Compiled Public Comments on the Draft NIH Genomic Data Sharing Policy

September 20, 2013 – November 20, 2013

Public Comments

Commenter: Abhinav

Date of comment: 10/2/2013 22:38

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: whenever personal (even deidentified) data is shared, the investigators should notify the people whose data they are sharing so that the people can (1) know what the data is being used for, (2) be aware if there is a breach in data security, (3) be themselves be aware that the data is being used for said purposes only.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: If the investigators or their institutions or related institutions are charging money for the samples or the data, the person whose data or samples are being 'sold' should be compensated.

Commenter: ACTG and IMPAACT Network Laboratories, Bob Coombs

Date of comment: 11/21/2013 9:52

Comment:

Dear Genomic Data Sharing Policy Team:

Regarding the draft NIH Genomic Data Sharing (GDS) Policy released September 20, 2013:

This policy indicates that it applies to all NIH-funded research projects that involve nonhuman organisms or human specimens that produces genomic data. As laboratory representatives of the AIDS Clinical Trials Group (ACTG) and the International Maternal Pediatric Adolescent AIDS Clinical Trials Group (IMPAACT) we support more open access to sequencing data; however, we wish to raise several points on how this policy could have untoward consequences in terms of the GDS policy and our domestic and international clinical trials:

- This policy seems to include HIV and HCV sequence data generated as part of clinical trials (*Sequence data from tens of isolates from infectious organisms*).
    - As the majority of our clinical trials continue for two years or more, dissemination of viral sequence data from ongoing clinical trials could adversely affect the studies by potentially unblinding the study arms and revealing information on antiviral drug resistance for individuals whose HIV or HCV is sequenced in real-time as part of the clinical trial (*Data release timeline: Up to 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first).*

    - The release of these sequence data before the clinical trial is completed will encourage "database mining," which may compromise the primary endpoints for the clinical trials by encouraging the rapid analysis and publication of study data before the protocol team is able to publish the primary protocol results. As such, it would make more sense to submit these data for public access one year after the study closes and the primary data analysis is published.

    - The release of these sequence data before the clinical trial is completed will also compromise confidentiality agreements established between collaborating pharmaceutical industry partners and the clinical trials networks.

- This policy could create a substantive burden on a clinical trials group to complete sequencing analysis if required to submit each sequence from an applicable study within 3 months after sequence data are generated (*Data submission expectation: Project specific, generally within 3 months after data generation*). From a scientific perspective, premature access to these

sequences before the sequence can be adequately evaluated in the context of the total study data would be a disservice to all.

  - The proposed submission requirement also would be much earlier than the ClinicalTrials.gov data posting of results based on the timeline for the overall study, which is generally no later than 1 year after the study completion date, which in turn is the date of last visit for determining the primary endpoint of the study. Thus, for clinical trials, a sequence submission timeline concordant with the ClinicalTrials.gov timeline for results submission would seem most appropriate.

- The informed consent process for study participation would need to be modified to harmonize between the requirement of the GDS policy and the conduct of our ACTG and IMPAACT clinical trials. This would be a burden to do retrospectively for our ongoing clinical trials.

- The proposed data sharing requirements will also put an undue burden on our international clinical trial studies by international Institutional Review Boards who will look most unfavorably on this expanded data sharing requirement outside of the clinical trial. This will make it even more difficult to transport study specimens across international borders, which will severely limit our ability to analyze clinical trial specimens and analyze protocol data and may also negatively affect protocol approval by some international ethics committees.

- There will be specific effects of human genomic data sharing for ACTG protocols A5128 and A5243; and IMPAACT protocols P1026S, P1058A, P1070, P1078, P1083, P1097, P1103, P1106, P1108, P1110, and P1111 because genomic sequence data can be potentially linked to individual subjects, thus compromising subject confidentiality. As a consequence the proposed GDS policy will severely impair our ability to acquire further knowledge into the host genomics of HIV pathogenesis and antiviral drug pharmacokinetics among different subject populations.

Thank you in advance for considering the ACTG and IMPAACT Network Laboratories comments on the draft NIH Genomic Data Sharing Policy.

With kindest regards,

Robert W. Coombs, MD, PhD | ACTG Network Laboratory Principal Investigator | University of Washington | Seattle, WA

Susan Fiscus, PhD | IMPAACT Network Laboratory Principal Investigator | University of North Carolina | Chapel Hill, NC

Ronald Bosch, PhD | ACTG and IMPAACT Statistical Data Management Center | Harvard University School of Public Health | Boston, MA

Robert W. Coombs MD PhD FRCPC
Professor of Laboratory Medicine & Medicine
Vice-chair for Research, Department of Laboratory Medicine
University of Washington Box 359690
Chief of Service, Laboratory Medicine
Harborview Medical Center
325-9th Ave, Room 706C
Seattle, Washington 98104-2499
Phone: 206-897-5205 or 5202
Fax: 206-897-5203
UWMC Paging: 206-598-6190

CONFIDENTIALITY NOTICE: This e-mail message, including any attachments, is for the sole use of the intended recipient(s) and may contain confidential, proprietary, and/or privileged information protected by law. If you are not the intended recipient, you may not use, copy, or distribute this e-mail message or its attachments. If you believe you have received this e-mailmessage in error, please contact the sender by reply e-mail and destroy all copies of the original message. Thank you.

Commenter: Stephanie Alessi, Emily Borgelt, Dr. Mildred Cho, Prof. Hank Greely, Hayden Harvey, Dr. Sandra Lee, Emily Liu, Dr. David Magnus, Dr. Marsha Michie, Pr

Date of comment: 11/19/2013 16:22

Comment:

Section II. Scope and Applicability: 1) For data or samples already collected, if the consent is neither consistent nor inconsistent with the data sharing policy, an IRB, privacy board or equivalent body should review the proposed data sharing plan and should conduct an Institutional Certification (as described in Section 5 of the draft policy). That is, the review should assure that: • The protocol for the collection of genomic and phenotypic data was consistent with 45 CFR part 46; Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained; • Risks to individuals and their families associated with data submitted to NIH-designated data repositories were considered; • To the extent relevant and possible, risks to groups or populations associated with data submitted to NIH-designated data repositories were considered; and• The investigator's plan for deidentifying datasets is consistent with the standards outlined in this PolicyThis review should also evaluate whether the donors had other concerns that would be violated by the GDS policy. 2) We believe that NIH should give guidance about interpretation of 45 CFR 46.101(b)4 to reconcile the position that NIH no longer considers genomic data to be deidentified with the exception for publicly available data. This clarification should include guidance about whether controlled access data are considered publicly available.

Section IV.A. Data Sharing Plans: In Section 5 of the draft policy (quoted below), we recommend that the default should be to not allow these data to be shared, but that an IRB, privacy board or funder could make exceptions. The policy should require that the IRB, privacy board or equivalent reviews the justification in addition to the funder, considering whether the scientific justification is compelling, and also considering the reasons for absence of consent. In addition, the sentence "cell lines or clinical specimens that were created or collected" should be changed to "cell lines that were created or clinical specimens that were created". "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. If there are compelling scientific reasons that necessitate the use of cell lines or clinical specimens that were created or collected after the effective date of this Policy and that lack consent for research use and data sharing, investigators should provide a justification for the use of any such materials in the funding request."

