February 28, 2022

Lawrence A Tabak, DDS, PhD
Acting Director
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Re: NIH Request for Information (RFI) on Proposed Updates and Long- Term Considerations for the NIH Genomic Data Sharing Policy

Comments submitted electronically via https://osp.od.nih.gov/rfi-updating-the-nih-genomic-data-sharing-policy

Dear Acting Director Tabak:
The American Medical Informatics Association (AMIA) appreciates the opportunity to respond to the RFI on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing Policy.

AMIA is the professional home for more than 5,500 informatics professionals, representing frontline clinicians, researchers, and public health experts who bring meaning to data, manage information, and generate new knowledge across the health and healthcare enterprise. As the voice of the nation's biomedical and health informatics professionals, AMIA plays a leading role in advancing health and wellness by moving basic research findings from bench to bedside, and evaluating interventions, innovations and public policy across settings and patient populations.

Eight AMIA member experts formed a work group to prepare this response on updates to the NIH Genomics Data Sharing Policy, and its content was reviewed by the AMIA President and Board Chair. AMIA has provided direct responses to many of the RFI provisions in the table below. In addition, AMIA members highlighted several overarching principles that must be addressed in an evolving regulatory framework for a Genomic Data Sharing Policy of the future.

1) Change Data Sharing “Expectations” to “Requirements”

There is a need for a paradigm shift in how data sharing is defined and managed across its lifecycle. The new paradigm requires stronger protections and requirements for the manner by which genomic information is collected, used, analyzed, stored, disclosed, transferred, and reused. This paradigm shift is essential to ensure that a person’s privacy and autonomy are preserved in any data sharing policy. AMIA believes that a person’s privacy protections must be consistently and continually maintained, with privacy preferences respected across clinical, research, community services, and commercial settings.

One component of such a paradigm shift is the necessity for stronger language to indicate NIH’s duty, authority, and intent to protect a person’s genomic information. For example, in the RFI section on Expectations for Alternative NIH-Supported Genomic Data Management and Sharing Resources that Store Human Genomic Data, AMIA strongly encourages NIH to replace “Expectations” with “Requirements.” Under each subsection on Data Submission, Data Access, and Data Security, “Should” would appropriately be replaced with “Must.”

2) Audit and Enforcement Policy

A paradigm shift would also entail a strong audit and enforcement policy with stringent civil and criminal penalties (as allowed by law) for violations of the NIH data sharing policy by individual investigators, institutions, data recipients, or other applicable entities. We encourage the NIH to promulgate such a policy specific to genomic data management and to advise on legislation as needed to implement the policy.

3) Informed Consent

Informed consent requires clearly worded, understandable explanations of how a person’s health data will be used and the circumstances in which it will be disclosed. In this RFI, NIH states:

> “The GDS Policy will continue to provide expectations regarding consent for broad sharing and future use of human genomic and phenotypic data.”

Again, AMIA urges NIH to change this language to state the policy will provide “requirements” regarding consent for broad sharing and future use of human genomic and phenotypic data.

Further, AMIA urges NIH to establish a comprehensive platform to facilitate a person's ability to manage longitudinally consent for use of genomic data across all entities with access to the data. Such a platform would foster efficient, dynamic access to the individual's signed consent forms and provide an expeditious pathway for a person to revoke consent for data use at any point in time.

4) Data Policy Harmonization
All approaches need to be in alignment with, but not limited, to the overall NIH Data Sharing Policy, and where there is an international dimension, with applicable policies.

Thank you for your consideration of these comments. Should you have any questions or require additional information, please contact Tanya Tolpegin, AMIA Chief Executive Officer, at ttolpegin@amia.org.

Sincerely,

Gretchen Purcell Jackson, MD, PhD, FACS, FACMI, FAMIA
President and Board Chair, AMIA
Vice President & Scientific Medical Officer, Intuitive Surgical
Associate Professor of Surgery, Pediatrics, and Biomedical Informatics, Vanderbilt University Medical Center

Please look at pages below (in landscape for tabular response).

