

Open Humans: Public Data Sharing Research Protocol

Protocol Title:	Open Humans: Public Data Sharing
Protocol Number:	Unassigned
Version:	2014-9-22
Version Date:	September 22 2014
Sponsor:	PersonalGenomes.org, a 501(c)(3) nonprofit organization 423 Brookline Ave #323 Boston, MA 02215-5410
Principal Investigator:	Madeleine Ball, PhD Email: madeleine@personalgenomes.org Phone: 857-600-1680
Study Site Contact Information:	Email: oh-publicdatastudy@personalgenomes.org

Table of contents

<u>Page number</u>	<u>Section</u>
1	Cover sheet
2	1. Background and purpose of the study
2	1.1 Background
2	1.1.1 Summary
3	1.1.2 How this study compares to existing projects
5	1.1.3 Purpose of the Open Humans Portal, distinct from the study proposed here
6	1.2 Purpose of the Open Humans: Public Data Sharing study
7	2. Description of the Study Population
8	3. Methods and Procedures
8	3.1 Background: Data and Profile Management as an Open Humans Member
9	3.2 Public data sharing as a Study Participant
9	3.3 Data collection, analysis, and monitoring
9	3.4 Data access and non-confidentiality
10	3.5 Data removal and withdrawal
10	4. Risks and Benefits
10	4.1 Risks
11	4.2 Benefits
11	5. Recruitment and Consent
11	5.1 Recruitment
12	5.2 Consent
13	5.3 No costs or payment for participation

Acronyms used in this proposal:

HIPAA/HITECH Act	=	Health Information Portability and Accountability Act and Health Information Technology for Economic and Clinical Health Act
OH	=	Open Humans

1. Background and purpose of the study

1.1 Background

1.1.1 Summary

The Open Humans Portal, www.openhumans.org, is a website operated by PersonalGenomes.org, a 501(c)(3) non-profit organization. As of August 2014, this website and associated projects are funded through awards from the Robert Wood Johnson Foundation and the John S. and James L. Knight Foundation.

Research data generated by human subjects research is often kept private to the study that collects it: this situation appropriately respects privacy of individuals, but also limits access, redistribution, and reuse of valuable data. In particular, the act of sharing data with participants is often unexplored, and could lead to diverse beneficial consequences for both participants and society. The Open Humans Portal is designed on behalf of individuals to empower them to liberate their own existing personal research data from these “research data silos” by (1) enabling individuals to privately integrate and manage their existing personal research data from those disparate sources all in one convenient location and, (2) pursuant to this proposed research study protocol, enabling individuals to share publicly their integrated research data profiles. The Open Humans Portal serves as a valuable online community ecosystem where interested individuals and researchers can discover and connect with one another.

Five key community roles exist in the Open Humans Portal (OH Portal):

1. OH Staff:
Those involved in creating and operating the OH Portal and the Open Humans: Public Data Sharing study (OH Public Data Sharing study)
2. OH Visitors:
Those who look at and/or use the open access resources on the OH Portal
3. OH Members:
Those who register an account to access features of the OH Portal (such as personal profile, private data import and management, and bulk data aggregation tools)
4. OH Study Participants:
Those who voluntarily elect to participate in the OH Public Data Sharing study to share their highly integrated, longitudinal data and profiles as “Open Humans”
5. OH Research Partners:
Those investigators of equitable, IRB-approved research projects who agree to accept and fulfill requests, initiated by individuals using OH Portal tools, to retrieve existing personal research data for the individual’s personal management on the OH Portal

While enabling rich, integrated, longitudinal public data profiles for future research is an important and desired outcome for the Open Humans Portal, the fundamental motivation behind the creation of Open Humans is to empower individuals to manage their own research data and to release those data as they see appropriate (*i.e.*, liberating the individual’s research data from data silos). As such, it is important to highlight that the interactions OH Members will have with OH Research Partners is on behalf of and for the primary benefit of OH Members to make it easy for each individual to claim, retrieve, import, and export their own research data. In this way, Open Humans will promote communications and discussions outside of any existing research protocols, which will nourish ideas and insights for future scientific endeavors and collaborations within the OH community.

Open Humans is seeking WIRB’s opinion on the initial phase of the Open Humans: Public Data Sharing study (*i.e.*, the public data sharing features of the OH Portal). OH specifically requests WIRB’s opinion on:

- A. the consent process to enable OH Members to become OH Study Participants
- B. the Study Participant-initiated public sharing of existing research data in an integrated profile

The initial phase of the OH Portal will be focused on supporting data import from four sources. Three of these are OH Research Partners: the Harvard Personal Genome Project (PI-George Church, Harvard Medical School), the American Gut Project (PI-Rob Knight, University of Colorado-Boulder), and the Flu Near You: GoViral Study (PI-Rumi Chunara, Boston Children’s Hospital). The fourth source will be raw genotyping data that participants import from one or more direct-to-consumer genetic testing providers (e.g. 23andme). OH Staff anticipate adding support for new Research

Partners and other data imports to the OH Portal at a later date. Eligibility for the OH Public Data Sharing study will be limited to individuals who are registered OH Members and have imported research data to the Open Humans Portal.