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: In Section 4 (quoted below), NIH should provide sample language and talking points for what "explicit consent for sharing their data through open-access mechanisms" means. In addition, in this

paragraph, for projects initiated after the effective data, the word “prominently” should be added where indicated.“For studies initiated after the effective date of this Policy, the NIH expects the informed consent process and documents to state [add prominently] that a participant’s genomic and phenotypic data may be shared broadly for future research purposes and also explain whether the data will be shared 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.”

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Russ Altman

Date of comment: 9/27/2013 23:07

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: My only comment is that you should create model language for human subjects consent forms that is compatible with this entire policy, so that those of us doing the work can be sure that we have consent forms that are compatible with this policy.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: American College of Medical Genetics and Genomics (ACMG), Michael S. Watson

Date of comment: 11/20/2013 22:59

Comment:

Section II. Scope and Applicability: see below

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: November 19, 2013Genomic Data Sharing Policy Team Office of Science Policy National Institutes of Health 6705 Rockledge Dr., Suite 750 Bethesda, MD 20892 http://gds.nih.gov/survey.aspxTo Whom it May Concern:The American College of Medical Genetics and Genomics (ACMG) is pleased to be able to respond to your request for comments on the Draft NIH Genomic Data Sharing (GDS) Policy. ACMG represents over 1400 clinical and laboratory geneticists in North America who are board certified in Medical Genetics. We appreciate how important it will be to capture the molecular and associated clinical information from those undergoing genome sequencing in order to inform our understanding of the clinical implications of genome variation and to improve the care of those with conditions with genetic influences. As such, our comments focus primarily on the sharing of human genome sequence data with a focus on its associated clinical information. As projects to clinically annotate and curate genome variation are rapidly emerging, it is important that we define the policies that will both protect those whose clinical and genomic information are made available to clinical investigators as well as the investigators who are funded to develop and analyze such data. It is our view that data sharing in this area has such significant potential benefit that we must find appropriate means by which potential harms to patients are minimized and the interests of grant funded data developers are protected rather than to step away from the greater good. That said, we believe that the model of rapid public data sharing that worked so well for capturing sequence data is much more challenging when it comes to the clinical data that allows for the clinical annotation of the genome sequence. Ensuring patient privacy and ensuring that researchers who are funded to do specified types of research will require robust controlled data sharing policies. ACMG considers the GDS policy in general to be an important addition to NIH's grant funding requirements. However, it is equally important that the NIH grant and contract review processes be aligned with these policies in order to ensure that the importance and adequacy of data sharing practices proposed by investigators be carefully reviewed and given significant weight in their prioritization for funding. Clinical research has not always been a priority of study sections and

grant reviewers which has been to the detriment of the development of robust genotype:phenotype relationships.With regard to human genomic data, including the associated phenotype data, it will be important to design policies that work for different sources of data. Phenotype data comes in many forms ranging from patient provided data for which validity in some contexts has not been validated to data generated by providers in the course of routine care delivery to data that has been cleaned and subjected to peer review that further enhances the data quality. Data that is developed by those working under specific grants with obligations to publish their data over time requires greater protection for the data developers than does data submitted from clinical environments. Further, it is often difficult to publish informative research on rare diseases within narrow time frames since cases may accrue slowly. When access to such data is controlled or restricted to investigators who appreciate the limitations inherent in small data sets that may not reflect the full spectrum of a disease, the chances for the publicly available information to be misused by those less familiar with these constraints is avoided. Allowances should be made for such circumstances. We recognize that data could also come into NIH funded databases from individuals without grant funding but with an interest in being able to access accumulating data to improve their care of patients with uniquely rare variations in genotype or phenotype. Physicians value access to data that can inform their management of their patients with similar rare findings. Increasing numbers of academic medical centers are now offering higher levels of access to sharable data when used in the course of care delivery than is provided to researchers who don't have immediate clinical needs for the data. A system that acknowledge the importance of sharing clinical data to improve patient care in real time will lead to different policies for those involved in research than those delivering care. Clear policies as to how such data is to be accepted, cleaned, and governed can enhance its utility. Data Submission Expectations and Timelines: ACMG agrees that any data that supports the conclusions in publications of research results should be readily publicly available to facilitate replication and independent assessment. Sequence data should continue to be made available rapidly to inform genome anatomy and variation. However, overly aggressive policies by NIH that require clinical investigators to make available unpublished data related to ongoing grant-funded clinical research could discourage the open submission of data to NIH databases. Some of the features of studies that may require longer periods of protection from secondary use include clinical history development studies of rare diseases. This may be most important in funded studies to define clinical histories of rare diseases, particularly when accrued in the routine course of clinical care that may require additional protection to allow investigators to accumulate sufficient patients to provide an accurate view of a particular condition. When very long periods of time are anticipated that may keep data hidden for longer than is desirable, strong policies for how such data can be secondarily used could maximize its availability. Research may also involve sensitive phenotypes that may require little individual level data sharing.In order to maximize the likelihood and value of open data sharing, a means for assigning global unique identifiers to case studies brought into databases can allow for credit to accrue to the data providers as the data they provide leads to publications. It is also quite possible that genomic data will eventually be considered to be subject to HIPAA privacy due to a possibility that it can be re-identified after de-identification. It should also be recognized that much ongoing clinical research is already being done under consent agreements with subjects that may not align with new policies for GDS. As re-consenting could jeopardize the ongoing research, allowances for such situations should be in place and remediated during competitive renewals and not applied to noncompetitive renewals of funding.Data Repositories: ACMG agrees that registration of studies

in dbGAP can have value similar to that realized by registration clinical trials. It enhances the possibility of developing collaborations and identifying gaps in the research pipeline. However, we do not conflate registration of studies with requirements for submission of individual level data. We will address our view on the sharing of individual level data in the following section.Tiered Systems for the Distribution of Human Data: As alluded to in our earlier comments, we believe that there are some types of data that need higher levels of protection. Data from patients with rare diseases when shown at the individual level increases the likelihood that identify of individuals can be reconstructed. Hence, care will be needed in determining the level at which rare disease and rare gene variant data is made public, even if de-identified. We appreciate how important this data is and are pleased that there are many who are well informed about the risks and benefits of making their own clinical data publicly available and agree to do so. The proposed tiers of open vs. controlled access to data allows for a range of options that fit with study constraints. The proposed allowance for data use limitations to accommodate patient preferences provides a reasonable balance. Informed Consent: We agree that these policies should be applied prospectively. Re-consenting of patients in ongoing time limited studies could greatly complicate the completion of their work. Clinical and other deidentified data that can increase the likelihood of reidentification of subjects should follow the agreements reached at the time of informed consent procedures. While we favor as much clinical data as possible being publicly available, we agree that data that is put into open-access databases should be explicitly consented. We have concerns about the proposal that clinical specimens to be used in studies be consented, even if de-identified. The NICHD has funded the creation of a newborn dried blood spot repository that now includes as many as 25 million specimens, many of which were acquired 10-15 years ago. Specimens have been obtained under State Public Health authorities. The vast majority represent a general population while a small minority represent specimens from babies with rare diseases identified by newborn screening. The practicality of requiring states to seek written consent for the deidentified use of these specimens in research would not be a viable proposal. It would require every birthing hospital in the US to offer quality informed consent to patients. If consent is to be required, an opt-out model is the only form of consent that is likely to work in our currently disjointed health care system. We agree that any studies that involve a newborn blood spot that is linked to an individual should be done with patient consent at the time the study is occurring. We are also concerned about the definition of research that may be applied to dried blood spot work. It is important to distinguish secondary uses of these spots for program development and quality assurance and improvement from “research” so that continued improvement of this important public health program is not jeopardized. We appreciate the opportunity to comment on this proposed policy. The final policy will have a great impact on what is likely the next most important genome study after the sequencing of the human genome, that being its clinical annotation.Sincerely, Michael S. Watson, MS, PhD, FACMG Executive Director

