

February 22, 2022

Lawrence Tabak, DDS, PhD
Office of The Director
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Submitted at <https://osp.od.nih.gov/rfi-updating-the-nih-genomic-data-sharing-policy>

Re: NOT-OD-22-029

Dear Dr. Tabak,

The Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard (MRCT Center) appreciates the opportunity to comment on the National Institutes of Health (NIH) draft “*Request for Information on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing Policy*,” published in the Federal Register on November 30, 2021. It is a timely and important proposal.

The MRCT Center is a research and policy center that addresses the ethics, conduct, oversight, and regulatory environment of international, multi-site clinical trials. Founded in 2009, it functions as a neutral convener to engage diverse stakeholders from industry, academia, patients and patient advocacy groups, non-profit organizations, and global regulatory agencies. The MRCT Center focuses on pre-competitive issues, to identify challenges and to deliver ethical, actionable, and practical solutions for the global clinical trial enterprise. Over the last five years, the MRCT Center has been intimately involved in data sharing, including (1) developing guidance for sharing aggregate plain language summaries for participants and the public, (2) developing guidance for sharing individual results with participants, (3) promoting principles of individual participant data (IPD) sharing including protections of patient/participant confidentiality and privacy of confidential commercial information, (4) developing template data use agreements and data contributor agreements for IPD and other data sharing, (5) crafting informed consent language to promote participant understanding of the implications of sharing de-identified data, (6) launching Vivli, a platform for global data sharing of IPD data, (7) furthering the establishment of credit for data sharing for those individuals who choose to share their data, among other efforts, and (8) working in collaboration with NIH, European universities, learned societies, and American research institutions in trying to establish legal pathways for the transnational shipment and secondary research uses of personal data under the EU General Data Protection Regulation (GDPR). Of note, the responsibility for the content of this document rests with the leadership of the MRCT Center, not with its collaborators, nor with the institutions with which its authors are affiliated.¹

¹ Brigham and Women's Hospital, Mass General Brigham HealthCare, Ropes & Gray LLP, Harvard Medical School, Harvard University, and Yale Law School.

The MRCT Center appreciates the NIH request for information and the efforts to update its policy on Genomic Data Sharing (GDS) to maintain currency with evolving technology and understanding.

- De-identification and evolving concept of “identifiability”

The MRCT Center supports the proposal to add the expert determination method (i.e., 45 CFR 164.514 (b)(1) of the HIPAA Privacy Rule) as an acceptable form of de-identification. This revision would have the additional advantage of harmonizing the HIPAA Privacy Rule with the NIH GDS policy, thus overcoming a current challenge for institutional compliance. This is especially the case given that the expert determination method is increasingly used for research studies that require certain identifiers (e.g., zip codes) that cannot be included in safe harbor de-identified data sets.

We are aware of additional privacy safeguards, such as differential privacy, blockchain, and other technologies that allow greater security and, in some cases, participant control of the use and/or granularity of their information. We encourage NIH to explore these technologies and, further, permit their use whenever the same or greater levels of privacy and security are demonstrated compared to HIPAA provisions (i.e., removal of 18 identifiers or expert determination method). NIH should consider funding opportunities for the continued development, testing, and dissemination of privacy and security methods, including research on participant understanding and preferences.

We support the changes to the GDS policy that are proposed in reference to the submission, sharing, and use of potentially identifiable information, under certain conditions:

- (1) If individual informed consent *has* been obtained from the participant, and there are no foreseeable third-party risks (e.g., to immediate or extended family members), there should be:
 - a. No limit to the *submission* of data to a qualified repository
 - b. No limit to the *use* of such information, providing:
 - i. A Data Access Committee and/or IRB has reviewed the original informed consent, and
 - ii. A Data Use Agreement has been executed, and
 - iii. Appropriate safeguards are in place (see below), and
 - iv. If the proposed research involves sensitive data or if the proposed research relates to vulnerable populations or discrete and insular communities, an IRB has reviewed and approved the research in advance of its performance.
- (2) If individual informed consent *has not* been obtained, then all the above requirements apply, but we recommend that IRB review and approval of the research be required (rather than recommended) or IRB waiver of the requirement to obtain informed consent (i.e., 45 CFR 46.116(f)), regardless of the sensitivity of the data or the nature of the research, in advance of its commencement.

- (3) For data collected before the NIH adopted the GDS policy or its proposed revision, consent for data sharing was in most cases not obtained, but in the MRCT Center's view, this should not prevent the sharing of previously collected data if those data are de-identified or are shared under the Limited Data Use criteria of HIPAA.

We recommend, additionally, that if any "higher resolution data" are used, those data should conform to the definition and be subject to the protections of HIPAA's Limited Data Set, which requires entry of a data use agreement meeting the requirements set forth at 45 CFR 164.514(e)(4) with the data recipient. If additional data are considered for use, specific participant protections should be in place, and the research should be reviewed by an IRB in advance of its commencement, as above.