AMIA Responses noted on right side of chart

Request for Information on Proposed Updates and Long-Term Considerations for the NIH Genomic Data Sharing Policy	
Notice Number:	NOT-OD-22-029
Key Dates	
Release Date:	November 30, 2021
Response Date:	February 28, 2022
Related Announcements	
NOT-OD-14-124	NIH Genomic Data Sharing Policy
NOT-OD-21-013	Final NIH Policy for Data Management and Sharing
Issued by	Office of The Director, National Institutes of Health (OD)
Purpose	NIH is seeking public input on potential updates to the NIH Genomic Data Sharing Policy to keep pace with evolving scientific opportunities and stakeholder expectations.

Background	
The NIH Genomic Data Sharing (GDS) Policy (NOT-OD-14-124), issued in 2014, set forth expectations for ensuring the broad, responsible, and timely sharing of genomic research data generated from NIH-funded or conducted research. A landmark policy at the time, the GDS Policy focused on striking an appropriate balance between accelerating scientific research through rapid genomic data sharing and minimizing risk through formalizing expectations of informed consent and appropriate privacy protections. The GDS Policy has served the research community well, facilitating tens of thousands of genomics studies while preserving public trust in the biomedical research enterprise.	A number of issues arose contemporaneously with the launch of the GDS Policy and since then, putting many assumptions into question. For example, even in 2013 the journal *Science* carried a paper on "surname inference," highlighting the fact that "Surnames are paternally inherited in most human societies, resulting in their cosegregation with Y-chromosome haplotypes," with the consequence that "... surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases."[Gymrek] An example that made national headlines and muddied the ethical waters was that of the so-called Golden State Killer, former policeman Joseph James DeAngelo Jr., whose identity was discovered through deep DNA sleuthing combined with some basic police methods. "... the DNA-matching effort that caught [DeAngelo] was more extensive than previously disclosed and involved covert searches of private DNA housed by two for-profit companies despite privacy policies, according to interviews and court discovery records accessed by The Times."[LA Times]

	My genome, data that uniquely defines me, should not become an entity's intellectual property and I should have a right to revoke its use by a researcher or third party. Consent for sharing genomic data should be considered temporal, subject to change by the person whose data it is. Structures should be in place to manage consent in an ongoing manner. Informed consent would include the risks associated with data sharing as well as a list of all third parties that will receive the data in managing the research. If the risks or third parties change, notice must be sent to persons who have given consent to allow them to opt out of further sharing. For example, if a researcher wanted to migrate the data from a smaller software entity that provided genomic management support to a competing Google service that was less expensive or even "free", every person who had previously granted access should be given the opportunity to say "don't share my data with Google". When there is no payment for the product, the data is the payment, and the risk profile changes if a person's data is in line to be monetized by Google. [Gymrek] Melissa Gymrek, Amy L. McGuire, David Golan, Eran Halperin, Yaniv Erlich *Identifying Personal Genomes by Surname Inference.* **Science** 2013 https://www.science.org/doi/10.1126/science.1229566

	[LA Times] Man in the Window https://www.latimes.com/california/story/2020-12-08/man-in-the-window
While the principles underlying the GDS Policy remain relevant for research today, genomic sequencing and related technologies are now considered integral to the conduct of biomedical research. Moreover, data sharing is widely recognized as a best practice for advancing research and the promise of societal benefit continues to evolve. While NIH has adjusted implementation of the GDS Policy to keep pace with these changes, several key developments affecting the conduct of NIH-supported genomic research warrant reassessment of aspects of the GDS Policy. These developments include:	

• A growing interest in using information with a potentially higher degree of identifiability, especially in combination with other data types, than is currently allowed to be shared, such as granular location or date of treatment information;	While this is arguably true for research, two observations must be juxtaposed here: (1) the risk of re-identification is now greater than ever, and (2) there are privacy preserving federated methods in advanced development that obviate the need for data to be aggregated in one place.
• An increasing capability to link participants' data from diverse datasets, such as electronic health records, with genomic information, thereby creating new opportunities – and challenges – for ensuring records linkage techniques sufficiently account for and respect consent, manage risk, and preserve privacy;	A taxonomy of previous attacks in a chart form with examples has been given in Bonomi, L., Huang, Y. & Ohno-Machado, L. *Privacy challenges and research opportunities for genomic data sharing.* Nat Genet 52, 646–654 (2020). https://doi.org/10.1038/s41588-020-0651-0
• The release of the new NIH Data Management and Sharing (DMS) Policy,[i] effective January 2023, which sets additional expectations for managing and sharing scientific data, including those data subject to the GDS Policy, by expecting the development of Data Management and Sharing Plans for all NIH-supported research; and	As noted already, *expectations* seem to leave too much room for interpretation, especially if formulated as "just-in-time" elements in funding applications. These should be made into firm requirements.
• The continued development of novel data types with high scientific utility that may be equally as sensitive as genomic data (e.g., proteomic or metabolomic data) but are not currently subject to the GDS Policy's protections.	We concur with this observation and in some circumstances would add aspects of Social and Environmental Determinants of Health data if coupled with geolocation information.

NIH remains committed to the principles espoused by the GDS Policy of maximizing scientific advances and public benefit by sharing genomic data and associated phenotypic data in a manner consistent with participants' informed consent. However, in an effort to ensure NIH policies keep pace with evolving scientific opportunities and stakeholder expectations, NIH is seeking public feedback on how to ensure the GDS Policy remains consistent with this changing landscape. Note that while potential updates are under consideration, the GDS Policy will remain in effect in current form until further notice.	
Request for Input	
I. Maximizing Data Sharing while Preserving Participant Privacy and Preferences	

Respect for and protection of the interests of research participants are central tenets of the NIH GDS Policy and are fundamental to NIH's stewardship of large-scale genomic data. Data derived from human research participants under the GDS Policy must be de-identified and provided with a random, unique code, the key to which is held by the submitting institution. NIH acknowledges that the concept of "identifiability" is a matter of ongoing deliberation within the scientific and bioethics communities. NIH relies on robust protections beyond de-identification, such as Institutional Review Board (IRB) consideration of risks associated with data submission, designating controlled access for certain data types, use of Data Access Committees to review requests, data use agreements to prohibit data disclosure and participant re-identification, and Certificates of Confidentiality[ii] to prohibit disclosure. As outlined in the NIH GDS Policy, the criteria for establishing de-identification are:	
• Identities of research participants cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 CFR 46.102(e) (Federal Policy for the Protection of Human Subjects); and	We understand the "expert determination" rule to be rather more stringent than is suggested by "... identities … cannot be readily ascertained … ." The statistical probability of re-identification must be insignificant even in the presence of complementary datasets from other sources.
• 18 identifiers enumerated at 45 CFR 164.514(b)(2)(the HIPAA Privacy Rule) are removed.	We note that the 18th identifier is "Any other element" that may identify a person. In the context of genomic data this

	has to be understood broadly, given that a person's genome may also identify close family members.
The reliance on the 18 identifiers enumerated at 45 CFR 164.514(b)(2) (the HIPAA Privacy Rule) as the only acceptable method under the GDS Policy for de-identification has recently presented several challenges. Certain data elements considered potentially identifiable, such as date ranges shorter than a year, may have scientific utility, especially when studying disease progression (e.g., with COVID-19) or higher resolution location data than the regulatory standard (e.g., full ZIP codes or mobile location data), which may be valuable for studying the social determinants of health or environmental risk.	We observe that even when dates are shifted, date differences provide a strong "candidate key" to a patient record, making re-identification possible when aligned with other data sets.

Challenges have also arisen recently around data linkage. It is difficult to know in advance which data sources may add scientific value when combined, so it is not always possible to tell participants about data linkage during their initial consent. Linking data refers to connecting two or more data sources (often multiple studies) to bring together information about a person, enabling researchers to learn more about a participant or small group of participants. For example, a participant might enroll in a study that uses their electronic health record as well as a separate study that uses a sample of their blood, and the data about them from those studies could later be linked in new research for more powerful analyses. This challenge in prospectively informing participants about data linkage raises questions about respecting individuals' autonomy and what participants understand about how their data will be used. Furthermore, data from multiple sources may not have been obtained under the same consent and de-identification expectations as the GDS Policy.	
NIH seeks input on:	

1. **De-identification.** The risks and benefits of expanding de-identification options, including adding the expert determination described at 45 CFR 164.514 (b)(1) (the HIPAA Privacy Rule), as an acceptable method for de-identification under the GDS Policy, and whether other de-identification strategies exist that may be acceptable in lieu of HIPAA standards.	Something here about differential privacy? I.e. that some mechanisms at preventing identification are more powerful than others, in particular more powerful than simply stripping certain fields. Proactive "white hat" activity to seek external datasets that would reveal the identity of a person if they were fused with personal health or genomic data. This should be considered an integral part of sound expert determination.
2. **Use of potentially identifiable information.** The circumstances under which submission of data elements considered potentially identifiable to repositories under the GDS Policy would be acceptable, any additional protections (including for security) that would be warranted, and whether there is certain potentially identifiable information that would not be acceptable to submit.	The NIH needs to acknowledge that there is almost always potential for re-identification of information and data elements if they are accessible by those who desire to re-identify the data. This risk carries implications not only for those that consent to participate, but also for their immediate family, particularly for their descendants. Therefore, the focus is on tight data access criteria, permissions, and strict, explicit limits on additional sharing from those who do have permission with anyone else for any use. A plan for auditing access and any additional sharing with others is necessary along with stringent enforcement of penalties. In addition, entities that use such information must clearly state in the consent form that confidentiality cannot be guaranteed by the data user and should not be expected by the individual providing consent.

3. **Data linkage.** Whether the GDS Policy should permit data linkage between datasets that meet GDS Policy expectations (e.g., data obtained with consent for research use and de-identification), and whether the GDS Policy should support such linkages to datasets that do not meet all GDS Policy expectations (e.g., data may have come from a clinical setting, may not have been collected with consent, may retain certain potentially identifiable information). Feedback is also requested on risks and benefits to any such approaches.	Any genomic data linked to datasets that meet GDS Policy expectations should be required to meet those expectations moving forward, or the flow of data should not be bi-directional. Periodic audit should also form part of the process.
4. **Consent for data linkage.** Whether data linkage should be addressed when obtaining consent for sharing and future use of data under the GDS Policy, as well as in IRB consideration of risks associated with submission of data to NIH genomic data repositories. And if so, how to ensure such consent is meaningful.	The policy must make it explicit what sorts of consent mechanisms for re-use, de-identification, and linkage are acceptable. To wit: A recent situation discussed in the AMIA Ethical, Legal, and Social Issues WG, as having been discussed in HISTalk on January 31, 2022, is as follows: Politico reported that a Crisis Text Line, a non-profit behavioral health entity that uses ML/AI shares its anonymized text conversations with a for-profit spinoff that sells customer service software. A 50-paragraph disclosure allows user data to be shared without further consent, including with Facebook Messenger. The former CEO stated that text conversations are predictive of self-harm activities, sexual orientation status, and assesses users on a variety of topics from COVID-19 status to various mental health conditions. It moves those “at risk” to the top of a cue for help within 39 seconds.” Deep concerns among

	the ELSI WG members are that a 50-paragraph consent, while legal, is likely to be disregarded and/or misunderstood by those under duress, so they may miss the statement that their information *will be sold* to others for profit. Those others may include entities that have the power to make linkages across many data sets to not only re-identify users, but to sell that re-identified, linked data in ways that would jeopardize employment, insurability, etc. Genetic data alone can identify conditions that can and are used to discriminate against people, and so we want to prevent any situation where further linking of data, for example to Social Determinants of Health/risk factor data elements, creates a pool for “redlining” against people for any purpose. An analysis of the strategy of the most prominent corporation offering a privacy-preserving record linkage (PPRL) solution suggests that it sees its business as that of a data aggregator. Without doubting their good intentions, we would ask whether it is wise to allow an entity that should function as an honest identity broker also to hold personal health information.

Anonymity Compromised

The balance between maintaining individual privacy and sharing genomic information for research purposes has been a topic of considerable controversy. Gymrek *et al.* (p. 321; see the Policy Forum by Rodriguez *et al.*) demonstrate that the anonymity of participants (and their families) can be compromised by analyzing Y-chromosome sequences from public genetic genealogy Web sites that contain (sometimes distant) relatives with the same surname. Short tandem repeats (STRs) on the Y chromosome of a target individual (whose sequence was freely available and identified in GenBank) were compared with information in public genealogy Web sites to determine the shortest time to the most recent common ancestor and find the most likely surname, which, when combined with age and state of residency identified the individual. When STRs from 911 individuals were used as the starting points, the analysis projected a success rate of 12% within the U.S. male population with Caucasian ancestry. Further analysis of detailed pedigrees from one collection revealed that families of individuals whose genomes are in public repositories could be identified with high probability.

Headpiece to Gymrek et al., Science

II. Expectations for Alternative NIH-Supported Genomic Data Management and Sharing Resources that Store Human Genomic Data	
The rapid advance of genomic technologies, available at increasingly accessible cost, has enabled a wealth of large-scale genomic data and other associated data types. NIH has traditionally provided substantial capacity to the community for storing and managing access to human genomic data under the GDS Policy through dbGaP and a small number of other NIH-operated repositories.	
To reduce the technical burden of analyzing genomic data, NIH has begun investing in a number of resources (i.e., beyond dbGaP) for storing, sharing, and analyzing human genomic and phenotypic data under the GDS Policy. These investments have resulted in an increasingly federated landscape of platforms and repositories, hosted both at NIH and awardee institutions. There is consequently a need to establish shared principles between NIH and external organizations that are supported by NIH to ensure that data protections are consistent with those provided by dbGaP and the terms of the GDS Policy.	

Accordingly, NIH proposes principles derived from the GDS Policy and dbGaP practices that have been used as criteria to ensure that NIH-supported alternative resources hosting human data generated and shared under the GDS Policy maintain appropriate standards and protections. Note that these principles would provide expectations only for NIH-supported resources, and NIH is not proposing at this time that sharing of human genomic data in non-NIH-supported repositories or platforms would satisfy the GDS Policy's expectations. These principles are also intended to be consistent with the criteria described in the supplemental information to the DMS Policy, "Selecting a Repository for Data Resulting from NIH-Supported Research" (NOT-OD-21-016). The principles include the following:	
Data Submission	
• Repository or platform should obtain a data submission agreement from the submitting institution that is consistent with the principles outlined in Section IV.C.5 of the GDS Policy [iii]	It is not clear that this is to protect the data rather than the data contributing institution. Replace "should" with "must".
Data Access	Recommend a paradigm that accounts for genomic data access, use, re-use and exchange.

	We need to address the transition, security and control of data during the sale, acquisition or transfer of data sets. Genomic data is at such a sensitive level, that the downstream use and re-use and exchange of the data takes on a different level of consideration. Data use, re-use and exchange across an ecosystem, over time – for research purposes, for commercial for profit use. What about data monetization?
• Repository or platform should execute a data access agreement with the requesting institution that is consistent with the principles outlined in Section V of the GDS Policy	Data access agreements must be easily available for public review, concise, and meet government standards for plain language to ensure that members of the public can understand them.
• Repository or platform should expect users to comply with the "NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy"[iv]	Language must be stronger here. It cannot be "should." It must state "Platform or repository users *must* comply with the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy"[iv]"
• Repository or platform should have systems for authentication of users (e.g., eRA Commons ID)	Language must be stronger here too, as above.

• Repository or platform should have procedures in place for handling data management incidents (DMI) (e.g., process to suspend users, penalty assessment criteria) and a communication plan to notify appropriate NIH staff of a DMI	Replace “should” with “must” to emphasize the importance of such procedures. Potential consequences need to be well defined, including Individual and institutional penalties, including criminal penalties. NIH should pursue any needed changes in law.
• Repository or platform should report data use statistics	Replace “should” with “must”. Use for research, for care and treatment innovation and for commercialization.
Data Security	
• Repository or platform should have FISMA[v] and FedRAMP[vi]Moderate Authority to Operate (ATO)	Replace “should” with “must.”
• Repository or platform should comply with the “NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy” as applicable	Replace “should” with “must”.
NIH seeks input on:	

5. **Data management and sharing principles for NIH-supported resources**	
1. Any aspect of the principles described for Data Submission.	
2. Any aspect of the principles described for Data Access.	NIH needs to strengthen the language throughout the data access points to require compliance with all restrictions regarding access to genomic data.
3. Any aspect of the principles described for Data Security.	The suggested standards should be the minimum required for security and the expectation should be that entities will proactively improve their Data Security protocols as stronger protections become available.
III. Policy Harmonization	

In October 2020, NIH released the **NIH Policy for Data Management and Sharing (DMS Policy)** to promote the management and sharing of scientific data generated from NIH-funded or conducted research. Please note that while it was released October 2020, it is not effective until January 25, 2023. The framework for DMS Plan submission and review, as well as specific considerations for data managing and sharing practices, shall also be the default practice for those proposing research that is subject to the GDS Policy. To this effect, NIH intends to harmonize the GDS Policy and GDS Plan elements, submission, and review with the DMS Policy. Harmonization of the GDS and DMS Policies will ensure consistency in data sharing and management expectations, reduce administrative burden on the scientific community, and streamline and enhance compliance with NIH data sharing policies, while maintaining the principles of sharing large-scale genomic research data and protecting research participants' interests and privacy.	
To harmonize these policies, NIH proposes to make the following changes to the GDS Policy, GDS Plans, and GDS Plan submission and review:	

● Harmonization of GDS and DMS Plans: Under the GDS Policy, the NIH currently expects a GDS Plan to be submitted in grant applications or R&D contract proposals. To avoid researchers having to submit two plans when the DMS Policy becomes effective, NIH proposes that for research subject to the GDS Policy:	
○ There will be one plan. Plans for sharing genomic data will be reported in the DMS Plan submitted at time of funding application or proposal, and not in a separate plan or at Just-in-Time;	We do not support Just-in-time
○ Elements recommended to be addressed in DMS Plans, provided in the "Elements of an NIH Data Management and Sharing Plan" (NOT-OD-21-014), will be expected to also cover genomic data sharing considerations;	… with appropriate protections.
○ As expected by the "Update to NIH Management of Genomic Summary Results Access" (NOT-OD-19-023), DMS Plans will also indicate whether a study should be designated as "sensitive" for purposes of access to genomic summary results, and for applicable applications, should be reported in the Access, Distribution, or Reuse Considerations section of the DMS Plan; and	

○ As with the DMS Policy, the budget for genomic data management and sharing will be commented on during peer review, and NIH Programmatic Staff will assess the adequacy of Plans.	This is too weak. Understanding that in certain circumstances "just-in-time" proposals are the only option, there is still every reason to consider them a make-or-break element in a proposal, or failing that, subject to a formally negotiable agreement.
● Timeline for data sharing: The supplemental information to the GDS Policy[vii] provides expectations for the timeline of data submission and release based on the level of data processing (e.g., submission of cleaned data within three months of data generation). For human data, these timelines are generally shorter than the DMS Policy, which states that shared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication or the end of the performance period, whichever comes first.	Noting that reproducibility of results is a crucial criterion in science, it is still necessary to allow those conducting primary research the opportunity to publish first based on data they have generated. While this should not be used as a reason to block data sharing, there should be an actionable undertaking on the part of the recipients not to usurp this opportunity.
In some cases, the GDS Policy's earlier timelines for sharing have posed challenges for compliance. NIH seeks comment on harmonizing these timeline expectations by modifying GDS Policy expectations to be the same as DMS Policy expectations (i.e., no later than the time of publication or end of the performance period for unpublished data, whichever comes first). NIH Institutes, Centers, and Offices (ICOs) and programs will continue to be able to set earlier timelines for data sharing for specific projects if warranted.	"... no later than the time of publication or end of the performance period for unpublished data, whichever comes first)" appears to be encouraging hasty publication or failure to allow enough time for data to be adequately processed for publication.

• Alignment of expectations for non-human genomic data with the DMS Policy The GDS Policy applies to research generating human or non-human genomic data. The GDS Policy and the supplemental information indicate that non-human data are generally subjected to fewer sharing expectations than for human data (e.g., data are generally expected to be shared no later than the time of initial publication through any widely used data repository). To clarify and simplify the expectations for research generating non-human genomic data, NIH seeks comment on sharing non-human genomic data consistent only with the expectations for "scientific data" in the DMS Policy.[viii]	There should still be a principle that "life cannot be patented," i.e. meaningful fragments of DNA cannot be owned as property even by the scientists who discover or sequence it.
This approach may have the consequence of sharing less data due to the definition of "scientific data" under the DMS Policy, which focuses on data of sufficient quality to validate and replicate findings, rather than the more expansive definition provided in the supplemental information to the GDS Policy. Through this change, NIH seeks to simplify compliance and focus sharing expectations on non-human genomic data that underlie research findings. While NIH ICOs and programs would be free to set more stringent expectations, NIH seeks input on the potential negative impacts of sharing non-human data consistent only with the DMS Policy.	

Elements that will remain in the GDS Policy: The following expectations would remain in the GDS Policy because they achieve particular goals that are more specific than those outlined in the DMS Policy:	
• Scope: The GDS Policy will continue to apply to NIH-supported or conducted research that generates "large-scale" human genomic data as well as the use of these data for subsequent research;	
• Data expected to be shared: The NIH will continue to expect sharing of large-scale genomic data described in the GDS Policy, supplemental information, and further NIH ICO expectations. Note that input is requested on the realignment of non-human genomic data sharing expectations with the DMS Policy and that the data shared may differ from those that meet the definition of "scientific data" under the DMS Policy, which are those that are of sufficient quality to validate and replicate research findings;	
• Informed Consent: The GDS Policy will continue to provide expectations regarding consent for broad sharing and future use of human genomic and phenotypic data;	"Expectations" should be strengthened to "requirements." There is a need for consent to be revocable, with the ability to remove data from data sets. Derivative products, e.g. AI models build on a data set from which records are subsequently removed on request by the individual, raise

	sticky questions as well as challenging technical and logistical capabilities. Consider a recommendation for publicly accountable and transparent infrastructure that supports a permanent consumer portal in these platforms, for informed consent management (including examples like SaaS), directory of location of bio-specimens, For example, children or elder adults, in health transitions having the support to transition their research over time. Consent should be time limited. Consumers should be contacted periodically to for re-consent
• Institutional Certification: The GDS Policy will continue to expect Just-in-Time submission of an Institutional Certification for human genomic data submitted to NIH supported data repositories;	

● Repository specifications: The GDS Policy will continue to articulate expectations for repositories (see also the "Expectations for Alternative NIH-Supported Genomic Data Management and Sharing Resources that Store Human Genomic Data" section for further discussion of these expectations);	
● Responsibilities for Investigators Accessing and Using Genomic Data: The GDS Policy will retain the expectations for requests for controlled-access data based primarily on the informed consent under which the data or samples were collected, and the terms and conditions for future research use of controlled-access data;	
● Intellectual Property: The GDS Policy will continue to encourage broad use of NIH-funded genomic data consistent with a responsible approach to management of intellectual property derived from subsequent discoveries; and	NIH should explore licensing, contractual, and IP models that permit some form of retention of downstream rights by the individuals whose data was used to create a valuable product, and/or society at large. This might take the form of royalty payments to a public trust, under the stewardship of NIH for example.
● Enforcement and Compliance: The GDS Policy will retain provisions for enforcement of the Policy as a term and condition of award.	

Additional changes to these provisions may be made to clarify or simplify language, harmonize with DMS Policy terminology or practices, or to reflect comments received from this request for information or other sources.	
NIH seeks input on:	
6. **Harmonizing GDS and DMS Policies.** Any aspect of the approach to harmonize GDS and DMS Policies and Plans described above, including for non-human genomic data.	
7. **GDS and DMS data sharing timelines.** Whether the continued use of earlier submission expectations for human genomic data in the GDS Policy (e.g., submission of human data within three months of data generation) is needed, or whether timelines should be harmonized with the DMS Policy expectations (i.e., sharing of data no later than the time of publication or at the end of the performance period, whichever comes first), as described in the proposal above.	
IV. Long-Term Consideration of the Scope of GDS Policy	

NIH recognizes that data types and analytical methods have advanced since the release of the GDS Policy in 2014. In some cases, non-genomic data types (e.g., proteomic and metabolomic data) may pose similar risks of re-identification as large-scale human genomic data and may warrant the additional protections afforded by the GDS Policy, such as the Policy's specific de-identification expectations. Furthermore, institutions submitting human genomic data to NIH repositories are to review associated informed consent materials via IRBs or equivalent bodies and provide an Institutional Certification to the funding NIH ICO. These same protections or sharing expectations could potentially be applied to other specific high-value and/or potentially sensitive data types. Additionally, because the scope of the GDS Policy (e.g., large-scale) does not apply to certain studies, the protections of the GDS Policy discussed here are not uniformly applied.	

With the implementation of the DMS Policy, NIH will soon expect researchers to maximize appropriate sharing of scientific data. However, some of the GDS Policy's expectations for the level of data to be shared and the speed of data sharing will go beyond those expected under the DMS Policy. While the DMS Policy outlines the scope of data sharing in terms of those data needed to validate and replicate research findings, the GDS Policy refers to the submission of large-scale genomic data and associated phenotypic data based on level of processing. The value of this volume of data, and its potential reuse for a multitude of additional analyses, were key factors in establishing this sharing expectation, but as large-scale data become more common, there may be other data types that possess similar value for advancing NIH's mission.	
As stated in the GDS Policy, "[a]t appropriate intervals, NIH will review the types of research to which this Policy may be applicable." As such, NIH seeks input on whether the protections of the GDS Policy should apply to research involving additional data types, and whether the expectations for the level and speed of data sharing are warranted for such research that would not otherwise be satisfied by the DMS Policy's expectations.	

As stated in the GDS Policy's preamble, the Policy applies to research funded in part or in whole by NIH if NIH funding supports the generation of the genomic data. To ensure collaborations are consistent with the Policy's goals, NIH seeks comment on clarifying that the GDS Policy applies to research funded in part or in whole by NIH that generates large-scale genomic data, even if NIH does not directly support the sequencing itself.	
NIH seeks input on:	
8. **Types of research covered by the GDS Policy.**	
1.	
1. Whether there are other types of research and/or data beyond the current scope of the GDS Policy that should be considered sensitive or warrant the type of protections afforded by the GDS Policy (e.g., with consent for future use and to be shared broadly, as well as IRB review of risks associated with submitting data to NIH), even when data are de-identified.	Infant blood spot tests required by state law to assess newborns for genetic conditions? In some states parents are not allowed to decline this testing, so if these tests are not excluded the result could be that sensitive data about family members is included in dataset(s) and used without consent by the family members affected.

2. Whether small scale studies (e.g., studies of fewer than 100 participants) and those involving other data types (e.g., microbiomic, proteomic) should be covered under the GDS Policy, and if training and development awards (e.g., F, K, and T awards) should be covered by the GDS Policy ("Implementation of the NIH Genomic Data Sharing Policy for NIH Grant Applications and Awards," NOT-OD-14-111).	Studies of all sizes should be covered under the GDS Policy. Participants in smaller studies, or studies that target rare conditions, are more vulnerable to re-identification and just as, if not more, deserving of the protections inherent in the GDS Policy.
3. Whether NIH-funded research that generates large-scale genomic data but where NIH's funding does not directly support the sequencing itself should be covered by the GDS Policy.	Any NIH-funded research should be covered under GDS policy. The protection of genomic data needs to be as comprehensive as possible, not limited by considerations regarding different agencies funding different portions of the process.
9. **Data sharing expectations under the GDS Policy.** Whether there are other types of research and/or data that warrant the data processing level and timeline expectations established by the GDS Policy (e.g., sharing lower levels of processed data, not just those of sufficient quality to validate and replicate findings as in the DMS Policy).	
How to Submit a Response	

Comments must be submitted at https://osp.od.nih.gov/rfi-updating-the-nih-genomic-data-sharing-policy. Responses will be accepted through February 28, 2022.	
Responses to this RFI are voluntary and may be submitted anonymously. You may also voluntarily include your name and contact information with your response. Other than your name and contact information, please do not include in the response any personally identifiable information or any information that you do not wish to make public. Proprietary, classified, confidential, or sensitive information should not be included in your response. After OSP has finished reviewing the responses, the unredacted responses may be posted to the OSP website.	
[i] Final NIH Policy for Data Management and Sharing (October 29, 2020). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html	
[ii] Notice of Changes to NIH Policy for Issuing Certificates of Confidentiality (September 7, 2017). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-109.html	
[iii] NIH Genomic Data Sharing Policy (August 27, 2014). https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html	

[iv] NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing (GDS) Policy (March 9, 2015). https://osp.od.nih.gov/wp-content/uploads/NIH_Best_Practices_for_Controlled-Access_Data_Subject_to_the_NIH_GDS_Policy.pdf	
[v] NIST Risk Management Framework. Federal Information Security Modernization Act (FISMA) Background. https://csrc.nist.gov/projects/risk-management/fisma-background	
[vi] FedRAMP Program Basics. https://www.fedramp.gov/program-basics/	
[vii]Supplemental Information to the National Institutes of Health Genomic Data Sharing Policy (August 27, 2014). https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf	

[viii] The DMS Policy defines scientific data as "The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens."	
[ix] Institutional Certifications. https://osp.od.nih.gov/scientific-sharing/institutional-certifications/	
Inquiries	
Please direct all inquiries to:	
NIH Office of Science Policy SciencePolicy@od.nih.gov	