As the Open Humans Portal develops, the OH Portal is anticipated to include additional features. These features may include, but are not limited to, (a) facilitating in-network recruitment of OH Members for their voluntary participation in new research projects (such as network advertisements, personal emails, a registry of members, participant profile filters for equitable researchers, and a catalogue of equitable research studies), (b) allowing peer-to-peer data sharing options among Members, (c) hosting surveys developed by Research Partners, the responses to which Members and Study Participants could integrate in their research profiles, (d) facilitating research data visualization and interpretation of imported data, including link-outs to external tools, (e) promoting new research collaborations between Members and Research Partners, (f) enabling OH Members to propose research hypotheses, and (g) polling OH Members for input on portal feature development and research priorities.

It is important, however, to keep the OH Portal conceptually distinct from the proposed OH Public Data Sharing study. OH is not requesting WIRB's opinion of the OH Portal features at this time. A comprehensive discussion of all facets of the OH Portal is not provided in this proposal. Rather, when details about the OH Portal are provided in this proposal, they are provided to facilitate an understanding of the context in which the OH Public Data Sharing study will be conducted.

1.1.2 How this study compares to existing projects

The efforts of this proposed study (Open Humans: Public Data Sharing) to host a valuable source of public data that researchers could use for future research and secondary analyses pursuant to distinct IRB-reviewed and IRB-approved protocols are similar in general spirit to the efforts of our colleagues at Sage Bionetworks. Therefore, we would like to draw WIRB's attention to the ways in which the OH Portal and proposed OH Public Data Sharing study compare to Sage Bionetworks' efforts for Portable Legal Consent (PLC), Common Genomic Research Environment (CGRE), and Synapse Commons (Synapse). Please see WIRB Protocol #20112068 and Proposal titled "Citizen-contributed Genomic data cohort" dated March 9, 2012. A summary of the similarities and differences is provided here to help present a clear picture of the present proposal for the OH Public Data Sharing study.

Like the above-referenced efforts of Sage Bionetworks, the OH Public Data Sharing study involves no physical procedure or direct contact with participants. The main function of this study is not to generate *new* data for research but, rather, to help individuals manage and share *existing data* for research. Like Sage Bionetworks, the OH Public Data Sharing study will use an online consent process in which (1) participants will voluntarily elect to consent or not consent based solely on information provided in the consent document; (2) participants can contact the study's Principal Investigator and/or other staff with any questions they may have, but there is no mandatory consent discussion scheduled; and (3) the decision to consent is not time-sensitive but allows the individuals to take as much or as little time as they need to make their decision.

Open Humans has important distinctions from the above-referenced efforts of Sage Bionetworks. These distinctions relate to (1) the conceptual framework and the timing of the informed consent process and (2) the terms of use and identifiability of data made available in the resulting Open Humans dataset. In contrast to the framework designed by Sage Bionetworks (where all individuals interested in importing their existing research data into the CGRE are taken through the PLC process and become participants of Synapse), OH provides space for its members to integrate and manage their existing research data without becoming participants in a new study. Only those members who subsequently wish to publicly share any of their integrated research data are taken through the informed consent process to become Study Participants of the OH Public Data Sharing study. To reiterate, the OH Portal is distinct from this study. (Contrast Figure 1 below with "Figure 1: PLC-CGR workflow" on page 8 of the above-referenced proposal by Sage Bionetworks). We provide some discussion of the data integration and management features in this proposal as necessary to promote clarity for how this study will operate with the OH Portal; however, these features are not to be considered part of the protocol for this study.

Individuals can be OH Members without encountering the informed consent process or becoming participants in this proposed study. It is anticipated that some individuals will create an account for convenience to privately manage their existing research data from multiple studies in which they participate or have participated. These individuals may not have any immediate interest in making those integrated research data publicly available. It is the attempt to share the integrated

research data publicly and openly that triggers the online consent process for this proposed study. Only those members who voluntarily elect to become OH Study Participants can use the public sharing features of the OH Portal.

Unlike Synapse (which uses a 3-tier data access system and imposes corresponding levels of data use restrictions for each tier), Open Humans Portal will impose no data use restrictions on public data and profiles. While the OH Portal will have Terms of Use and Data Use Guidelines for users and members, the Open Humans data and profiles publicly shared by the OH Study Participants are unrestricted public domain materials, being either released with a CC0 Public Domain Dedication (“No Rights Reserved”) or as non-copyrightable material. Another important distinction between Synapse and the Open Humans public data is that OH does not guarantee that data will be stripped of direct HIPAA identifiers. This is in line with the approach taken by the Harvard Personal Genome Project, an approach that (1) acknowledges privacy, confidentiality, and anonymity are impossible to guarantee with the public sharing of genetic data and (2) seeks to engage only those individuals who are willing to forgo assurances of privacy in order to make lasting contributions for scientific research. Within the OH ecosystem, the individual initiates and authorizes the importing of data and the data sharing or publication (regardless of whether those data may have included “protected health information” for purposes of HIPAA/HITECH Acts).¹ When individuals initially enter the area of the OH Portal for managing research data, the individuals will encounter a link leading to the informed consent process for this proposed OH Public Data Sharing study. Additionally, each time Study Participants initiate the feature to publicly share research data from an additional study, OH will remind them of data sharing risks and require the individuals verify the authorization for that data sharing transaction.

