World Privacy Forum Comments

To: National Institutes of Health

Re: The National Institutes of Health draft Genomic Data Sharing Policy that promotes sharing, for research purposes, of large-scale human and nonhuman genomic data generated from NIH-supported and NIH-conducted research.

Via email <GDS@mail.nih.gov> and web <http://gds.nih.gov/survey.aspx>

Genomic Data Sharing Policy Team
Office of Science Policy
National Institutes of Health
6705 Rockledge Drive
Suite 750
Bethesda MD 20892

November 14, 2013

Re: Draft NIH Genomic Data Sharing Policy, Request for Public Comments

The World Privacy Forum welcomes the opportunity to comment on National Institutes of Health's (NIH) draft Genomic Data Sharing (GDS) Policy that promotes the sharing for research purposes of large-scale human and nonhuman genomic data generated from NIH-supported and NIH-conducted research. The draft policy appeared at 78 Federal Register 57860 (Sept. 20, 2013) at http://www.gpo.gov/fdsys/pkg/FR-2013-09-20/pdf/2013-22941.pdf.

The World Privacy Forum is a non-profit public interest research and consumer education group. We focus our activities on privacy issues, and much of our work focuses on technology and health-related privacy issues. You can see our publications and more information at www.worldprivacyforum.org.

Sharing of genomic data for research, just like the sharing of other health data, can be acceptable from a privacy perspective under the right terms and conditions. The WPF recognizes and appreciates the sensitivity shown by NIH in addressing the rapidly changing landscape of genomic research. We see the proposed policy as another positive action intended to keep policy, practice, and technology properly aligned. The proposed policy applies to research involving

nonhuman genomic data as well as human data, but our comments are concerned exclusively with human data.

There is much that we like in the proposed policy, but we also have some suggestions for change and comments on some aspects of the policy and how it relates to other NIH policies.

I. Scope

We support application of the policy to all NIH-funded research, including NIH intramural research. We very much support enforcement of the policy through the withholding of research funding as well as other methods.

II. Certificates of Confidentiality

We are pleased that NIH encourages researchers and institutions to obtain certificates of confidentiality. **We believe that obtaining a certificate should be <u>mandatory</u> and that NIH should amend the policy to say so.**

Further, no data should be eligible for deposit in a repository unless the repository itself has a certificate of confidentiality covering the data. We believe a repository can qualify under 42 U.S.C. § 241(d) because those who maintain research data are "engaged in biomedical, behavioral, clinical, or other research" as provided in the statute. We support allowing a transition period so that any existing repositories will have an opportunity to meet the certificate requirement. It should not be difficult for any repository to obtain a certificate.

We recognize that there are many uncertainties about the scope and value of certificates of confidentiality, and we encourage NIH (or HHS) to undertake a thorough review of certificates with an eye toward recommending statutory or other improvements. We do not suggest, however, that a review is a prerequisite to proceeding with the GDS policy.

We note that some of the recommended repositories are located in other countries. A certificate of confidentiality provided by the Secretary of HHS is not likely to have any value in a jurisdiction outside the United States. Whether data in a foreign repository will actually have any privacy protection or possibly greater privacy protection than would be available in the United States is hard to say and variable from country to country. Issues about differing legal regimes and different degrees of privacy protection cannot be readily or immediately resolved. These issues are worthy of more attention by NIH in the near future. We do not suggest that attention to international issues is a prerequisite for proceeding with the GDS policy.

III. Informed Consent

We note NIH's view of the importance of informed consent.

> Respect for and protection of the interests of research participants is fundamental to the NIH's stewardship of human genomic data. The informed consent under which the data or sample were collected is the basis for the submitting institution

> to determine the appropriateness of data submission to NIH-designated data repositories, and whether the data should be available through open or controlled access.
>
> ****
>
> If human genomic data are to be shared in open-access repositories, the NIH expects that participants will have provided explicit consent for sharing their data through open-access mechanisms. For studies proposing to use cell lines or clinical specimens, the NIH expects that informed consent for future research use and broad data sharing will have been obtained even if the cell lines or clinical specimens are de-identified.

We want to call attention to a Privacy Act of 1974 System of Records that NIH maintains. The system is *Clinical, Basic and Population-based Research Studies of the National Institutes of Health* (HHS/NIH/OD, 09–25–0200). http://www.gpo.gov/fdsys/pkg/PAI-2003-HHS_AHCP/html/PAI-2003-HHS_AHCP-SYSTEMOFRECORDS-9-25-0200.htm. This appears to be the main system of records for NIH research activities. We assume that this system of records applies to NIH research activities using genomic data, although it does not apply to research by NIH grantees.