In all scenarios, appropriate security standards for transmission and storage of data should be required.

- Data Linkage

There is a balance between scientific utility of data and individual privacy interests. Scientific utility increases when data linkage between different data sets that contain common patients or research participants is permitted, including significant opportunities for advances in public health and diagnosis, treatment, and prevention of disease and conditions that affect humankind. At the same time, data linkage increases the risk of reidentification. Given the potential benefit of research enabled through data linkage, MRCT Center supports the proposal that data linkage be allowed under certain conditions:

- (1) consent for research use has been obtained and the data are deidentified, or as an alternative to the informed consent requirements, a broad consent for use of identifiable private information or identifiable biospecimens has been obtained,
- (2) a data use agreement is executed,
- (3) a certificate of confidentiality is issued, and
- (4) the research is conducted in a controlled access environment from which results, but not data, can be exported.

If consent for research use has not been obtained (as, for example, with data collected before the NIH has adopted any new standards), then data linkage should be permitted under the following additional conditions:

- (5) a waiver to obtain informed consent has been provided by an IRB (i.e., 45 CFR 46.116(f)),
- (6) the proposed research has been reviewed and approved by a Data Access Committee, and
- (7) the proposed research has been reviewed and approved by an IRB.

The MRCT Center agrees with the proposal that the risks and benefits of data linkage should be part of obtaining consent for sharing and future use of data. It is clear, however, that the range of potential uses of data cannot be predicted, nor even envisioned, at the time of consent. It is for this reason that the additional protections above are offered.

We further recommend that an ethical framework for IRB (and Data Access Committee) review and approval of the sharing and use of deidentified data should be developed by a convened expert group that includes patient and public representatives of appropriate diverse populations. This expert group should be coordinated with the Office for Human Research Protections and the US Food and Drug Administration and provide guidance.

- Data management and sharing principles for NIH-supported resources

The MRCT Center agrees that the principles enumerated within the RFI are reasonable if they are applied specifically for NIH-supported resources. Those principles include requiring a data submission agreement consistent with Section IV.C.5; a Data Access Agreement consistent with Section V of the GDS Policy; compliance with the “[NIH Security Best Practices](#)”; systems for user authentication (e.g., [eRA Commons ID](#)) with procedures in place for handling data management incidents (DMI); data security procedures (e.g., FISMA, FedRAMP, and Moderate Authority to Operate (ATO)), and requirements in place that the repository or platform comply with the “[NIH Security Best Practices](#)” as applicable.

- Harmonization of GDS and NIH Policy for Data Management and Sharing (DMS) Plans

The MRCT Center believes that harmonization of GDS and the NIH Policy for Data Management and Sharing (DMS) policy is important for the regulated community. The DMS plans should be submitted prior to proposal review, not at funding or Just-in-Time, and should be reviewed by the grant review committee

The RFI requests comments on the timeline for data sharing, since the expectations differ between the NIH GDS and DMS policies. On the one hand, the submission of cleaned data within three months of data generation is required, while on the other, data must be shared not later than the time of publication or end of the performance period for unpublished data, whichever comes first. The MRCT Center believes that the first is potentially too short (but only potentially too short in the instance of genomic sequencing), while the second is ambiguous, vague, and imprecise. In addition, “reasonable” timing depends in part on the nature of the data being submitted, in that DNA sequence differs from clinical trial or EHR data relevant to the results of a study, and it differs depending upon the nature of the research (e.g., data relevant to pandemic responsiveness). If a timeline triggered by study completion is considered, the MRCT Center recommends that NIH account for situations in which there is no clear “study completion date,” and that alternative triggering events (e.g., submission for publication, current award end date) be substituted. Such situations include long-term studies (e.g., Framingham Heart Study), registries, adaptive trials, platform trials, and others.

Whatever final timeline is chosen, the MRCT Center recommends that:

- (1) NIH clearly identify the person or entity responsible for data submission.
- (2) A process for requesting exceptions be established, and that such process involve written documentation and NIH approval and be understood to be rare.
- (3) Submitted data be annotated to indicate whether and how they have been validated and cleaned, and that there be a mechanism not only to add additional data but to correct data that are subsequently modified, all with appropriate metadata.
- (4) NIH post consequences for failing to abide by its policies (see below).

The RFI requests comments on harmonizing the GDS and DMS policies related to non-human genomic data. Currently, the GDS policy is less rigorous for non-human than human data: the timeline for non-human data is the time of initial publication (versus within three months of data generation) and, further, the NIH is proposing to conform to the expectations for “scientific data” in the DMS policy (e.g., “data of sufficient quality to validate and replicate findings”). The DMS policy is much less stringent and, we believe, will have a significant negative impact on the utility of the shared data. The MRCT Center does not believe that fewer data should be shared and would strongly oppose diminishing the expectations for sharing to those data necessary to validate and replicate findings.

- Types of research covered by the GDS Policy

There are many additional data types that have value and are currently beyond the current scope of the GDS Policy, some of which are potentially sensitive (e.g., proteomic and microbiomic data, electronic health record (EHR) data). However, we recommend that the term “sensitive data” be clearly defined in the GDS Policy. The MRCT Center supports inclusion of these additional data types in data sharing requirements and believes that sharing and use of these data warrant protections, including:

- (1) Review by the data submitter for concordance with informed consent provisions, and
- (2) IRB and/or Data Access Committee review of risks associated with submitting data to NIH, even when data are de-identified, and
- (3) IRB review and approval of the proposed research if consent has not been obtained.

We agree that small scale studies should be covered under the GDS Policy, particularly for rare diseases, rare patient populations (e.g., pediatric patients), and other situations. However, we recommend caution in including other types of studies, given the lack of consensus in the identifying nature of certain data types. The use of the data from the inclusion of these populations should be subject to the additional protections discussed above, including the use of controlled access environments. Similarly, training and development awards (e.g., F, K, and T awards) should be covered by the GDS Policy if the award covers the research project but not if only salary and benefits are provided (and see next paragraph).

The MRCT Center believes that NIH-funded research that generates large-scale genomic data but in which NIH's funding does not directly support genomic sequencing should be subject to the GDS policy if:

- (1) NIH funds supported the collection, annotation, processing, or storage of the data;
- (2) Sequencing the data was proposed in the grant application for funding; or
- (3) Any part of subsequent use of the sequenced data is supported by the NIH.²

- Other Considerations and comments

The MRCT Center recommends that the NIH consider several additional issues in its revision to the GDS policy.

- (1) Withdrawal of consent:

The MRCT Center encourages the NIH to clarify what will happen to data in the event of participant withdrawal of consent, especially because with only de-identified data, it will in many cases not be possible to re-identify the individual to whom the genomic data pertain.

- (2) Third party risks:

Risks to persons other than the individual from which data were derived, such as immediate and extended family members, are well appreciated and will only increase with increased availability of data. We recommend that NIH consider deidentification of data relating to potential family members, and that protections are extended to immediate and extended family members.

- (3) Harmonization with other federal and international agencies

The costs and challenges of compliance with the GDS policy are increased insofar as the GDS policy differs from the policies of other federal or international agencies. Efforts to harmonize this policy with other US and ex-US policies should be considered. Harmonization or, at a minimum, lack of inconsistency (so that the policies enable data sharing) will promote scientific discovery. The methods involved, considering new methods of data protections and privacy, controlled access provisions, etc., will need to be considered; a higher bar than "readily identifiable" to help ensure privacy, confidentiality, and protections from re-identification may need to be imposed.

- (4) Inability to "future-proof" re-identification:

² If the researcher making the secondary use supported by NIH funding differs from the researcher who initially generated the data without NIH support, the researcher making the secondary use should be subject to the data sharing requirement.

We are also concerned about the challenge of attempting to predict the risk of reidentification in the future. The concern undergirds our suggestion that future research be subject to review and approval by a Data Access Committee and/or IRB committee before the time of the research.

(5) Sanctions for misuse of data:

As data become more granular, data linkage is permissible, and additional sources of data become more available, the risks of reidentification are significantly increased. In addition to the suggestions above, we recommend sanctions (e.g., fines, debarment) and/or criminal liability for those individuals or institutions who share or misuse data inappropriately, knowingly, intentionally, or negligently. NIH GDS should work with the other branches of government to invigorate enforcement provisions, civil and criminal liability for such actions.

(6) Public education:

Finally, the MRCT Center recommends that NIH endeavor to promote a public educational campaign to illuminate the benefits of data sharing and data linkage, while informing the public of (a) the potential small risk of reidentification and (b) any potential recourse in the event of personal harm.

(7) Effectiveness and impact of the GDS policy:

As the spirit of this guidance is to encourage increased sharing of data, it would be helpful for NIH to identify what metrics NIH is using or will use to determine effectiveness and impact of the policy.

Thank you again for the opportunity to comment on this important issue. We believe that the NIH is in a unique position to promote genomic data sharing for the public good, but only if NIH uses this opportunity to advance the culture of, and infrastructure to support, data sharing coupled with adequate participant protections and engagement.

We are available to discuss our comments with you if that would be helpful and would be happy to work with you on any of the aforementioned items. Please feel free to contact the MRCT Center at bbierer@bwh.harvard.edu, sawhite@bwh.harvard.edu, and mark.barnes@ropesgray.com.

Respectfully submitted,

Barbara E Bierer, MD
Faculty Director, MRCT Center

Mark Barnes, JD, LL.M.
Faculty Co-Director, MRCT Center

Sarah A White, MPH
Executive Director, MRCT Center

David Peloquin, JD
Senior Advisor, MRCT Center