⁶ This estimate is open to debate and likely to change, so it would be useful to investigate whether researchers could be more forthcoming about re-ID without compromising participation. At the same time, we worry that PGP’s detailed discussion of cloning and hypothetical use of DNA at a crime scene – or even re-ID from genetic data – distracts attention from the easy-to-implement, more likely possibility of re-ID from medical records data and ZIP code.

Discussions with Family

Because genomic information about an individual reveals data about families as well, the PGP encourages participants to discuss their participation with family members before they join.²⁷ Most participants (n = 31) did have this discussion, although often it was after they enrolled:

It wasn’t really a “should I do this” debate. Just a “here’s what it is.” Here’s what the risks are associated with it and I want you to know that I’m doing this.

Family members were sometimes supportive, including some who decided to join PGP, but others were concerned about the potential implications for themselves:

I have a sister and a brother who are sort of concerned. They operate under the – they’d rather not know what they don’t know. So to some extent they give me a bit of a hard time about participating in this.

These conversations between participants and their families raise an emerging ethical challenge for widely shared data, since current practices allow individuals to take on risks that affect others as well as themselves.

A few participants noted strengthened family connections because of a shared interest in PGP and positive experiences of sharing health-related information. For example, a participant spoke about learning from a brother’s profile:

I... clicked on him [my brother’s profile] and went, ‘Oh. Look at that. We both have that.’ Went and talked about that... because we actually both have mild IBS, so not really talking about our digestive systems to each other... I guess that’s the odd precipitating event out of the PGP for me so far.

In general, human research has not solved ethical problems that arise when the autonomous decision of an individual to participate in research has consequences for others. The PGP practice of recommending a discussion with family members is a good step forward and a reasonable one, given that the risks are similar to a family member personally disclosing private information outside of research.

Limitations for Diverse Populations

PGP participants are predominantly well-educated and white, more male than female (about 65% male), and engaged in science and technology. However unintentionally, the PGP model, with an extensive tutorial and screening test, tends to over-represent these groups. This screening strategy makes sense as a starting point for novel research with risks that are difficult to foresee when privacy threats and genomics knowledge are evolving rapidly. However, health research cannot ethically exclude less-educated or vulnerable populations in the long run, particularly as data sharing becomes essential and ubiquitous.

Two types of barriers to diversity need to be addressed: the potential for some groups to face greater risks of economic harm or discrimination from privacy loss, and the challenges in explaining risks in order to obtain valid consent. Both of these issues were raised in the PGP interviews. PGP participants identified younger people, who have their careers ahead of them, and parents, whose children might be affected by data disclosure, as groups that might be less able to participate in open data projects. Individuals who are already vulnerable to discrimination, for example because of their race, ethnicity, or gender identity, might also be hesitant to increase their risk. In addition, PGP participants were concerned that people with less education or poor mental capacity would not be able to navigate the PGP consent process and, thus, would be barred from the benefits of the study. New strategies to resolve these problems will serve the goals of beneficence and justice.

Conclusion

Our interviews with Personal Genome Project participants are the first, to our knowledge, with participants in an open consent study and provide insight into the decision-making process about consent to widely share personal research data for unspecified purposes. PGP experiences can provide empirical input to discussions about new forms of data sharing and consent. The PGP participants are carefully screened, more-educated than the general population, and likely to differ in other ways, but we can think of them as representing an optimistic scenario for obtaining valid consent and, thus, a starting point from which data sharing practices can be developed to responsibly inform people about risks and benefits.

We found that the participants in the PGP were readily able to discuss the risks of sharing health and genomic data online and describe how they evaluated these risks and balanced them against benefits. They regarded re-identification as technically feasible but thought others were unlikely to be motivated to re-ID them. Those who were older, childless, self-employed, or wealthy felt less vulnerable to harms. Employment and insurance discrimination were considered the most likely risks, but participants felt ready and able to seek legal remedies. Futuristic risks, like cloning, were frequently mentioned but not with alarm. Benefits were discussed with considerable excitement. Participants envisioned their data contributing to advances in personalized medicine to benefit society as a whole. They viewed their own data with curiosity, and some anticipated personal benefits from learning about genetic risks, so they could target preventive behaviors and tailor medical decisions. Without exception, interviewees stated that they would participate in PGP even if they knew they would be re-identified, though several removed data to reduce their vulnerability. From the rich discussion of benefits and risks, we conclude that these select individuals were well-

informed and ethically consented to forego their research privacy and share extensive personal data online. Thus, our results support the PGP model of online training and an entry quiz as strategies for informed consent in some circumstances.