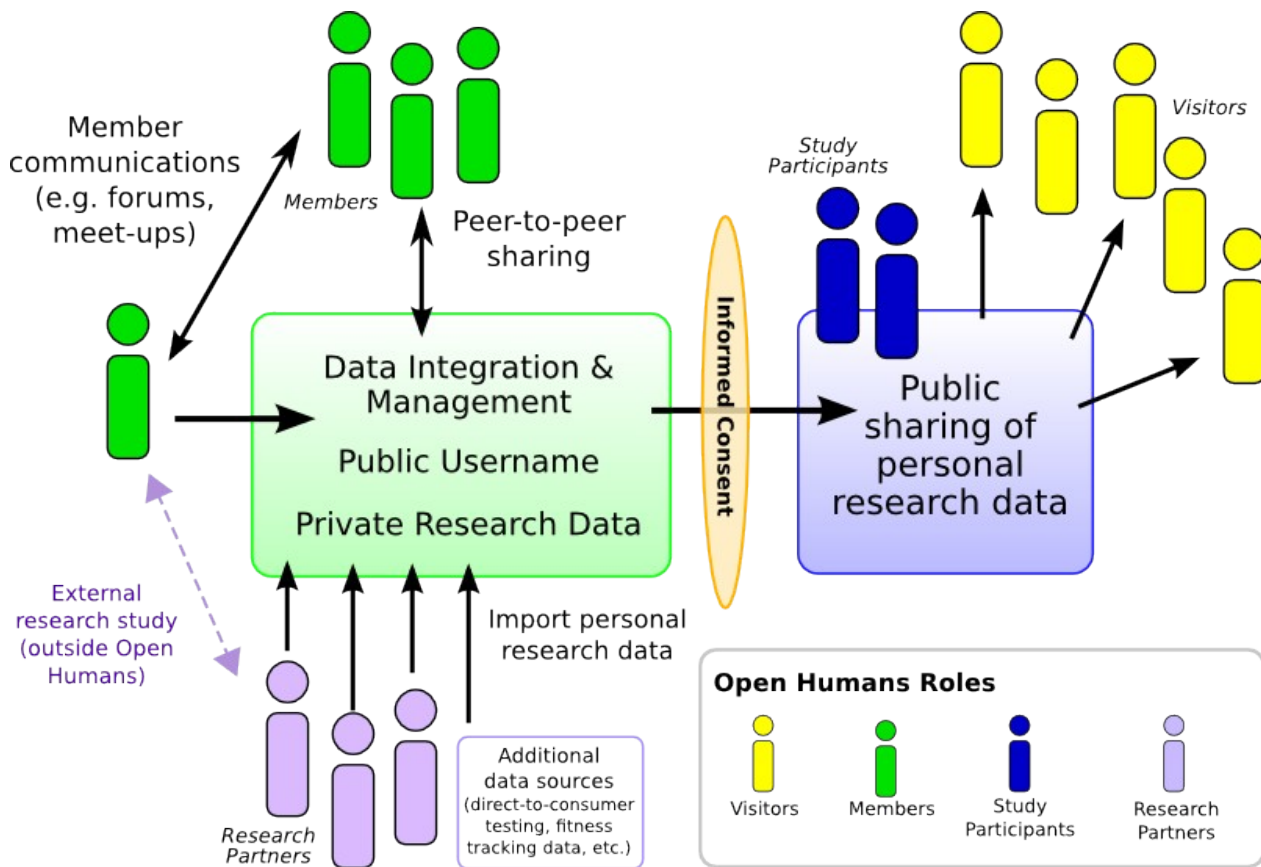


Figure 1. Schematics for the OH Portal and OH Public Data Sharing study

Information flow and community roles in OH Portal, illustrating that only those members who provide informed consent to become OH Study participants are able to use public data sharing features of the OH Portal.

¹ As explained in section 3.1, it is important to keep in mind that the Open Humans Portal is not a covered entity or business associate required to comply with the HIPAA/HITECH Act..

1.1.3 Purpose of the Open Humans Portal, distinct from the study proposed here

Solving some of our society's most serious public health and medical problems requires bringing together a critical mass of people, ideas, and data in open research. Collective intelligence cannot be leveraged when data are confined to a single academic lab, commercial company, or health care system. The value of sharing data is abundantly clear, and technical barriers are now surmountable thanks to the Internet and advances in information technology. The remaining barriers are legal and ethical, not technical. For example, while privacy promises must be tempered in light of the growing amount of research data now recognized to be re-identifiable, data sharing must ultimately be restrained so that researchers do not break privacy promises that were made to participants when initially collecting or generating the data. Current research practices create data silos that prevent us from realizing the full potential of the generous contributions that individuals have made to research (either directly through actual participation in the projects or indirectly through public taxpayer dollars supporting the projects). After design and funding, each study recruits and enrolls a cohort, collects samples, and generates private data to produce publications and intellectual property. The resulting data silos are economically inefficient and fail to enable the integration of data, participants, and insight. The OH Portal is an alternative paradigm that focuses on open data and research as a collaborative effort from the outset.