AMERICAN COLLEGE OF RHEUMATOLOGY
EDUCATION • TREATMENT • RESEARCH

Specialists in Arthritis Care & Research
2200 Lake Boulevard NE • Atlanta, GA 30319-5312
Phone: (404) 633-3777 • Fax: (404) 633-1870
www.rheumatology.org • info@rheumatology.org

November 20, 2013

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

Re: Input on the Draft NIH Genomic Data Sharing Policy (NOT-OD-14-018)

Dear Ladies and Gentlemen:

The American College of Rheumatology, representing over 9,000 rheumatologists and health professionals, appreciates the opportunity to comment on the draft NIH Genomic Data Sharing Policy to inform the design of a policy to describe the responsibilities of investigators and institutions for the submission of non-human and human genomic data to data repositories, the secondary research use of such data, and expectations regarding intellectual property. These challenges and opportunities will require significant investment in synergistic, collaborative funding mechanisms, as well as infrastructure and technological innovations.

Over the years, the NIH has made great strides in facilitating the sharing of valuable resources, including human and non-human data and biologic samples. The ACR's own Foundation has adapted these guidelines and is currently considering recommendations to increase its efforts in these areas.

Members of the ACR and its Committee on Research respectfully provide the following comments for your consideration as you continue to revise the policy related to sharing genomic data:

- Regarding instructional certification, the draft states, "The responsible Institutional Signing Official of the submitting institution should provide an Institutional Certification to the funding IC prior to award." It would be extremely helpful if the NIH could develop a template for all institutions to use. This is especially true for investigators at academic centers, especially those with multiple collaborators and/or funders at institutions across the country. The ACR is highly in favor of this approach, and would be happy to recommend investigators and administrators who may be helpful to you.

- Regarding consent, we recommend that investigators and sponsoring organizations be strongly encouraged and/or required to use broad, common language in relevant study consent forms that will ensure data will be allowed to be shared as intended. Suggested language from the NIH, consistent with new interpretations of the common rule, would be most helpful. With respect to protected populations, we advise that careful consideration is given to children who reach the age of majority who have provided assent and their parents have provided consent.

- With respect to data integrity: Depending on the type of data that is deposited, will it be possible to avoid duplication of data, since the data are de-identified? How will this be addressed to maximize the value of the repositories while protecting the privacy of individuals?

- Careful consideration should be given to the timelines proposed, which may need to be lengthened. Otherwise, how does one protect the intellectual property of the study investigators that comes from the analysis of the data? If study investigators cannot maintain control of their data or have a say in how their data are used, requiring that they allow broad access to these enormous datasets within 9 months (or less) of generating them may not be reasonable. Is there a way to make this a collaborative process? That is, can the investigators that generated the controlled-access data have a say in how it is used and who uses it and participate as collaborators in certain projects if they wish?

We present these proposals to the NIH to improve an outstanding body of work, and to assure that the future of rheumatology research is bright and that our patients benefit from ongoing research into the rheumatic diseases.

The American College of Rheumatology greatly appreciates the work you do and the opportunity to provide these comments. Thank you for your consideration and please feel free to contact us for further discussion. If you have questions, please contact Mary Wheatley, senior director of research and training, at (404) 633-3777 or mwheatley@rheumatology.org.

Sincerely,

Bruce N. Cronstein, MD
Chair, Committee on Research

233 N. Michigan Ave.,21st Fl.
Chicago, IL 60601

phone » (312) 233-1100
fax » (312) 233-1090
web » www.ahima.org

American Health Information Management Association®

November 20, 2013

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

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

Dear Genomic Data Sharing Policy Team:

On behalf of the American Health Information Management Association (AHIMA), I am pleased to provide comments on the "Draft NIH Genomic Data Sharing (GDS) Policy" as drafted and published by the Office of Science Policy of the National Institutes of Health (NIH).

AHIMA is the national non-profit association of health information management (HIM) professionals with component state associations in all 50 states, the District of Columbia and Puerto Rico. There are more than 71,000 members nationally who are dedicated to effective and efficient health information management. HIM professionals work for more than 40 different settings including hospitals, physician offices, long term care organizations, clinics, health information technology vendors and developers, consulting firms, life science companies, and government and education systems. AHIMA's members can be found in numerous and diverse roles with a wide range of responsibilities. Individual members are hospital administrators; deans of universities; lawyers; privacy and compliance officers; data stewards; government officials; coders and data analysts; and consultants and industry professionals.

Our members typically manage electronic health record (EHR) systems and oversee increasingly complex and vital health information management principles and processes in various care delivery settings. AHIMA and its members help assure quality, cost effective, and efficient health and healthcare through data and information governance and stewardship. As the data and information custodians of healthcare organizations' health records (whether paper or electronic) and leaders in the effective management of health information, ensuring the privacy, security, and confidentiality of personal health information has been a fundamental principle for the health information management (HIM) profession throughout its 85-year history. Today, HIM professionals continue to face the challenge of maintaining the privacy and security of patient information, an effort that grows in complexity as information becomes more and more distributed in electronic systems. The challenge of this responsibility has also increased due to the constantly changing legislative and regulatory environment. Ongoing efforts to share and exchange health data and information continue to present policy, operational, technological and ethical issues.

In providing input, we will address selected topics identified in the request for comment, as well as provide some general comments regarding the ongoing need to ensure the appropriate use of health data and information for research, based on AHIMA's previous and ongoing work in this area.[1] [2]

NIH Draft GDS Policy

AHIMA acknowledges that the draft GDS Policy supports the NIH's mission to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability.[3]

Data Submission Expectations and Timeline

AHIMA agrees that human data that are submitted to the NIH-designated data repositories should be de-identified according to the standards set forth in the HHS Regulations for the Protections of Human Subjects and the HIPAA Privacy Rule. We are generally supportive of the proposal that the de-identified data should be assigned a random, unique code and the key to that code be held by the submitting institution. However, AHIMA further recommends that the GDS specifically delineate what happens to the human data that is received that is not de-identified as required. Processes to address such instances should include quarantining the identifiable data, returning it to the institution, and prohibiting its use unless or until it is de-identified as required.

AHIMA agrees that a Certificate of Confidentiality as defined by NIH[4] could serve as an additional safeguard to prevent compelled disclosure of any personally identifiable information that it holds in NIH-designated data repositories.

Data Repositories

Although AHIMA agrees that investigators who elect to submit data to a non-NIH-designated data repository should confirm that the appropriate data security, confidentiality, and privacy measures are in place, we recommend that NIH define the process for this confirmation. In addition, AHIMA believes that the GDS should specifically articulate and include the appropriate requirements for information governance, data security, confidentiality, and privacy (http://www.ahima.org/topics/infogovernance).

Informed Consent

AHIMA supports the specificity required for the informed consent including that it states that "a participant's genomic and phenotypic data may be shared broadly for future research purposes and also explain whether the data will be shared through open or controlled access." AHIMA recommends that the GDS policy should state that the NIH "…***requires*** that participants will have provided explicit consent…" rather than "…***expects*** that participants will have provided explicit consent…" Additionally, rather than an "expectation" for future research use, it should be a "requirement" that "…the 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." In this regard, informed consent is a tool to provide the subject with knowledge of how their information will be used and how their privacy will be protected. Therefore, the privacy protection provided by the informed consent and the NIH should be a requirement, not an expectation.