While our interviews show PGP participants understand the risks that were communicated during consent, our interviews also highlight participants' hopes and trust that data sharing will significantly advance science and health, so evaluating how researchers shape these expectations and develop concrete practices to meet them becomes an important additional, inter-related ethics issue. This issue of promises on the benefits side is particularly relevant for consent to share data indefinitely and for unspecified purposes, because it requires a high degree of long-term trust and accountability.

As we enter the "big data" era, it is important to consider what aspects of the PGP model can be broadly carried forward in other studies. For example, the PGP tutorial is not an appropriate format for participants who lack experience with academic culture, but the central content describing re-ID is widely relevant. The GDS policy prohibits data users from attempting to re-ID people, but the privacy risks of broad consent may, in practice, substantially or completely overlap with open consent. As data sets become larger and include nearly everyone, research staff have increasing opportunities to stumble upon or intentionally search for people they know or public figures, and as more data are shared online, security breaches lead to significant privacy risks. The current ability to discover and police privacy violations is limited. Thus, the descriptions of risk developed for open consent are useful models to adapt for broad consent. For the future, the GDS policy notes that re-ID experiments by researchers may be allowed in narrow circumstances, and we hope that this type of research will be developed to empirically evaluate privacy threats and solutions, and to inform the description of risk in broad consent.

The research community is at an early stage in developing new ethical practices in a "big data" context. Our results lead us to key recommendations for next steps in broad data sharing:

1. The re-ID risks should be described with clarity that goes beyond characterizing re-ID as "unlikely" and describes common data linkage vulnerabilities and their possible risks. PGP participants were not concerned about futuristic scenarios, but a not-negligible fraction of the group experienced social gossip or family tension, so these possibilities should be described. We are reassured by our interviews that researchers will be able to be forthcoming about risks without dampening participation, and greater transparency may benefit research by improving trust in the long run.
2. IRBs should ensure that structures are in place to realize the social benefits from data sharing, since these are the motivators for participants to risk their privacy. At minimum, IRBs should hold researchers accountable for meeting technical standards that make data usable when it is shared.
3. New studies should be undertaken to inform consent for data sharing in diverse and more vulnerable populations. We hypothesize that diverse populations will be able

to understand privacy risks if they are appropriately described and will be willing to contribute data to benefit health science, but this hypothesis needs to be tested.

Beyond the relevance to broad consent, the PGP begins to connect academic and laboratory-based technical science with emerging citizen science, because it envisions open consent as part of an engaged collaboration between biomedical researchers and citizen scientists with iterative communication, for example through the GET Conference, GET-Evidence website, and online blogs and forums, and our interviews show that participants do view themselves as part of a proactive community. Some are using personal sensors to collect and upload measurements about themselves. This context adds a further challenge for IRBs to rethink the boundaries between participants and researchers, and calls on IRBs to consider “reflexive research ethics,” the self-conscious, interactive, and iterative reflection upon researchers’ relationships with research participants.²⁸ In the digital era, as personal information becomes more readily available, research data becomes self-collected, and the expectations and reality of privacy evolve, data sharing and consent practices will also need to evolve. The best practices will allow volunteers to participate in accelerating science through data sharing, protect their well-being, remain truthful about risks and benefits, and maintain their trust.