Participatory research studies that share raw research data with their enrolled volunteers are a new practice and an enabling force toward more meaningful research collaborations. They create opportunities for participant-initiated data sharing. However, most studies continue to withhold data from volunteers, sometimes because of inertia ("we've always done research this way") and other times because of practical challenges ("we don't know how to do research that way"). Public data sharing continues to be stymied by one-size-fits-all approaches to data privacy issues, even though research participants have diverse interests and priorities. As a result, access to valuable data resources is mediated by a few who decide what to investigate, publish, and share. When data sharing does happen, the sharing typically occurs through restrictive controlled access databases. The OH Portal will make it possible for individuals to (1) locate equitable research studies that return data to participants, (2) aggregate their own research data over time, and (3) maximize the impact of their contribution to science by facilitating public access to their own data. By highlighting the success of IRB-approved research projects that share data with participants, the OH Portal will promote future equitable research projects by enabling the highlighted protocols to become valuable roadmaps (*i.e.*, design guidelines) for those setting up new projects.

Highly integrated, longitudinal biomedical data are extremely valuable for science. Because detailed, individual-level data can be impossible to anonymize, traditional privacy assurances are problematic. These assurances create structures that limit the utility of the data, reduce data sharing, inhibit data aggregation and collaboration, and compromise discovery. Anonymization practices often isolate participants from one another, preventing them from contributing their full value, and prevent them from learning of and participating in the distribution of results of the studies in which they participate. The OH Portal will empower individuals to connect with other individuals (and researchers) with shared interests and to mediate access to their existing research data in ways they see fit.

Participants in Open Humans: Public Data Sharing will have the ability to publicly share their aggregated data. The OH Portal will include a public interface for accessing, exploring, and downloading this publicly shared data; a member-facing interface that allows individuals to manage their private and public research data sets; and design guidelines and resources for those who want to design equitable research projects. Not all aspects of the OH Portal are part of this proposed study, so it is important to delineate those boundaries of the OH Public Data Sharing study. Figure 1 illustrates the various aspects of the OH Portal, including the timing of the informed consent process. OH Members not enrolled in this study will have access to many data aggregation and management features not part of this study. The Public Data Sharing study is limited to the public data sharing feature, which will be available only to OH Study Participants.

Many aspects of the OH Portal are not part of a systematic investigation and have not been designed to develop or contribute generalizable knowledge. The communication features that may be designed for the OH Portal (e.g., message boards, lists of equitable research studies, links to external websites, and administrator-member and member-to-member forums, email, or other messaging) are not features intended for systematic investigation, development of generalizable knowledge, or research recruitment (unless such recruitment is done pursuant to separate IRB-approved protocols). The "Data Integration and Management" features (shown in blue in the above Figure 1b) are for the individual and not data collection for any research purpose (although those data were generated for research originally and copies may be available elsewhere for research purposes). Any information imported or uploaded by the individuals in this way is intended to be for the individual's personal benefit and use (and not intended for use by OH investigators).

1.2 Purpose of the Open Humans: Public Data Sharing study

The proposed OH Public Data Sharing study involves the public sharing of pre-existing research data online. This study is predicated on the development of the Open Humans Portal at www.openhumans.org and the Study Participants' use of the OH Portal as an OH Member to integrate and manage their existing personal research data.

The OH Portal is inspired by and will collaborate with the Harvard Personal Genome Project. Founded in 2005, the Harvard Personal Genome Project has championed a model that returns research data to participants and enables them to share those data publicly. To promote data aggregation, integration, and discovery, the Harvard Personal Genome Project has turned the privacy problem on its head: rather than jeopardize the trust and integrity of the researcher-participant relationship with weak promises about anonymity, the Harvard Personal Genome Project specifically recruits individuals who are comfortable with public data sharing and willing to assume risks of re-identifiability. The practice is called "open consent." To date, approximately 3000 Harvard Personal Genome Project participants have contributed extensive data and biospecimens with the goal of advancing scientific knowledge and discovery. The critical importance of such public resources is now readily recognized, and the OH Portal aims to enhance existing efforts to provide high quality reference material (such as those by the National Institute of Standards and Technology, the U.S. Food and Drug Administration, and the Genome in a Bottle Consortium).

While enabling rich, integrated, longitudinal public data profiles for future research is an important and desired outcome for the Open Humans Portal, the fundamental motivation behind the creation of OH is to empower individuals and revolutionize biomedical research by enabling individuals to manage their own research data and to release those data as they see appropriate (*i.e.*, liberating the individual's research data from data silos). The Open Humans Portal serves as a valuable online community ecosystem where interested individuals and researchers conducting equitable research projects (*i.e.*, truly collaborative and participatory research projects in which raw research data are returned to the participants) can connect with one another.

The main purpose of the OH Public Data Sharing study specifically (as opposed to the aforementioned purpose of the OH Portal generally) is to improve our understanding of public sharing of research data by participants. We are interested in better understanding (1) factors that affect the decision to publicly share data, (2) the consequences this data sharing has for the participants themselves, and (3) the effect it has on data re-use beyond the study for which data was originally generated. This study involves no interventions. This study is observational and longitudinal (sometimes referred to as "hypothesis generating research") and hopes to address the following questions:

- What are the pros and cons of sharing individual-level research data beyond the scope of the initial study in which those data were generated?
- To what extent does public research data sharing by participants lead to re-use of research data?
- What data compatibility and quality control problems do current research data silos create, and how can those issues be resolved?
- What are the positive and negative consequences of public sharing of research data?
- How do people decide to share data publicly, and what prompts them to change their minds?
- What do people understand about data sharing risks and benefits?
- What awareness, understanding, and attitudes do people have about the tradeoffs between various factors when establishing research data sharing norms? For example...
 - How do individuals weigh privacy interests with their desires to maximize the potential impact any given contribution to science can have?
 - How does understanding and attitudes vary for different data types?
 - What demographic aspects of individuals are correlated to decisions to publicly share data? How do these factors relate to attitudes and evaluations of risk?
- Can individual-initiated data sharing accelerate biomedical research discoveries, and how?

2. Description of the Study Population

OH seeks to enroll 1,000 Study Participants in the OH Public Data Sharing study's initial phase. The OH seeks a diverse study population to include adult individuals of all genders and racial and ethnic backgrounds who are willing to share

their integrated research data publicly. The participants of the OH Public Data Sharing study must be individuals who are Members of the OH Portal and who have imported raw research data (from a Research Partner or other compatible research data). Individuals must be citizens or documented permanent residents of the United States. OH participants must be at least 18 years of age. Individuals must have legal capacity to provide autonomous, voluntary consent (*i.e.*, legal guardians are not permitted to enroll individuals lacking mental capacity). Individuals must be willing to share data without any promises or assurances of privacy, confidentiality, or anonymity and must demonstrate they are knowledgeable about benefits and risks of public data sharing.

Participation in the OH Public Data Sharing study is completely voluntary and does not affect the individuals' participation in other traditional research endeavors (including their participation in the underlying studies conducted by Research Partners) and does not affect the individuals' ability to use the OH Portal to integrate and manage their own existing research data.

Some vulnerable groups are not included in our exclusion criteria: prisoners, poor/uninsured, institutionalized, pregnant women, students of the PI or study staff, employees directly supervised by the PI or sub-investigator, and employees of the research site or sponsor. Our reasons for this is as follows:

- Recruitment occurs online and will be available to these vulnerable groups.
- No recruitment is made specifically targeting these groups.
- Exclusion does not prevent individuals from taking the same data sharing risks: they may engage in public data sharing external to the Open Humans Portal. In such cases, the main consequence of exclusion is that these individuals' public data is less likely to be seen and re-used by researchers and impacts of sharing will not be tracked.
- Exclusion from participation may make individuals feel excluded from the Open Humans community.
- We believe inclusive treatment of all Open Humans Members is important for increasing the empowerment of participants in research.
- As described in section 4.2, this research is not expected to personally benefit participants, but it is expected to produce broader benefits to research and society; exclusion of vulnerable groups may result in excluding their demographics from access to the benefits of this research.

Additional specific reasoning applies to the following groups:

- **Pregnant women**
This study involves no physical procedures, and thus is not representing any direct risk to the fetus. There are no additional risks of participation due to a participant's pregnancy.
- **Students and Employees of the PI, research site, and sponsor**
It will be made clear to employees and students that participation in this study is entirely voluntary, and no specific recruitment will be performed of employees and students. However, employees and students are likely to be attracted to PersonalGenomes.org for community reasons; notably several employees in our organization are currently members of the Harvard Personal Genome Project and are already publicly sharing their identifiable/sensitive genetic data. Thus, we believe it is especially likely some employees and students will strongly wish to be participants in this study, and that excluding these individuals would be unfairly deny them the opportunity to participate.

3. Methods and Procedures

The proposed OH Public Data Sharing study involves the public sharing of pre-existing research data online. This study is predicated on the development of the Open Humans Portal at www.openhumans.org and the Study Participants' use of the OH Portal as an OH Member to integrate and manage their existing personal research data.

3.1 Background: Data and Profile Management as an Open Humans Member

OH enables Members to bring their existing research data into the OH ecosystem and manage them in an integrated profile. This management feature is not research and is, thus, outside of the scope of the OH Public Data Sharing study. OH Members are not asked to provide informed consent before using the data integration and management features of the Open Humans Portal.

Members will have the ability to import research data from Research Partners and also upload research data they may have in their own possession in a variety of formats (*e.g.*, data from direct-to-consumer companies and wearable fitness tracking devices, such as 23andMe and FitBit, respectively). Data integration and management is at the individual's initiation and for the individual's benefit. OH may impose basic restrictions on the file formats and sizes supported and may limit the amount of total space for an individual to manage personal research data. Such rules would be specified in the OH Portal's Terms of Use and/or Community Guidelines. Data imported to an OH Member Profile are kept private.

It is important to keep in mind that OH is not a covered entity or business associate required to comply with the HIPAA/HITECH Act. The OH has been developed on behalf of and for the primary benefit of the individuals (not on the behalf of any HIPAA covered entities or business associates). Any interactions the OH has with HIPAA covered entities or business associates are on behalf of the individuals, with OH acting simply as a gofer to claim, retrieve, and/or share existing research data at the direction and authorization of the individuals. OH, while not required to comply with HIPAA/HITECH Act, will voluntarily take reasonable efforts to maintain security and privacy for the OH data management and integration features of the OH Portal, including usage of passwords, encryption, and authorization keys to secure access to sensitive information. Sensitive account modification actions taken by Open Humans Members and Study Participants (*e.g.* password reset or changing a user's email address) will require password confirmation and will trigger a notification email.

As described in our Terms of Use, the Open Humans Portal will track various interactions with the site (*e.g.* clicking links, downloading data, and profile viewing) including those performed by Visitors, Members, and Study Participants. Privately collected data will include IP addresses and, when possible, user account identifiers. This data collection has several purposes: (1) to monitor the site for usability, and (2) to give us information for potential investigations and corrective actions should we discover violations (legal or illegal) of the site's Terms of Use, Data Use Guidelines, and other site policies and (3) to collect statistics for internal use and that we can share with Members for their education (*e.g.* number of site visitors, number of data downloads).

3.2 Public data sharing as a Study Participant

Participation in the OH Public Data Sharing study enables a Study Participant to share some or all existing research data from multiple studies as an integrated dataset, public and associated with their Member username and account. The data from the Research Partners (*i.e.*, Harvard Personal Genome Project, American Gut, and Flu Near You) will be the seed data from which the OH ecosystem will take shape.

The consent process is described separately below. Once an OH Member provides informed consent and becomes an OH Study Participant, the public data sharing features will be unlocked and enabled for that Member account (now also a Study Participant account). The OH Study Participant may choose to publicly share research data they have imported and privately stored. Public data sets are shared associated with the individual's public username, and may be shared on an individual basis (*e.g.* a Study Participant may choose to publicly share their American Gut data and not publicly share their imported 23andme genotyping data). Prior to publishing the profile and integrated data as an Open Human, the Study Participant will get reminder prompts indicating the data sharing risks and will be asked to verify the action.

3.3 Data collection, analysis, and monitoring

To better understand public data sharing behaviors, practices, consequences, and impacts, the investigators will track the interactions that OH Study Participants have with the OH Portal. These interactions could include (but are not limited to) the following:

- the content and frequency of questions submitted to OH support staff,
- what types of OH data files are shared,
- the frequency and extent to which particular types of OH data files are shared,
- extent of Study Participant usage of features on the OH Portal,

Also, to better understand the downstream use of publicly shared data, Open Humans will track:

- access to and downloading of OH data, based on de-identified data collected by the Open Humans Portal (as described in section 3.2)
- publicly described derivative usage of the data, *e.g.* publications

To better understand Study Participant attitudes, preferences, behaviors, and experiences related to public data sharing, the investigators will administer impact surveys after 12 months of public data sharing. These impact surveys may contain questions such as (but not limited to) those provided in the exemplar attached to this proposal.

To better understand the consequences of public data sharing, the investigators of the OH Study will also monitor access to and downloads of Open Humans datasets through IP address and/or OH account.

3.4 Data access and non-confidentiality

There is no data confidentiality for information publicly shared by Open Humans Study Participants. Members of the Open Humans Portal will have publicly shared usernames, which may be their real name (but may not impersonate another individual). Study Participants will share data associated with that username. Participants will not be assigned a study ID number; based on the Harvard Personal Genome Project experience, this might cause confusion or mislead the individuals to believe their anonymity, privacy, and confidentiality is protected in ways that it is not.

Data published publicly on Open Humans are available to any OH Visitor. You do not need to be a registered OH Member to access and download publicly shared data. The Data Use Guidelines are a non-legal set of community norms, and include a guideline to not attempt to re-identify Open Humans Members who have not intentionally shared their name/identity.² As mentioned in section 3.1, the Open Humans Portal will privately monitor interactions with the site (for the purposes of improving user experience and for security reasons), including access and download of publicly shared research data.

OH maintains strict information technology procedures to prevent improper access of private data integrated and managed by OH Members. However, the OH Study will intentionally publish, without any restrictions, data the OH Study Participants have elected to publicly share.

3.5 Data removal and withdrawal

Participants can elect to have their data removed from public sharing, either individually or in whole (*i.e.* no longer publicly shared on the Open Humans Portal). This option will be presented to Study Participants on their data management page.

Participants may also entirely withdraw their enrollment in this study with a written request to oh-publicdatastudy@personalgenomes.org. Former participants will no longer have the ability to perform a public release of currently-private research data in their account, but may choose to leave their currently-public data as publicly shared. They will also retain the ability to remove public sharing status, but after doing so will not be able to re-release that data without re-entering this study.

Although the participant's data will no longer be publicly shared on the Open Humans Portal, other copies of the data may exist elsewhere (*i.e.* external to our project), if the data were copied while they were available publicly. Participants will be informed when they sign up that we can only remove data from the Open Humans Portal and cannot destroy copies stored elsewhere (such as on the Research Partner sites) or copies that may have been downloaded already by OH Portal Visitors and Members.

4. Risks and Benefits

4.1 Risks

No physical procedures will occur with Study Participants as part of this study. With the exception of the monitoring activities described in section 3.3, the data involved in this study are already in existence. Risks derive from the potential consequences associated with making identifiable and/or sensitive data public. These include risks of re-identification (if the participant does not elect to attach their identity to their data), discrimination, embarrassment, and identity theft. In

² An exception to this guideline is made for re-identification studies, provided they inform Open Humans of their study and follow other responsible study practices.

addition, specific data types come with specific risks (e.g. sharing location data introduces a risk of being approached unexpectedly by someone).

These risks are described in detail in our consent document (submitted with this protocol), as we seek to ensure that all Study Participants have a strong understanding of the risks of participation. Not assigning a study ID number and encouraging use of real names or identifiable usernames (such as Twitter handles) also makes the risk of identifiability more transparent to the individuals before they release any potentially sensitive and/or identifiable data in an integrated format.

A data safety monitoring board (DSMB) is not required for a study such as this one that is observational and does not involve research interventions. OH Public Data Sharing study will not create and use a DSMB. The OH Portal will provide a support button and support email address where OH Study Participants can contact the OH Public Data Sharing study staff and investigators to report an incident of harm or discuss any questions and concerns. Within a reasonable amount of time, OH Public Data Sharing study investigators will review any reports and support requests submitted by participants and will respond when necessary. The impact surveys administered after 12 months of public sharing may also incorporate questions to elicit information regarding the participants' perceived risks and experiences of actual harms that may have occurred as a result of the public data sharing. Periodically and in response to incidents, the investigators will re-evaluate whether these procedures are sufficient and whether a DSMB should be created for the study.

4.2 Benefits

Personal benefits of participation in the OH Public Data Sharing study include the self-satisfaction, education, and entertainment that individuals may get from liberating their existing personal research data from research data silos and controlling the scope of data about themselves that is available open access in one integrated, longitudinal data set. Personal benefits of research might not be direct, immediate, proximate, or tangible. For example, this research project is unlikely to help an individual find a pharmaceutical drug that is tailored for a health condition he/she has currently or be used by doctors in the near future to improve his/her health care. While this research may not make a significant difference in any particular Study Participant's life, health, or well being, it might make a difference to society by encouraging equitable research; helping us better understand genetic and phenotypic variation; advancing biomedical scholarship and technologies; helping us reach public health goals; and helping us end health disparities and discrimination. Participants may benefit from these and other societal, collective benefits this research enables.

5. Recruitment and Consent

5.1 Recruitment

Advertising the OH Portal and recruiting individuals to become OH Visitors and Members is outside the scope of this OH Public Data Sharing research protocol.

Users of the OH Portal may encounter the OH Public Data Sharing Study through the presentation of "public sharing" as a disabled data management feature, only available to participants in this study. When individuals who become OH Members and begin importing and managing their data on the OH Portal, they will see a locked public data sharing feature for their imported research data. Attempts to open or unlock the public sharing feature will provide the Member with brief information about this study, including eligibility requirements (e.g. Figure 2 below), and a link to the study information page (see recruitment materials).

Our recruitment materials are submitted in a separate document and include:

- Study information webpage on the Open Humans Portal. Includes a link to begin the enrollment process.
- Study information webpage on our nonprofit organization's website
- Online advertising (e.g. Google Ads)
- Printed study information

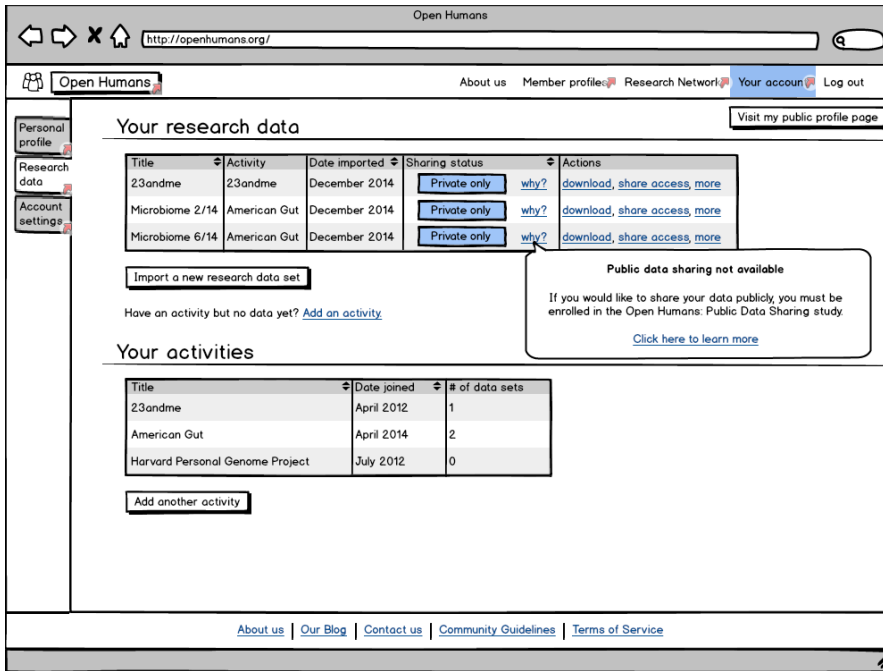


Figure 2. Study presented to users through “disabled” feature in the Open Humans Portal

5.2 Consent

When the informed consent process is triggered OH Members will be asked to review the Research Subject Information and Consent Document, which provides important information about data sharing risks and benefits (e.g. Figure 3). The ability to sign the consent document is disabled at this stage; before being invited to sign the consent document, the Member must complete a quiz testing their understanding of the study’s risks and protocols as described by the consent document (e.g. Figure 4). This process is similar to that of the Harvard Personal Genome Project, but the information and quiz are focused specifically on data sharing risks.

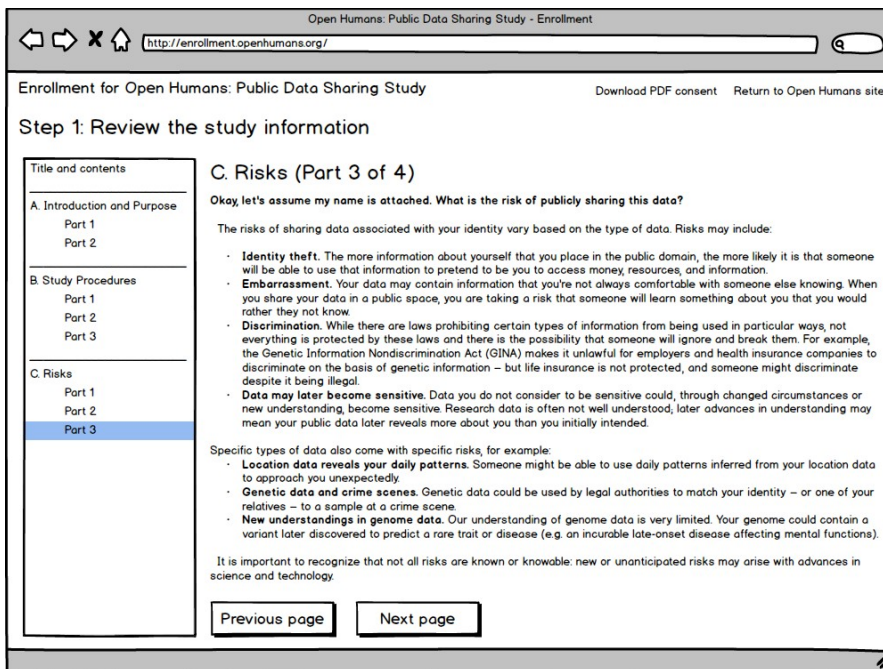


Figure 3: Example page for the online presentation of the consent document

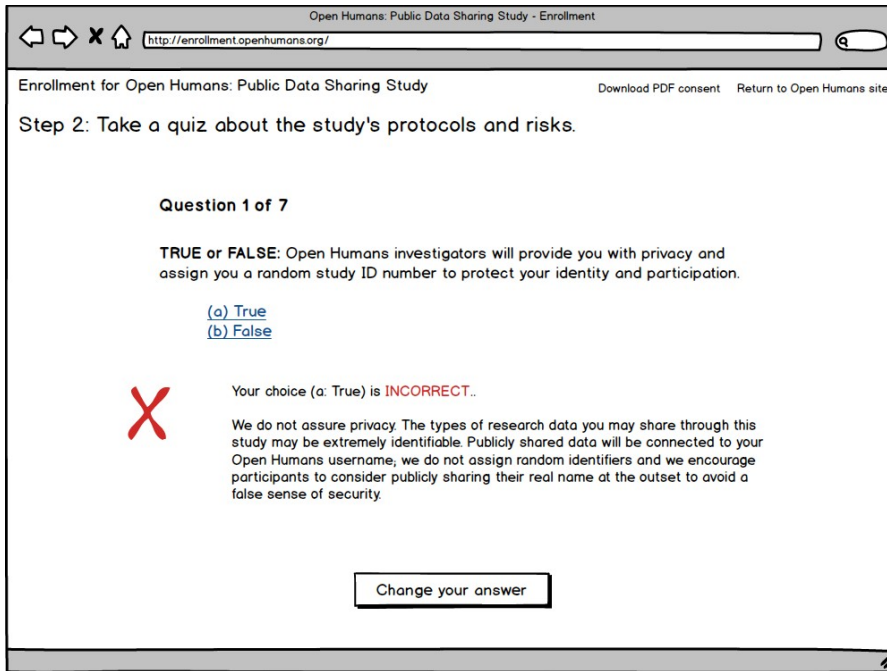


Figure 4: Example page for the quiz portion of the enrollment process

After answering each question, the Member's answer will be evaluated and a section of the consent document will be quoted that explains why the answer was correct or incorrect. The Member must change any incorrect answer to a correct answer before progressing to the next quiz question. After completing all quiz questions in this manner, the Member is returned to the consent document and invited to electronically sign it.

Copies of the consent document and quiz are attached to this proposal.

5.3 No costs or payment for participation

There are no costs to participate, Study Participants are not paid and no medical care or treatment is provided to participants as part of this study. No compensation or payment will be provided to participants.