Further, AHIMA believes that informed consent should include:

1. Notification that even despite due diligence and strong privacy protections that take steps to de-identify information, genetic information can sometimes be re-identified; and,
2. That individuals donating DNA or participating in research should be aware that privacy implications may exist that extend beyond themselves to family members or those who share their genome, including parents, siblings, children, adoptees, birth parents, sperm donors, and others.

When one donates DNA, he/she is also donating information about others. The donors need and should be made aware of this in the informed consent.

AHIMA agrees that data use limitations should be specified in the Institutional Certification submitted to the NIH prior to the research award. It should further specify the requirements regarding any permissible/allowable data use.

Institutional Certification

AHIMA recommends adding the following bullet point:

- The data submission and ultimate use is in accordance with the ethical use of health information.

According to AHIMA's Code of Ethics,[5] the ethical obligations of the health information management (HIM) professional include the safeguarding of privacy and security of health information; disclosure of health information; development, use, and maintenance of health information systems and health information; and ensuring the accessibility and integrity of health information.

AHIMA believes that healthcare consumers are increasingly concerned about security and the potential loss of privacy and the inability to control how their personal health information is used and disclosed. Core issues include what data should be collected, how the data should be handled, who should have access to the data, under what conditions the data should be disclosed, how the data are retained and if/when data are no longer needed, and how data are disposed of in a confidential manner. Harmonization with other federal as well as applicable state regulations is also critical.

Ethical obligations are central to the professional's responsibility, regardless of the employment site or the method of collection, storage, and security of health information. In addition, sensitive information (e.g., genetic, adoption, drug, alcohol, sexual, health, and behavioral information) requires special attention to prevent misuse. In the increasing complex world health services and biomedical research, interactions with consumers, requires expertise in the protection of the information.

Exceptions to Data Submission Expectations

As in the Informed Consent section, we recommend elevating "expected" to "required" in this section. Additionally, as in the Institutional Certification section, we recommend adding the following bullet:

- Submitting and using data in accordance with their ethical use.

Acknowledgment Responsibilities

As in earlier sections, we recommend elevating "expected" to "required."

Conclusion

AHIMA appreciates and supports the use of health information for research and scientific value and agrees that date use must be conducted done in accordance with all applicable laws, regulations, institutional polices and *ethical requirements.*

Thank you for providing the opportunity to comment and we look forward to working with you on the further development of the Draft GDS Data Sharing Policy. If we can provide any further information, or if there are any questions regarding our feedback, please feel free to contact me or Meryl Bloomrosen, Vice President, Thought Leadership, Practice Excellence, and Public Policy at Meryl.bloomrosen@ahima.org. Please let us know if we can be of further assistance to you in your efforts.

Sincerely,

Lynne Thomas Gordon, MBA, RHIA, CAE, FACHE, FAHIMA
Chief Executive Officer

---

[1] http://perspectives.ahima.org/critical-issues-in-bioinformatics-and-computing/#.Uo0qapWA3IU
[2] http://perspectives.ahima.org/flexible-approaches-for-teaching-computational-genomics-in-a-health-information-management-program/#.Uo0q5JWA3IU
[3] Draft NIH Genomic Data Sharing Policy Request for Public Comment. September 20, 2013.
[4] http://grants.nih.gov/grants/policy/coc/
[5] AHIMA Code of Ethics. Revised and Adopted 2011.

AMERICAN SOCIETY FOR HISTOCOMPATIBILITY & IMMUNOGENETICS

15000 Commerce Parkway • Suite C • Mt. Laurel, NJ 08054
Phone: 856.638.0428 • Fax: 856.439.0525 • Web site: www.ashi-hla.org • E-mail: info@ashi-hla.org

November 20, 2013

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

To the Genomic Data Sharing Policy Team,

We are clinical scientists and researchers in the fields of transplantation, histocompatibility, immunogenetics and immunogenomics, responding to the Request for Public Comments on the Draft NIH Genomic Data Sharing (GDS) Policy issued on September 20th 2013. Collectively, we have extensive clinical and research experience with genes in the Major Histocompatibility Complex (MHC) and Leucocyte Receptor Complex (LRC) regions of the human genome. These regions (on chromosomes 6 and 19, respectively) are central to the study of disease etiology, diagnosis and therapy, but are poorly represented in current genome assemblies due to high levels of polymorphism and structural variation. Our comments, while specific to these immunogenomic regions, apply to all regions of the genome that display high levels of polymorphism and structural variation.

We would like to address the human genomic data submission expectations outlined in section IV.C.1 of the GDS Policy. Appendix A of the GDS Policy states that submission of Level 1 processed data (aka, initial sequence reads) is not expected for human data if those reads are included in a Level 2 aligned sequence file (e.g. BAM format), and states that Level 2 processed data (i.e., DNA sequence aligned to a reference sequence or *de novo* assembly) are expected for submission. However, Level 1 processed data that have not been aligned to a reference sequence or that have been excluded from *de novo* assembly do not appear to be expected for submission. As we detail below, the continual discovery of new HLA and KIR sequence polymorphisms can rapidly invalidate the interpretation of aligned and *de novo* assembled sequences for these genes, reducing their utility for future studies. Furthermore, the current state of the primary and alternative alignments for the MHC and LRC regions of the genome is insufficient to permit reliable and accurate alignment of initial sequence reads for these regions. The sharing of unmapped or unaligned initial sequence reads is critical for the investigation of these regions.

The MHC region on human chromosome 6p21.3 is the most medically relevant region of the human genome. More than 100 infectious, autoimmune and pharmacological disease phenotypes and cancers are associated with genetic variation in the MHC[1-9], and in particular with the Human Leucocyte Antigen (HLA) genes. HLA molecules are critical components of the adaptive immune system, mediating the specific destruction of infected cells and production of antibodies. In addition, HLA molecules interact functionally with Killer-cell Immunoglobulin-like Receptor (KIR) molecules, key components of the innate immune system that also play critical roles in transplantation and disease[10-20]. The HLA and KIR (chromosome 19q13.4) regions are the most polymorphic in the human genome[2, 21, 22, 23, 24]; both are polygenic and highly dense with homologous genes[2, 21, 25, 26], and both display extensive structural variation[27, 28]. Due to extensive genetic variation observed for these genes among

PRESIDENT Marilyn S. Pollack, PhD, D(ABHI) 210.567.5698
PRESIDENT-ELECT John A. Gerlach, PhD, D(ABHI) 517.432.3467
VICE PRESIDENT OF OPERATIONS Judith E. Baker, CHS, I(ASCP) 214.820.8863
SECRETARY Dorothy M. Levis, MT(ASCP), CHS(ABHI) 734.922.1056
TREASURER Elizabeth R. Trimble, MT(ASCP), CHS(ABHI) 316.268.5447
PAST PRESIDENT Amy B. Hahn, PhD, D(ABHI) 518.262.5574
EXECUTIVE DIRECTOR Katherine Miranda 856.793.0907

human populations, study of the HLA and KIR genes is also a model for health disparities research[29].