References

1. Paltoo DN. Data Use under the NIH GWAS Data Sharing Policy and Future Directions. *Nature Genetics*. 2014; 46(9):934–938. [PubMed: 25162809]
2. Homer N. Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLOS Genetics*. 2008; 4(8):e1000167. [PubMed: 18769715] Gymrek M. Identifying Personal Genomes by Surname Inference. *Science*. 2013; 339(6117):321–324. [PubMed: 23329047]
3. Final NIH Genomic Data Sharing Policy, 79 Fed. Reg. 51,345.
4. Maschke KJ. Wanted: Human Biospecimens. *Hastings Center Report*. 2010; 40(5):21–23. [PubMed: 20964268]
5. National Human Genome Research Institute. Informed Consent Elements: Considerations for Genomics Research and Sample Language. last updated March 10, 2015, at <http://www.genome.gov/27559024>
6. Hull SC. Patients’ Views on Identifiability of Samples and Informed Consent for Genetic Research. *American Journal of Bioethics*. 2008; 8(10):62–70. [PubMed: 19003716] Callier SL, Schmidt H. Managing Patient Expectations About Deidentification. *American Journal of Bioethics*. 2010; 10(9):21–23. [PubMed: 20818551]
7. Lunshof JE. From Genetic Privacy to Open Consent. *Nature Reviews Genetics*. 2008; 9(5):406–411. Steinsbekk KS, Myskja BK, Solberg B. Broad Consent versus Dynamic Consent in Biobank Research: Is Passive Participation an Ethical Problem? *European Journal of Human Genetics*. 2013; 21(9):897–902. [PubMed: 23299918] Kaye J. Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks. *European Journal of Human Genetics*. 2015; 23:141–146. [PubMed: 24801761]
8. Wilbanks JT. Portable Legal Consent Overview. at <http://sagecongress.org/WP/wp-content/uploads/2012/04/PortableLegalConsentOverview.pdf>.
9. de Montjoye YA. openPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLOS ONE*. 2014; 9(7):e98790. [PubMed: 25007320]
10. Ball MP. Harvard Personal Genome Project: Lessons from Participatory Public Research. *Genome Medicine*. 2014; 6(2):10. [PubMed: 24713084]
11. Lunshof. From Genetic Privacy to Open Consent. 409

12. Ness M. GENOME: The Future Is Now WEBISODE 2. YouTube video. <http://www.youtube.com/watch?v=2r9DpthvNKM>.
13. Personal Genome Project. Personal Genome Project Consent Form. at http://www.personalgenomes.org/static/docs/harvard/PGP_Consent_Approved_02212013.pdf.
14. Personal Genome Project. Why Participate in the PGP? at <http://www.personalgenomes.org/organization/why-participate>.
15. Personal Genome Project. About the PGP. at <http://www.personalgenomes.org/organization/pgp>.
16. Bobe J. Group Photo of PGP Participants at 2013 GET Conference. Personal Genome Project Blog. <http://blog.personalgenomes.org/2013/05/10/group-photo-of-pgp-participants-at-2013-get-conference/>.
17. Data Privacy Lab. How Unique Are You? at <http://aboutmyinfo.org/>.
18. Ball. Harvard Personal Genome Project: Lessons from Participatory Public Research.
19. GET-Evidence is an open information resource developed to help PGP participants interpret their genomes; see Ball MP. GET-Evidence Guides. Personal Genome Project Blog. <http://blog.personalgenomes.org/2012/04/02/get-evidence-guides/>. see GET-Evidence GET-Evidence: About. at <http://evidence.pgp-hms.org/about>.
20. Quantified Self. About the Quantified Self. at <http://quantifiedself.com/about/>.
21. Kohane IS, Altman RB. Health-Information Altruists — A Potentially Critical Resource. *New England Journal of Medicine*. 2005; 353:2074–2077. [PubMed: 16282184]
22. Brown P. Measuring the Success of Community Science: The Northern California Household Exposure Study. *Environmental Health Perspectives*. 2012; 120(3):326–331. [PubMed: 22147336]
23. Michie M. ‘If I Could in a Small Way Help’: Motivations for and Beliefs About Sample Donation for Genetic Research. *Journal of Empirical Research on Human Research Ethics*. 2011; 6(2):57–70. [PubMed: 21680977]
24. U.S. Chief Information Officers Council. [May 2015] Project Open Data. at <https://project-open-data.cio.gov/>
25. G8. G8 Open Data Charter. at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf
26. National Human Genome Research Institute. Informed Consent Elements, Considerations for Genomics Research and Sample Language.
27. Personal Genome Project. Personal Genome Project Consent Form.
28. Corder A. Reflexive Research Ethics for Environmental Health and Justice: Academics and Movement Building. *Social Movement Studies*. 2012; 11(2):161–176. [PubMed: 22690133]
Morello-Frosch R. Communicating Results in Post-Belmont Era Biomonitoring Studies: Lessons from Genetics and Neuroimaging Research. *Environmental Research*. 2015; 136:363–372. [PubMed: 25460657]

