November 20, 2013

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

Dear Sir and/or Madam:

This letter is in response to the Request for Public Comments on the Draft Genomic Data Sharing (GDS) Policy That Promotes Sharing, for Research Purposes, of Large-scale Human and Nonhuman Genomic Data Generated from NIH-supported and NIH-conducted research on behalf of the International Society for Biological and Environmental Repositories (ISBER).

The International Society for Biological and Environmental Repositories (ISBER) is an organization that addresses the technical, legal, ethical, and managerial issues relevant to repositories of biological and environmental specimens (see www.isber.org for additional information). Although not restricted to repositories of human specimens intended for research, the great majority of ISBER members focus on providing annotated human tissues for research, either procured for research purposes, or from residual clinical specimens obtained during the course of routine medical care. ISBER membership and expertise in the area of human tissues used for research is extensive, longstanding, and representative of the best practices in the field. ISBER's thought leaders in this area are worldwide. As such, we have a keen interest in the matter of data sharing policy development and the implications for research.

The risk of developing the vast majority of non-infectious human diseases such as cancer, heart disease and diabetes involves a complex interplay of environmental influences along with the underlying genetic background of an individual. Studies aiming to unravel these interactions at a population level so as to deliver public health measures to prevent disease or to identify novel drugs to modify risk or treat conditions require enormous datasets across hundreds or thousands of individuals. Biobanks of human biospecimens linked to epidemiological and clinical data are a core infrastructure for this important research.

Data Availability

ISBER agrees that data arising from research involving human specimens should be made available to the broadest number of researchers to enable discoveries to be made. Indeed, it is clear from the vast quantities of data now being generated by Next Generation Sequencing that discoveries will only be made through international collaborations of researchers using information from many thousands of participants in a 'crowdsourcing' manner.

Data Standards

The draft GDS policy will require the upload of additional genomic data set types for shared access from publicly funded research. The breadth of the increase in data that will be required to be shared back to the dbGAP database is likely thousands of times more than the data currently required, yet there are no data standard provisions outlined in the draft policy. There is also a marked increase in the number of study co-variates that must be uploaded with the accompanying genomic data. It is unclear that the storage infrastructure to support this expanded data requirement will be successful if there are no data standards for the upload of the genomic and co-variate data. At each research site, there will be a requirement to house this data ready for upload without any funding provided to support the data collation and upload process.

Protecting Participants

ISBER supports only controlled access to all human genomic data because of potential for abuse of genomic data and uses that some subpopulations may find unacceptable. Of note, single-nucleotide polymorphisms (SNPs) may permit grouping subpopulations across multiple public genomic datasets that would permit data based on grouping race, ethnicity, sex and age. This could permit purposeful or accidental stigmatization of some subpopulations.

In addition, the policy does not adequately address how concerns regarding indigenous populations would be handled. The sharing of genomic research data from study participants is of global interest both as a means of rapid progress in health and of misuse. The mechanisms in place, including policy and guidelines governing data access must be inclusive of global concerns. Without that, there is a risk of data sets being biased by the non-participation of subpopulations whose concerns have not been addressed. Indigenous peoples are one such population where the loss of cultural oversight with the secondary use of their genetic data is a concern. A number of indigenous peoples regard their tissue and data as being collectively owned. Consent may be obtained for a specific study(s) with the participant gaining approval from their ethnic group, but this cultural oversight is lost when data is submitted to open access repositories and the potential for misuse is increased. The question arises “can an individual give consent for collectively owned data sharing for unknown purposes?” This issue is at the core of indigenous participation in research.

ISBER is concerned that the NIH Data Access Committees (DAC) in some cases may not have enough expertise to recognize all potential issues and problems with requests for data. In addition, given the extent of broad data sharing, sufficient resourcing will be necessary. ISBER suggests that the DAC’s confer widely when necessary to address unique concerns about data access requests and that after DAC approval, requests to use data be referred to the IRBs responsible for the individual datasets to be utilized for concurrence with DAC approval.

Additionally, ISBER notes that while there are currently multiple laudable terms and conditions listed for secondary research using controlled access data, the statement that investigators are "expected" to abide by NIH User Code of Conduct is insufficient for protection of controlled access data. Because of the prior abuse of genetic data, there should be significant penalties against both investigators and their institutions if the terms and conditions of secondary research are violated. Such penalties should be specified in the Genomic Data Sharing Policies.

Informed Consent Requirement

There is concern by ISBER that after the effective date of the policy that even de-identified specimens will require informed consent for all genomic studies funded by NIH. While ISBER recognizes that informed consent is desirable, it may not be feasible to obtain consent and it is a major undertaking to implement consent for all specimens collected surgically. Biorepositories in some cases collect specimens based on IRB approved waiver of consent; in many cases institutions/biorepositories cannot afford financially to consent all patients from whom tissue specimens are obtained in the course of routine care. Some larger institutions, for example, may perform up to 40,000 operations per year at which tissues are removed. With advances in genomic techniques, archival paraffin blocks are increasingly being used for genomic studies. Thus, any surgically removed specimen might potentially be appropriate for genomic studies. The cost to consent even 20,000 patients would be one million dollars at the low cost of $50 per consent and approximately two million at the more likely cost of $100 per consent. Such costs would be prohibitive at most institutions and valuable research opportunities would be lost. Because repositories collect specimens for future unknown uses, a requirement for consent for genomic sequencing of all de-identified specimens may severely limit the specimens available for genomic research and some research could simply not be done. While the policy suggests that exceptions for the use of clinically collected specimens without consent would still be considered by the NIH, the IRBs are in the best position to make determinations about the risks and benefits of the research. IRB waivers of consent should still be permissible for the use of clinically collected specimens for genomic research so that important research is not impeded.

General Document Composition

Clarification is needed regarding which actions are mandatory for policy compliance, and which are intended to be optional. Inclusion of permissive language and qualifiers such as "should" and "expects" are more confusing than helpful.

Provision of definitions rather than examples is highly suggested. Examples function best to supplement clear guidelines and definitions but are by themselves less than instructive. What, for example, will constitute a "widely-used data repository" or a "large-output" sequencing instrument or genotyping platform under this policy?

Summary

Each of the concerns expressed regarding the current policy draft is certain to lead to difficulties within the context of international collaborations. ISBER therefore respectfully suggests that the draft policy be restructured to ensure that it is suitable to the target audience and that it does not present a potentially confusing array of requirements.

ISBER appreciates the opportunity to respond to the Request for Public Comments on the Draft Genomic Data Sharing (GDS) Policy and would like to extend an offer of assistance in the further development of this policy.

Respectfully submitted,

Fay Betsou, DrSc HDR
ISBER President 2013-2014