The investigation of HLA and KIR polymorphism is an active and ongoing pursuit; as of October 2013, 678 unique KIR gene sequences and 9,945 unique HLA gene sequences have been identified[23, 24]. These sequences are housed in the IPD-KIR[24] and IMGT/HLA[23] Databases, which are updated on a regular basis. Both were most recently updated in October of 2013, with the addition of 79 new KIR sequences, and 439 new and extended HLA sequences. The IMGT/HLA and IPD-KIR Databases are the primary resources for the alignment of HLA and KIR initial sequence reads and for the validation of *de novo* sequence assemblies of these reads, and are the only resources available to relate gene sequences to the HLA or KIR allele nomenclature[30, 31], complex naming systems that are key for investigations of these genes[32, 33].

The constant increase in sequence polymorphism knowledge at these loci means that any *de novo* and aligned sequence assemblies for these genes will rapidly become obsolete, as the initial sequence reads need to be realigned and *de novo* assemblies revalidated in the context of each database update. The GDS Policy's expectation of the sharing of aligned or assembled Level 2 processed data alone for these genes means that those data can never be reevaluated in the context of future IMGT/HLA or IPD-KIR database updates. The cultural, medical and scientific ramifications for this information loss are unacceptable; this loss alone should be sufficient reason to reexamine the GDS Policy.

Because of sequencing and assembly challenges posed by the high level of polymorphism and structural variation within the MHC and LRC regions, these regions are poorly represented in Genome Reference Consortium (GRC) assembly GRCh37.p13; seven alternate locus assemblies are available for the MHC region, and eight alternative haplotypes are available as novel assemblies for the LRC region. All of these reference and alternative assemblies describe haplotypes common only in European populations[28, 34, 35, 36, 37, 38, 39] and do little to represent the extensive divergence and polymorphism observed in the USA and worldwide[40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50]. Moreover, these alternative alignments are largely incomplete and insufficiently reflect even basic levels of established structural variation and polymorphism for the HLA and KIR genes. We anticipate that a large number of complete reference assembly sequences will be needed to enable reliable genomic investigations and personalized medical applications for the MHC and LRC regions.

For example, of the five major structural variants (the DR1, DR51, DR52, DR53 and DR8 haplotypes) recognized for the HLA-DRB genes[27], only sections of the DR52 and DR53 haplotypes are represented in the alternative MHC alignments, while a section of DR51 is represented in the primary assembly. The DR1 and DR8 haplotypes, which constitute 38% of known HLA-DRB polymorphisms[23], are not represented in GRCh37.p13.

The primary assembly for the LRC includes a complete KIR haplotype representing a single genomic structure common in European populations [28]. Although many KIR haplotypes of differing gene content are known, of the eight novel assemblies for the LRC, only one (RefSeq NW_003571055.1) includes a KIR haplotype of alternative genomic structure. These LRC haplotype structures are medically important; their variation is associated with reproductive disorders, as well as decreased relapse after

bone marrow transplant, the bone marrow graft versus leukemia reaction and bone marrow graft versus host disease[13, 14, 16, 20].

In addition, large gaps in the alignments for these regions omit characteristic genes. Four of the seven alternative locus assemblies for the MHC omit the HLA-DRB1 gene and three omit the HLA-B gene, both of which are present in all individuals. None of these alternative locus assemblies contain the ~70kb segment that includes HLA-Y and associated genes. As noted above, many KIR genes are not represented in GRCh37.p13. However, these missing genes are represented in UCSC's hg19 assembly, and are present in one of four chromosome 19 unlocalized genomic contigs included in GRCh37.p13. This discrepancy between the h37 and hg19 assemblies illustrates the limitations of current alignment methods for this important genomic region.

The high levels of similarity among HLA genes and pseudo genes and among KIR genes and pseudogenes pose challenges for their *de novo* assembly and complicate the use of the alignments in assembly GRCh37.p13. The sequences of many HLA pseudogenes and gene fragments can be erroneously mapped to HLA genes[51], and it is likely that this has occurred in the primary and alternative locus assemblies. This degree of error makes it impossible for meaningful information to be obtained from Level 2 processed data.

Overall, the sharing of aligned or assembled Level 2 processed data alone will severely limit the research community's capacity for novel investigations and meta-analyses of immunogenomic data. For example, the potential introgression of HLA polymorphisms from archaic human species in the modern human population[52] could not have been detected through the analysis of Level 2 processed data; only the availability of all initial sequence reads for the Neanderthal and Denisovan genomes made this work possible.

Given the ongoing detection of new HLA and KIR polymorphisms and the current state of the genome assembly, it is our opinion that acceptance of aligned or assembled Level 2 processed data for the MHC and LRC regions is insufficient to meet the GDS Policy's stated goal of ensuring the responsible sharing of research data. We recommend that the GDS Policy require that any shared Level 2 processed data permit the complete regeneration of Level 1 processed data, making initial sequence reads that have not been aligned to a reference sequence or that have been excluded from *de novo* assembly available for future studies. This will allow shared genomic data to be reevaluated in the context of future improvements in the genomic assembly and alignment methodologies, and future expansions of relevant reference polymorphism databases. Ultimately, this approach to data sharing will represent an investment in the future of genomic investigation, stimulating novel research efforts and fostering improved clinical outcomes.

Sincerely Yours,

The ASHI Scientific Affairs Committee

This letter of comment on the Draft NIH Genomic Data Sharing Policy has been authored by the undersigned on behalf of the Scientific Affairs Committee of the American Society of Histocompatibility and Immunogenetics, the Immunogenomics Data Analysis Working Group, and the Immunogenomic Next Generation Sequencing Data Consortium.

Attila Berces, PhD
Chairman, Omixon

Nezih Cereb, MD
CEO, Histogenetics

Professor Frans H.J. Claas, PhD
Director, Eurotransplant Reference Laboratory
Professor Transplantation Immunogenetics, Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, The Netherlands

Lisa Creary, PhD
Post Doctoral Research Scientist, Histocompatibility and Immunogenetics Department, NHS Blood and Transplant, United Kingdom

Sue Davey, MSc, DipRCPath
HLA NGS project leader, Histocompatibility and Immunogenetics Department, NHS Blood and Transplant, United Kingdom

Professor Peter Donnelly, PhD, FRS
Director, Wellcome Trust Centre for Human Genetics
Professor of Statistical Science, Department of Statistics, University of Oxford, United States

Nadim El-Awar, PhD, MBA
Research Scientist
Transplant Diagnostics, Thermo Fisher Scientific

Henry A Erlich, PhD
Senior Scientist, Center for Genetics, Children's Hospital & Research Center Oakland, United States

Professor Marcelo A Fernández-Viña, PhD
Chairman, 17th International HLA and Immunogenetics Workshop
Organizer, Immunogenomic Next Generation Sequencing Data Consortium
Professor, Department of Pathology, Stanford School of Medicine
Director, Histocompatibility, Immunogenetics, and Disease Profiling Laboratory, Stanford Medical School Blood Center, United States

Professor Gottfried Fischer, MD
President, European Federation for Immunogenetics
Associate Professor, Department for Blood Group Serology and Transfusion Medicine, Medical University of Vienna, Austria

Professor Howard M. Gebel, PhD
Vice chair, ASHI Scientific and Clinical Affairs Committee
Professor of Pathology, Emory University Hospital, United States

Daniel Geraghty, PhD

Clinical Research Division, Fred Hutchinson Cancer Research Center, United States

Benjamin Gifford
Research and Development, Life Technologies

Professor Pierre-Antoine Gourraud, PhD MPH
Dept. of Neurology, School of Medicine, University of California San Francisco, United States

Professor Wolfgang Helmberg, MD
Professor of Medicine, Department of Serology and Transfusion Medicine, Medical University Graz, Austria

Jill A Hollenbach, PhD MPH
Chair, ASHI Scientific and Clinical Affairs Committee
Co-Chair, Immunogenomic Data Analysis Working Group
Organizer, Immunogenomic Next Generation Sequencing Data Consortium
Staff Scientist, Center for Genetics, Children's Hospital & Research Center Oakland, United States

Susan H Hsu, PhD
Director, Histocompatibility / Molecular Genetics, American Red Cross

Uma Kanga, PhD
Senior Research Scientist, Department of Transplant Immunology and Immunogenetics, All India, Institute of Medical Sciences, India

Steven J Mack, PhD
Editor In Chief, Human Immunology
Co-Chair, Immunogenomic Data Analysis Working Group
Organizer, Immunogenomic Next Generation Sequencing Data Consortium
Staff Scientist, Center for Genetics, Children's Hospital & Research Center Oakland, United States

Martin Maiers
Organizer, Immunogenomic Next Generation Sequencing Data Consortium
Director of Bioinformatics, National Marrow Donor Program, United States

Professor Steven GE Marsh, BSc PhD ARCS
President, International HLA and Immunogenetics Workshop Council
Deputy Director of Research, Anthony Nolan Research Institute
Professor of Immunogenetics, Cancer Institute, University College London, United Kingdom

Professor Diogo Meyer, PhD
Professor, Department of Genetics and Biology, Institute of Biosciences, University of São Paulo, Brazil

Professor Derek Middleton, DSc, PhD, FRCPath
President, International HLA and Immunogenetics Workshop Council
Consultant Clinical Scientist, Transplant Immunology, Royal Liverpool University Hospital Trust
Professor, Institute of Infection and Global Health, University of Liverpool, United Kingdom

Professor Dimitri S Monos, PhD

Director, Immunogenetics Laboratory
Professor of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, United States

Professor Cristina Navarrete, PhD, FRCPath
Head of Histocompatibility and Immunogenetics Research, NHS Blood & Transplant, United Kingdom

Paul J Norman PhD
Senior Research Scientist, Department of Structural Biology, Stanford University School of Medicine, United States

José M Nunez
Laboratory of Anthropology, Genetics and Peopling History, Department of Genetics and Evolution, Anthropology Unit, University of Geneva, Switzerland

John Ord
Research Associate, Medical School, University of Cambridge, United Kingdom

Professor Daniel Ramon, PhD, D[ABHI], HCLD[ABB]
Clinical Assistant Professor
Director, Histocompatibility Laboratory, Department of Pathology, University of Michigan Medical School, United States

Bryan Ray, PhD
Director Research and Development | LIFECODES Products

James Robinson
Senior Bioinformatics Scientist, Anthony Nolan Research Institute, United Kingdom

Professor Alicia Sanchez-Mazas
Director, Laboratory of Anthropology, Genetics and Peopling History, Department of Genetics and Evolution, Anthropology Unit, University of Geneva, Switzerland

Professor David Senitzer, PhD, ABMLI, ABHI
Clinical Professor of Hematology and Hematopoietic Stem Cell Transplantation
Director, Histocompatibility Laboratory, City of Hope National Cancer Center, United States

Professor Richard M. Single, PhD
Assistant Professor, Department of Mathematics and Statistics, University of Vermont, United States

Professor Emeritus Glenys Thomson, PhD
Department of Integrative Biology, University of California Berkeley, United States

Professor Elizabeth A Trachtenberg, MS, PhD, D(ABHI)
Director, HLA, Immunogenetics and Disease Profiling Laboratory, Stanford Blood Center
Professor, Department of Pathology, Stanford University School of Medicine, United States

Christien Voorter, PhD
Transplantation Immunology, Tissue Typing Laboratory, University Hospital Maastricht, The Netherlands

Literature Cited

1. Johnson, A.D. and C.J. O'Donnell, *An open access database of genome-wide association results.* BMC Med Genet, 2009. **10**: p. 6.
2. Shiina, T., et al., *The HLA genomic loci map: expression, interaction, diversity and disease.* J Hum Genet, 2009. **54**(1): p. 15-39.
3. Watson, C.T., et al., *Estimating the proportion of variation in susceptibility to multiple sclerosis captured by common SNPs.* Sci Rep, 2012. **2**: p. 770.
4. Isobe, N., et al., *Genetic risk variants in African Americans with multiple sclerosis.* Neurology, 2013. **81**(3): p. 219-27.
5. Hildesheim, A., et al., Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst. 2002. 94(23): p. 1780-9..
6. Chapman, S.J. and A.V. Hill, *Human genetic susceptibility to infectious disease.* Nat Rev Genet, 2012. **13**(3): p. 175-88.
7. Leish, G.E.N.C., et al., *Common variants in the HLA-DRB1-HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis.* Nat Genet, 2013. **45**(2): p. 208-13.
8. Kirino, Y., et al., *Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1.* Nat Genet, 2013. **45**(2): p. 202-7.
9. Nomura, T. and T. Matano, *Association of MHC-I genotypes with disease progression in HIV/SIV infections.* Front Microbiol, 2012. **3**: p. 234.
10. Santin, I., et al., *Association of KIR2DL5B gene with celiac disease supports the susceptibility locus on 19q13.4.* Genes Immun, 2007. **8**(2): p. 171-6.
11. Brown, M.A., *Genetics and the pathogenesis of ankylosing spondylitis.* Curr Opin Rheumatol, 2009. **21**(4): p. 318-23.
12. King, A.L., et al., *A genome-wide family-based linkage study of coeliac disease.* Ann Hum Genet, 2000. **64**(Pt 6): p. 479-90.
13. Martin, M.P. and M. Carrington, *KIR Locus Polymorphisms: Genotyping and Disease Association Analysis*, in *Methods in Molecular Biology: Innate Immunity*, J. Ewbank and E. Vivier, Editors. 2007, Humana Press Inc.: Totowa, NJ. p. 49-64.
14. Hollenbach, J.A., et al., *Susceptibility to Crohn's disease is mediated by KIR2DL2/KIR2DL3 heterozygosity and the HLA-C ligand.* Immunogenetics, 2009. **61**(10): p. 663-71.
15. Ahlenstiel, G., et al., *Distinct KIR/HLA compound genotypes affect the kinetics of human antiviral natural killer cell responses.* J Clin Invest, 2008. **118**(3): p. 1017-26.
16. Miller, J.S., et al., *Missing KIR ligands are associated with less relapse and increased graft-versus-host disease (GVHD) following unrelated donor allogeneic HCT.* Blood, 2007. **109**(11): p. 5058-61.
17. Kulkarni, S., M.P. Martin, and M. Carrington, *The Yin and Yang of HLA and KIR in Human Disease.* Seminars in Immunology, 2008. **In Press**.
18. Martin, M.P., et al., *Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS.* Nat Genet, 2002. **31**(4): p. 429-34.
19. Khakoo, S.I. and M. Carrington, *KIR and disease: A model system or system of models?* Immunol Rev, 2006. **214**: p. 186-201.
20. Carrington, M. and M.P. Martin, *The impact of variation at the KIR gene cluster on human disease.* Curr Top Microbiol Immunol, 2006. **298**: p. 225-57.
21. Mungall, A.J., et al., *The DNA sequence and analysis of human chromosome 6.* Nature, 2003. **425**(6960): p. 805-11.

22. Trowsdale, J. and P. Parham, *Mini-review: defense strategies and immunity-related genes.* Eur J Immunol, 2004. **34**(1): p. 7-17.
23. Robinson, J., et al., *The IMGT/HLA Database.* Nucleic Acids Research, 2013. **41**: p. D1222-7.
24. Robinson, J., et al., *IPD - the Immuno Polymorphism Database.* Nucleic Acids Research, 2013. **41**: p. D1234-40.
25. *Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium.* Nature, 1999. **401**(6756): p. 921-3.
26. Vandiedonck, C. and J.C. Knight, *The human Major Histocompatibility Complex as a paradigm in genomics research.* Brief Funct Genomic Proteomic, 2009. **8**(5): p. 379-94.
27. G. Andersson, *Evolution of the human HLA-DR region.* Front Biosci, 1998. **3**: p. d739-45.
28. Hollenbach, J.A., et al., *Killer cell immunoglobulin-like receptor (KIR) gene content variation in the HGDP-CEPH populations.* Immunogenetics, 2012. **64**(10): p. 719-37
29. Noble, J. A., et al., *HLA class II genotyping of African American type 1 diabetic patients reveals associations unique to African haplotypes.* Diabetes, 2013. **62**(9): p. 3292-9.
30. Marsh, S.G.E., et al., *Nomenclature for factors of the HLA system, 2010.* Tissue Antigens, 2010. **75**(4): p. 291-455.
31. Marsh, S.G.E., et al., *Killer-cell Immunoglobulin-like Receptors (KIR) Nomenclature Report, 2002.* Human Immunology, 2003. **64**: p. 648-654.
32. Hollenbach, J. A., et al., *A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement.* Tissue Antigens, 2011. **78**(5): p. 333-44.
33. Mack, S. J., and J. A. Hollenbach. *Allele Name Translation Tool and Update NomenCLature: software tools for the automated translation of HLA allele names between successive nomenclatures.* Tissue Antigens, 2010. **75**(5): p. 457-61.
34. Gragert, L., et al., Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. Hum Immunol, 2013. **74**(10): p. 1313-20.
35. Maiers, M., L. Gragert and W. Klitz. *High-resolution HLA alleles and haplotypes in the United States population.* Hum Immunol, 2007. **68**(9): p. 779-88.
36. Vierra-Green, C., et al., *Allele-level haplotype frequencies and pairwise linkage disequilibrium for 14 KIR loci in 506 European-American individuals.* PLoS One, 2012. **7**(11):e47491.
37. Mack, S. J., et al., *HLA-A, -B, -C, and -DRB1 allele and haplotype frequencies distinguish Eastern European Americans from the general European American population.* Tissue Antigens, 2009. **73**(1): p. 17-32.
38. Mack, S. J., et al., *Human leukocyte antigen-A, -B, -C, -DRB1 allele and haplotype frequencies in Americans originating from southern Europe: contrasting patterns of population differentiation between Italian and Spanish Americans.* Hum Immunol, 2011. **72**(2): p. 144-9.
39. Hollenbach, J. A., et al., *16(th) IHIW: population global distribution of killer immunoglobulin-like receptor (KIR) and ligands.* Int J Immunogenet, 2013. **40**(1): p. 39-45.
40. Norman, P. J., et al., *Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans.* Nat Genet, 2007. **39**(9): p. 1092-9.
41. Norman, P. J., et al., *Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes*. Genome Res, 2009. 19(5): p. 757-69.

42. Norman, P. J., et al., *Co-evolution of Human Leukocyte Antigen (HLA) Class I Ligands with Killer-Cell Immunoglobulin-Like Receptors (KIR) in a Genetically Diverse Population of Sub-Saharan Africans*. Plos Genet, 2013. 9(10): e1003938.
43. Tu, B., et al., *HLA-A, -B, -C, -DRB1 allele and haplotype frequencies in an African American population.* Tissue Antigens, 2007. **69**(1): p. 73-85.
44. Ellis, J. M., et al., *Diversity is demonstrated in class I HLA-A and HLA-B alleles in Cameroon, Africa: description of HLA-A*03012, *2612, *3006 and HLA-B*1403, *4016, *4703.* Tissue Antigens, 2000. **56**(4): p. 291-302.
45. Solberg, O. D., et al., *Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies*. Hum Immunol. 2008, **69**(7): p. 443-64.
46. Nunes, J.M., et al., *Analysis of the HLA population data (AHPD) submitted to the 15th International Histocompatibility/Immunogenetics Workshop by using the Gene[rate] computer tools accommodating ambiguous data (AHPD project report).* Tissue Antigens. 2010, **76**(1): p. 18-30.
47. Buhler, S. and A. Sanchez-Mazas. *HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events.* PLoS One. 2011, **6**(2): p. e14643.
48. Riccio, M.E., et al., *16(th) IHIW: analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report)*. Int J Immunogenet, 2013, **40**(1): p. 21-30.
49. Gonzalez-Galarza, F. F., et al., *16(th) IHIW: extending the number of resources and bioinformatics analysis for the investigation of HLA rare alleles.* Int J Immunogenet, 2013, **40**(1): p. 60-5.
50. Mack, S. J., et al., *Common and well-documented HLA alleles: 2012 update to the CWD catalogue.* Tissue Antigens, 2013. **81**(4): p. 194-203.
51. Major, E., et al., *HLA Typing from 1000 Genomes Whole Genome and Whole Exome Illumina Data*. PLoS ONE, 2013. **8**(11): p. e78410.
52. Abi-Rached, L., et al., *The shaping of modern human immune systems by multiregional admixture with archaic humans.* Science, 2011. **7**;334(6052): p. 89-94.

American Society for Investigative Pathology
Investigating the Pathogenesis of Disease

MARK E. SOBEL, MD, PHD
EXECUTIVE OFFICER
9650 Rockville Pike, Bethesda, MD 20814-3993 (USA)
Tel: 301-634-7130 • Fax: 301-634-7990 • Email: mesobel@asip.org • www.asip.org

November 15, 2013

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

Dear NIH Genomic Data Sharing Policy Team:

The American Society for Investigative Pathology (ASIP) is pleased to comment on the Request for Information: Input on the Draft NIH Genomic Data Sharing Policy (NOT–OD–13–119). ASIP is a nonprofit educational 501(c)(3) organization primarily representing the academic pathology research community. We are a society of biomedical scientists who investigate disease, linking the presentation of disease in the whole organism to its fundamental cellular and molecular mechanisms. Our members use a variety of structural, functional, and genetic techniques, seeking to ultimately apply research findings to the diagnosis and treatment of patients. ASIP advocates for the practice of investigative pathology and fosters the professional career development and education of its members.

As a professional pathology society, we support efforts to increase our fundamental knowledge about the nature and behavior of living organisms and to apply this knowledge in a beneficial manner. We believe that data sharing can effectively utilize information gathered from a single research protocol to address additional questions. Concurrent with our interest in advancing scientific knowledge, we remain committed to ongoing human subjects research protections.

While ASIP is supportive of genomic data sharing, we are deeply concerned about many aspects of the recent Draft Genomic Data Sharing Policy. We appreciate the opportunity to present our thoughts on this important issue. As a member of the Federation of American Societies for Experimental Biology (FASEB), we are fully in support of the comment letter that FASEB has submitted. We encourage the Policy Team to reassess the proposed policy in light of both FASEB's concerns and the issues highlighted below.