Table 1

PGP Participants' Understanding of Risks in Comparison with Informed Consent Statements

Type of risk n (number of participants who mentioned the risk)	Example participant quotes	Excerpts from PGP informed consent ¹
Discrimination. Effects on employment, health insurance, other insurance n: 29	"Somebody could discriminate against you based on your genetic information...then you'd have to go through a legal challenge to say, 'No, you can't do that.'" "...what if I'm predisposed to some sort of breast cancer or just any health risks...I could be discriminated against for employment."	"Whether or not it is lawful to do so, you could be subject to actual or attempted employment, insurance, financial, or other forms of discrimination or negative treatment due to the public disclosure of your genetic and trait information..." (Article VI, Section 6.1, Part a)
Learning distressing information about yourself. Disease risk, paternity, racial history. n: 6	"I think the risk to self – there are obviously psychological implications of finding out you're prone to a disease." "To have information come out that you have genetic defects or that your children do, I think, I mean that could make life more difficult for somebody...I think it would destroy their image of themselves as being perfect."	"Your...data may contain potentially alarming information, such as potentially harmful genetic variants. This may cause you to experience anxiety or stress." (Article VI, Section 6.3, Part c) "... disclosure of your genetic and trait data could cause you to learn [or make]...unexpected genealogical features about you and/or your family...inferences of non-paternity, as well as inferences or allegations of paternity made by individuals you did not previously know..." (Article VI, Section 6.1, Part a)
Distressing use of biological samples or genetic data. Cloning, creation of children, placing DNA at a crime scene n: 11	"The technology to recreate portions of published genomes could down the line cause problems..." "PGP warns you that at some point in the future there may be the technology to synthesize your DNA, given your full genome and leave that at a crime scene for example. Or to create a clone of you or, you know. Things that are less feasible now but may, might become more feasible."	"It may one day be possible for a third party to use...biological materials...for new or unexpected reproductive or other purposes, including cloning." (Article VI, Section 6.2, Part a) "Anyone with sufficient knowledge...could take your DNA sequence data...[to] make synthetic DNA and plant it at a crime scene..." (Article VI, Section 6.1, Part a)
Social effects. Effects on social relationships from revealing sensitive data, such as mental health status, sexually transmitted disease, sexual orientation. Gossip or publicity. n: 13	"There's always the potential for stigma if something comes out. I'm thinking sort of mental health issues or things, maybe sexually transmitted diseases or things that have a societal misunderstanding..." "You know how people gossip. Gossip can hurt a lot of people."	"The risks of public disclosure of your genetic and trait data, including your DNA sequence data, or other information you provide, could affect the employment, insurance and financial well-being or social interactions of you and your immediate family." (Article VI, Section 6.1, Part a) "Data...may be used to identify you, resulting in contact from the press and other members of the public... This could mean a significant loss of privacy and personal time. (Article VI, Section 6.1, Part a)
Risks to family members. Consequences to family members from inferences from the participant's data n: 5	"...we might learn things about my genome that they [my family] might not be comfortable with...my genome is shared with their genomes...so a chunk of their information would be out there..."	Your publicly available...information will include certain information that applies to your family members. Some people may draw conclusions...about what such information might reveal about you and your family members. (Article VI, Section 6.1, Part a)
Use of biological samples or data for profit n: 4	"Somebody could use that to make a commercial drug that they make money off of that neither I nor the PGP get any money from." "Now, of course, then it gets into the whole profit thing, and then people within that industry become competitive, and they wanna patent to make money, and then it becomes a little seedy and, frankly, disgusting at times."	"...information and materials that you provide...may be made available to third parties for research, patient care, commercial or other purposes, and these third parties may commercially profit from the data or other information that you contribute to the PGP." (Article VIII, Section 8.4)

¹Personal Genome Project, "Personal Genome Project Consent Form," February 21, 2013, at http://www.personalgenomes.org/static/docs/harvard/PGP_Consent_Approved_02212013.pdf.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript