Data Sharing Requirements for Individual Participant Level Data

Purpose

The purpose of this guidance is to inform researchers on how to respond to individual (de-identified) participant-level research data sharing conditional mandates from publications, funding agencies, and consortiums. This guidance focuses on data sharing requirements for individual participant level data regardless of funding. Individual participant level data is raw data from individual participants. It includes demographic information for each participant such as age, sex, nature of their health condition, as well as information about evaluations/treatments or tests received as part of the research and the outcomes observed.

Please note NIH has very specific requirements and associated timing. This information is included at the end of the document.

Policy

Data sharing must balance the data sharing requirements with privacy and human subject considerations for each protocol on a case by case basis. Investigators need to plan for data sharing as part of the initial IRB review process and include this possibility in the informed consent, as appropriate. Investigators should take into consideration any expectations of funding agencies and journals requirements before beginning the research. Increasingly journals and funding agencies have an expectation of placing individual participant level data into repositories when the research is completed.

The protocol and consent form should address the sharing of summary, aggregate or statistical analyses as well as individual participant-level research data. Investigators working on research covered by the NIH Genomic Data Sharing (GDS) Policy must obtain participants' consent for their genomic and phenotypic data to be used for future research purposes and to be shared broadly. This also includes specimens used to generate a cell lines as well. The IRB template consent form contains language which meets the requirement of the GDS policy.

The IRB office will be asked to confirm that the approved protocol and consent allow sharing of individual level data as proposed by the investigator. If sharing participant-level data externally is not approved as part of an initial protocol review, a protocol amendment may be required at a later date and the IRB will review whether the sharing of data is consistent with the scope of consent originally obtained from participants. Additional consent may be required if not originally addressed.

IRB review will not be necessary for sharing of methods, protocol-related documents, summary, aggregate or statistical analyses.

Procedure

Information suggested to be included in a IRB protocol application.

- What applicable requirements exist for sharing individual level data?
- What individual level data might be shared?
- Who would you likely share data with?
- What would be the likely mechanism to share data for example a controlled or open access system?
- Do you anticipate there may be sensitive data that would not be shared at the individual level? If so please describe*.
- Is there an increased risk of de-identification based on same numbers or rare conditions*?

* In cases of small cohorts or sensitive data where this is an increased risk of de-identification, funders or journals may require a letter from the IRB to justify why individual level participant data cannot be shared. For this reason, it is important to think about and address this possibility in the initial protocol application.

The IRB has developed template consent language that may be used to help guide investigators in considering appropriate language for the informed consent document. Please refer to the English Informed Consent Template.

Mechanisms for Sharing Information

There are multiple mechanisms to share individual level data.

Deposit Data into a Central Repository – a site which provides access to data submitted from multiple sites or researchers. There are two types of repositories: controlled and open access.

i. **Controlled Access** – is a restricted access process in which a scientist requests access to query the data and signs a data use agreement. An example of this includes dbGap

ii. **Open Access** is an unrestricted access process in which anyone can request or download the data without any vetting of the use or scientist, or an agreement. Examples of this may include a foundation website or NIH GEO(Gene Expression Omnibus).

Recommendations for Considering Data Repositories

Many publications or funding agencies now have specific requirements to deposit data in a repository. Sometimes open access is mandatory. Privacy, consent and other regulatory issues must be considered before sharing participant level data to central repositories, especially if the data is open access.

The preferred method for sharing in a data repository is controlled access as previously described. Completely unrestricted public posting of data on an open access website can be approved under specific conditions, including, but not limited to: participants prospectively provided consent and the individual participant-level data is 'limited,' so there is no risk of re-identification.

Any investigator who plans to submit individual level data into an open access database must be certain the informed consent document describes the plan to share data in an open access database; otherwise the investigator may be limited to use of controlled access databases.

Alternative Method for Sharing Individual Level Participant Data:

Data Available Upon Request

This process is used when a PI provides contact information and state the data is "available upon request." This request is vetted through the PI. A data use agreement or equivalent agreement process may be executed through Clinical Trials Business Office (CTBO).

How to document a data sharing plan for protocol submission or external entity

If you need specific language for a data sharing plan, consider customizing the following

> We are committed to making resources and data from the proposed research available to other investigators in the research community. All data collected for this research will be obtained with IRB review/approval and a waiver of consent/informed consent of study participants to sharing of de-identified data. We will submit (insert type of data to be shared as well as relevant associated data for example phenotype and exposure data) to XXX data repository in a timely manner, as indicated by the XXX policy; we acknowledge the registry will have (insert restricted /unrestricted access). We will also submit any information reasonably necessary to interpret the data, such as study protocols, data instruments and survey tools. The identities of research participants will not be disclosed to the repository. We will take appropriate steps to de-identify datasets according to the HIPAA privacy law. https://privacyruleandresearch.nih.gov/pr_08.asp
>
> Aggregate Data (will/will) not be available for submission/general research use.
>
> Insert a paragraph describing any data sharing restrictions: e.g., limited to cardiovascular disease or breast cancer; any restrictions described in the informed consent.

Other Resources and Information

- **The 2003 NIH Data Sharing Policy** is current and in effect until January 25, 2023. The 2023 policy will require NIH funded researchers to prospectively submit a plan outlining how scientific data from their research will be managed and shared. The current data sharing policy is available here. For more information about the 2023 NIH Policy for Data Management and Sharing, please refer to https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

- **NIH Data Sharing Policies by Institute**: https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html

- **Clinical Trials Business Office**
  http://web2.tch.harvard.edu/researchadmin/mainpageS2700P141.html

- **Data Use Transfer Questionnaire**
  http://web2.tch.harvard.edu/researchadmin/mainpageS2700P142.html

- **Office of Sponsored Programs**
  If you have questions about data sharing requirements for applications, JIT submissions, or progress reports, please contact your OSP grant officer.

- **Link to template Informed consent language (insert link)**
  You may refer to section on data sharing and adapt the template language as appropriate

Additional NIH Specific Information

NIH endorses the sharing of final research data to serve important scientific goals and expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. "Timely release and sharing" is defined as no later than the acceptance for publication of the main findings from the final data set.

Applications

All investigator-initiated applications with direct costs of $500,000 or more (excluding subaward F&A costs) in any single year are expected to address data sharing in their submissions. Additionally, some Funding Opportunity Announcements (FOAs) may require all applications to include data sharing information regardless of the dollar level. These applications should include a brief, one paragraph, description of how final research data will be shared, including any limitations, or explain why data sharing is not possible.

Applicants proposing projects that will **generate large scale genomic data** as detailed in the NIH Genomic Data Sharing Policy (GDS) will need to include the following information in their submissions:

- Cover Letter: Include a statement that the proposed project will generate large-scale human and/or non-human genomic data.
- Budget: As applicable, budgets may include costs, as appropriate, to prepare data for submission to repositories.
- Research Strategy Attachment: If accessing human genomic data from NIH-designated repositories (dbGaP) is known at the time of submission, briefly describe your plans and state your intentions to abide by the NIH Genomic Data User Code of Conduct.
- Resource Sharing Attachment: Briefly describe the data, plans for sharing, and include assurance that they conform to NIH policy. If the sharing of human data is not possible, provide a justification explaining why the data cannot be shared and present an alternative data sharing plan.

NIH recognizes that data sharing may be complicated or limited, in some cases, by institutional policies, local IRB rules, and local, State and Federal laws and regulations, including the HIPAA Privacy Rule. The rights and privacy of individuals who participate in NIH-sponsored research must always be protected. Thus, data intended for broader use should be free of identifiers that would permit linkages to individual research participants and variables that could lead to deductive disclosure of the identity of individual subjects. When data sharing is limited, applicants should explain such limitations in their data sharing plans. Investigators must exercise great care to ensure that resources involving human cells or tissues do not identify original donors or subjects, directly or through identifiers such as codes linked to the donors or subjects.

Just-In-Time Phase (JIT)
When applications are likely to be funded, NIH requests time-sensitive project information, which includes a Certification of IRB Approval. If the proposed project involves human subjects research, a certification to NIH that all non-exempt human subjects research has been reviewed and approved by an appropriate IRB must be submitted. Make sure to address data sharing plans and requirements in your IRB protocol(s); pending or expired approvals are not acceptable.

Projects proposing to **generate large scale genomic data** are also required to submit Institutional Certifications assuring the NIH that the data submission and plans for sharing is appropriate and consistent with the GDS policy as well as applicable laws and regulations. Research Administration, including the Office of Sponsored Programs and the Institutional Review Board, will coordinate efforts to provide these certifications and will request the following information from BCH research teams:

- The date range of when the samples/data were used (before or after the 01/25/2015 GDS implementation date)
- Describe what data will be included in the dataset submitted to the NIH-designated repository
- Describe how the dataset will be de-identified
- Copies of all consent forms used for the prospective collection of genetic information with identification of:
    - Relevant language that discusses whether data can be shared
    - Appropriate research use of the data
    - Any limitation on the use of the data
- The applicable IRB protocol number(s)

Researchers seeking Institutional Certifications should complete the Application to Submit NIH-designated Genomic Data Repository form and email it to OSP.

Annual and Final Research Performance Progress Reports (RPPR)
NIH expects that Principal Investigators and grantee organizations will make the results and accomplishments of their activities available to the research community and to the public at large. If the grant is funded, the resource sharing plan, including formal plans for sharing final research data, genomic data sharing, and/or other project-specific data, will be referenced as a Special Terms and Condition in the Notice of Award. BCH investigators will be required to include updates on their progress in implementing the applicable data sharing plan(s) as part of their annual and final progress report (RPPR) submissions. If the sharing plan is fully implemented, provide a final statement on data sharing.

Related Content

IRB Policy

NIH Genomic Data Sharing Policy and Procedures

Document Attributes

Title	Data Sharing Requirements for Individual Participant Level Data		
Author	Susan Kornetsky	**Dates Reviewed/ Revised**	8/17/2021
Reviewed/ Revised by	Susan Kornetsky Theresa Applegate Fernando Valles Kris Bolt		
Copyright	©Boston Children's Hospital, 2021	**Last Modified**	8/17/2021
Approved	Susan Kornetsky, MPH Director, Clinical Research Compliance August Cervini, MBA Vice President, Research Administration		