The individuals in this System of Records include anyone who might be the subject of any type of research study, as follows:

> Adults and/or children who are the subjects of clinical, basic, or population-based research studies of the NIH. Individuals with disease. Individuals who are representative of the general population or of special groups including, but not limited to: normal controls, normal volunteers, family members and relatives; providers of services (e.g., health care and social work); health care professionals and educators, and demographic sub-groups as applicable, such as age, sex, ethnicity, race, occupation, geographic location; and groups exposed to real and/or hypothesized risks (e.g., exposure to biohazardous microbial agents).

One of the routine uses that allows for **non-consensual** disclosure of data from any research study for any "**research purpose**" is long and complex:

> A record may be disclosed for a research purpose, when the Department: (A) Has determined that the use or disclosure does not violate legal or policy limitations under which the record was provided, collected, or obtained; e.g., disclosure of alcohol or drug abuse patient records will be made only in accordance with the restrictions of confidentiality statutes and regulations 42 U.S.C. 241, 42 U.S.C. 290dd–2, 42 CFR part 2, and where applicable, no disclosures will be made inconsistent with an authorization of confidentiality under 42 U.S.C. 241 and 42 CFR part 2a; (B) has determined that the research purpose (1) cannot be reasonably accomplished unless the record is provided in individually identifiable form, and (2) warrants the risk to the privacy of the individual that additional

> exposure of the record might bring; (C) has required the recipient to (1) establish reasonable administrative, technical, and physical safeguards to prevent unauthorized use or disclosure of the record, (2) remove or destroy the information that identifies the individual at the earliest time at which removal or destruction can be accomplished consistent with the purpose of the research project, unless the recipient has presented adequate justification of a research or health nature for retaining such information, and (3) make no further use or disclosure of the record except (a) in emergency circumstances affecting the health or safety of any individual, (b) for use in another research project, under these same conditions, and with written authorization of the Department, (c) for disclosure to a properly identified person for the purpose of an audit related to the research project, if information that would enable research subjects to be identified is removed or destroyed at the earliest opportunity consistent with the purpose of the audit, or (d) when required by law; and (D) has secured a written statement attesting to the recipient's understanding of, and willingness to abide by, these provisions.

Because this is not a comment on that system of records, we do not propose to unpack the terms of the routine use or evaluate the details. The routine use includes a variety of limitations and conditions on disclosure of research data.

What we want to point out is that the proposed GDS policy is quite clear on respecting informed consent of data subjects and on obtaining explicit informed consent for data sharing, **even when the stakes only involve de-identified data**. The existing NIH policy, as reflected in its research system of records, is considerably different and much less rigorous. A routine use by its definition is an exception to the statutory rule that an agency must obtain the consent of a data subject before disclosing the data subject's *identifiable* data.

We see a significant contradiction here. If de-identified genomic data can be shared with a repository only with explicit informed consent, why can fully identifiable research data be shared under a routine use without any notice to the data subject or consent from the data subject? If all genetic repositories were open access, that might justify the emphasis on explicit informed consent, but not all deposited data will necessarily be public.

We pose three questions about the policy differences suggested by the existing routine use and the proposed GDA policy.

1. **Is the NIH research System of Records policy inconsistent with the proposed GDS policy?**

Technically, the answer is no. The routine use quoted above requires that a "use or disclosure does not violate legal or policy limitations under which the record was provided, collected, or obtained." Thus, if NIH adopts the GDS policy, it should not be possible to rely on the routine use to justify a disclosure of data that the GDS policy restricts.

But NIH has not proposed an amendment to the routine use that refers to the GDS policy, although admittedly an amendment prior to adoption of the GDS policy may be premature. We believe that the routine use should expressly include a reference to and citation of all NIH policies that limit the scope of the routine use.

After NIH adopts the GDS policy, we ask that NIH amend the routine use in question to refer to the GDS policy and any other similar administrative policies that serve to limit the scope of the routine use. No one who relies on the routine use as a justification for a disclosure should have to look elsewhere to find the relevant NIH policy statements. However, we invite your attention to the remaining two questions before amending the routine use.

2. Should genomic data be treated differently than other health data?

This is not a simple question to address or answer. We acknowledge that there are some categories of health data (substance abuse data, for example) that are subject to different privacy standards than other health data. Some concerns about genomic data extend beyond the interests of the data subject alone, more so than for other health data. Yet we are not prepared to support a bright line around genomic data or to agree that different treatment of genomic data is appropriate in all circumstances.

We are doubtful about approaches based on genetic exceptionalism. We suggest that standards for the research uses and disclosures of all health data need to be considered together. Establishing strict standards for one category of health data will create difficult definitional issues and discontinuities. This already is the case with substance abuse information. NIH needs to look at the proposed GDS policy through a more open lens and to consider how the policy matches other policies governing research use of identifiable health data.

3. Should <u>all</u> health data be shared for research only with explicit informed consent?

That seems to be the position reflected in the proposed GDS policy for de-identified genomic data. If sharing de-identified requires explicit consent, how can any other policy be justified for other health data, much of which is shared in identifiable form for research activities? We are not prepared to suggest that the sharing of de-identified health data (or other data, for that matter) for bona fide research should occur only with explicit informed consent. We note that legal and policy considerations support some non-consensual data sharing for research. Given the routine use reproduced above, NIH seems to share that view as well. HHS clearly supports research disclosures of protected health information under the HIPAA health privacy rule. See 45 C.F.R.§ 164.512(i).

We observe that while there are admirable aspects of the proposed GDS policy, NIH may be painting itself into an informed consent corner that will prove to be uncomfortable in the future. We observe generally that there are many available methods that protect the confidentiality of data subjects, including encryption, certificates of confidentiality (despite their shortcomings), criminal and civil penalties, data use agreements or contracts that give a data subject the ability to pursue legal remedies against those who misuse data, denial of research funds to data misusers, and other measures. These other methods may in some circumstances provide better

options and strike a better societal balance between privacy and research interests than a reliance on explicit consent.

We want to offer another observation about explicit informed consent. Especially with respect to open repositories of data, there is a need for clearer standards for what constitutes explicit informed consent. We are disinclined to leave this entirely to institutional review boards.

Individuals often do not understand consent forms. Asking for consent to make genomic data openly available calls for more careful explanation of the risks and consequences. As recent history demonstrates clearly, today's de-identified genomic data is tomorrow's identifiable data. Individuals must be warned about all possibilities, even though this is a difficult thing to do in reality. It is hard to ensure that genuinely informed consent exists.

In this regard, we have considerable concerns about the long-term identifiability consequences of open repositories of genomic data. The trend toward greater identifiability of smaller and smaller portions of an individual genome will continue and may accelerate. We anticipate the eventual creation of a central databank with genomic information on everyone. Informed consent should not act as a privacy fig leaf of protection when other measures would be more effective, particularly in the long run. It could take but one single front page "horror story" for public attitudes to harden in ways that would not be favorable to research. The stakes are high.

IV. De-Identification and Long-Term Preservation

The draft policy states:

> Human data that are submitted to NIH-designated data repositories should be de-identified according to the standards set forth in the HHS Regulations for the Protection of Human Subjects and the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. The de-identified data should be assigned a random, unique code, and the key held by the submitting institution.

The requirement for a "random, unique code" is far from clear to the casual reader. Is the intent that each individual record will have its own code? Will each dataset have its own code? The policy needs to be more explicit about the requirements for the code as well as for maintenance of the key. As long as a code is subject to misuse, to a security breach, or to a court order that requires the institution to produce the code, the data is not truly de-identified.

Without more directions, institutions will lose, misplace, and improperly protect the keys. Institutions should be directed to plan to preserve the key and make it appropriately available for a minimum specified period. There may be a justification in some or perhaps even all cases for destruction of the code after a set time or event. Institutions may also need instructions about what to do with keys if they disband, merge, change function, or otherwise transition in form or function.

In the same regard, we wonder about the long-term existence of repositories themselves. We admit to a lack of detailed knowledge about the policies of repositories to know whether standards are needed for data preservation and destruction by repositories. It is a question worth looking at.

V. Data Withdrawal

The proposed policy states:

> Submitting investigators and their institutions may request removal of data on individual participants from NIH-designated data repositories in the event that a research participant withdraws his or her consent. However, data that have been distributed for approved research use cannot be retrieved.

This is fine as far as it goes. However, if data distributed for research use includes the code for each individual, then retrieval of the data will be possible. We fail to see a reason that withdrawal of consent is effective only at the repository level and not at all times if retrieval is possible. NIH should reconsider the proposed limit on data withdrawal. As long as an activity is justified on the basis of explicit informed consent, everyone has a responsibility to respect the withdrawal of that consent, even if inconvenient. As long as data withdrawal is practically possible, it should be required. Extraordinary measures are not needed, but with informed consent, close calls should belong to the data subject.

VI. Policy Changes

It is fair to suggest that the draft GDS policy will not be the last NIH policy statement in this area. NIH needs to confront the likelihood of future policies and the application of future policies to data already in a repository and, perhaps, to data already in the hands of researchers. We note, for example, that the HIPAA health privacy rule requires a covered entity to inform individuals that a future policy change may apply to protected health information maintained prior to the effective date of that policy change. 45 C.F.R. § 164.520(b)(1)(v)(C). We do not suggest that all future policy changes must necessarily be retroactive, but the possibility that changes might be retroactive should be considered. Of course, a legal change may affect data activities that occurred prior to the effective date of a law. That possibility might also warrant a mention.

Thank you again for the opportunity to submit these comments. We welcome the opportunity to discuss these issues with you further.

Respectfully submitted,

Pam Dixon
Executive Director,
World Privacy Forum
www.worldprivacyforum.org