Additional clarification is needed regarding what types of research are covered by the policy. The policy currently states that it applies to "… research involving non-human genomic data as well as human data that are generated through array-based and high-throughput genomic technologies."[1] Such technologies

[1] Request for Information: Input on the Draft NIH Genomic Data Sharing Policy, p. 2.

can be used for a variety of applications such as confirmatory testing that generate little new information but are essential to the scientific process. ASIP recommends a basic definition of what is included under the policy that is less tightly linked to the technology used to gather the information and specifically speaks to genomic testing for research purposes.

Appendix A of the draft provides four examples of data that would be subject to this policy. Providing examples without a more substantive framework will lead to significant confusion. FASEB has proposed that NIH issue a workflow diagram or charts to assist investigators and institutions in understanding what is subject to this protocol. Providing such information, along with more detailed scenarios, would increase the likelihood of a uniform understanding at the institution level.

Including projects where more than one gene or gene size region is sequenced and more than 100 participants are involved, would encompass research studies examining limited regions to determine correlations between regions (e.g., for variant X, does variation Y on another region correlate with clinical expression?). It is unclear why NIH has chosen the threshold of one gene or gene sized region when the other examples for humans represent far more substantial sequencing. ASIP encourages NIH to consider a higher threshold for information reporting. A clear definition of what is encompassed under this policy should assist in clarifying this concern.

The Draft Genomic Data Sharing Policy represents a significant administrative burden and may, as a result, impact overall grant funding. ASIP remains concerned about unintended administrative burdens throughout the research process. Given the current funding environment, a substantial expansion of research funds is not anticipated. While it is appropriate to allow proposals to include the additional funds needed to meet the requirements of the Genomic Data Sharing Policy, the level of funds required for each grant will increase. Given a constant, if not decreasing, funding level, NIH may no longer be able to fund the same number of research grants given the additional funds needed to meet the requirements of the data sharing policy. While ASIP supports the need for facilitating scientific discovery through information sharing, we remain deeply concerned that additional regulatory requirements will result in a further contraction of research funding.

Additional information is needed to understand the role of the IRB, the researcher and the data repository in determining which information will be housed in the open access area versus the controlled access site. ASIP is unclear on the criteria that will be used to designate open access versus controlled access of human data. The Data Sharing Policy should clearly spell out what issues, even issues beyond consent, that may be used in IRB deliberations to determine open versus controlled access. In addition, further information is needed to understand how NIH will handle any disputes related to institutional decisions. If there is the potential for overruling an organizational decision, an appeals process should be established and include clear criteria for accessing the process and criteria to be used in determining the outcome of the appeal.

Failing to allow for donor requests on research restrictions may have a substantial chilling effect on the ability to conduct quality scientific research. ASIP supports a researcher's ability to honor the requests of donors, providing assurance that the researcher and his or her institution are trustworthy entities. Specific individuals and/or subpopulations may be concerned that their specimens could be used for broader, yet to be named, research. Yet, study of that individual or specific subpopulation may yield highly relevant information for future care and treatment. Requiring that donors agree to broad data sharing may dissuade donors from research participation because they do not wish to have their samples used in unknown research. While the Draft Data Sharing Policy includes the submission of an Institutional Certification that could include restrictions on specimen use, it is unclear whether and how NIH will honor

specific exclusions. A policy limiting a researcher's ability to respect donor requests may have a substantial impact on the ability of researchers to recruit individuals to participate in both clinical trials and for research in highly sensitive areas.

A more reasonable timeframe for depositing data would encourage broader data submission, as well as maintaining a level playing field between large and small laboratories. The draft Policy indicates that data will be made available without restrictions on publication no later than six months after the initial date of submission or at the time of acceptance for publication, *whichever occurs first.*[2] The problem arises when this requirement is coupled with the expectation that data will be submitted within three months of data generation.[3] Researchers must be allowed adequate time to confirm data before deposition, ensuring confidence in the information contained within the repository. Repetition and correlation are fundamental to good science; these actions require time and attention and cannot also be performed within three months of original data generation.

Furthermore, smaller laboratories may reasonably expect a longer analysis period between data generation, confirmation, analysis and ultimately the submission for publication. The timeframes outlined in the draft Policy would significantly advantage larger laboratories with additional resources. This may ultimately be detrimental to the advancement of knowledge. Also, it is important to note that some journals may take longer than six months to review a paper. The need for appropriate and rigorous peer review should not be discouraged through the establishment of an unrealistic turnaround timeframe.

ASIP proposes that the policy be amended to indicate that the data be made available without restrictions on publication no later than six months after the initial date of submission or at the time of acceptance of the publication, *whichever comes later.* In addition, flexibility should be provided for investigators to submit data up until the time of submission of a manuscript for publication. This will improve the quality of the data submissions and ensure that they have been properly annotated and analyzed for quality control prior to submission to a repository. It would also be appropriate to include guidelines addressing situations in which either: data are not submitted for publication and are not appropriate for a repository; or data are not submitted for publication and are appropriate to be included in a data repository.

Clarification is needed on the Intellectual Property section to ensure that the language encourages and recognizes the unique contributions made by individual researchers. The draft Policy currently indicates that data should be made freely available through the repository including "all conclusions derived directly from them."[4] This, by its very nature, is what each researcher contributes to the scientific process. For example, how phenotype and genotype correlate and the resulting implications are the crux of much research. Scientific contributions should generally be released to the scientific community upon publication. We are unaware of a precedent that would suggest that genetic information should be treated differently than other information gathered during the course of scientific research. The scientific process would be undercut by mandating the use of a data repository containing sufficient information for another researcher to replicate a research study and failing to allow the original researcher adequate time for publication and/or protection of their intellectual property.

The additional statement in that paragraph is also problematic as it indicates that the data submitted to the repository should include "certain related information (e.g.,…)." While NIH has provided

[2] Ibid, p. 5.
[3] Ibid, p. 10.
[4] Ibid, p. 9.

examples of what might be included under "certain related information," it has not provided guidelines for how a researcher might determine what does or does not fall under this umbrella. It is currently a vague standard and requires additional clarification.

Harmonization is needed between the current Office of Human Subject Protection standards on what may qualify as an exemption from consideration as human subjects research and the consent requirements outlined in the draft Genomic Data Sharing Policy. Under current Office of Human Subject Protection (OHRP) guidelines, pathology specimens may qualify for exemption from consideration as human subjects research if they: (a) have not been obtained specifically for the current research project through an interaction or intervention with a living person; and (b) preclude the investigator from ascertaining the identity of the donor in any manner. ASIP believes that requiring patient consent as specified in the Draft NIH Genomic Data Sharing Policy is in conflict with current OHRP recommendations. Specifically, the prospect of data sharing and potential for use in secondary research may shift specimens gathered under the human subjects research exemption to non-exempt status. This is a substantial administrative burden as it would both dramatically increase the cost of current pathology research endeavors and would conflict with existing regulations.

Furthermore, in order to achieve the required inability to identify the donor as required by OHRP guidelines, researchers must work with already de-identified information. It is unclear how a "key" can be held by the submitting institution as described in the policy[5] as many researchers currently utilize completely de-identified information and agree to not attempt to ascertain the identity of the donor. No "key" is possible in this situation.

The genomics community as a whole is currently grappling with significant technological advances such that traditional de-identification methods may not be sufficient in the near future.[6] The ability to identify an individual from genomic specimens is evolving as technology advances. At one end of the spectrum is research involving a limited amount of genetic information such that it would be virtually impossible to identify an individual. At the opposite end of the continuum is research specifying a significant portion of the genome such that an individual could readily be identifiable. Re-identification is possible primarily by matching an identified reference sample to the sample being analyzed. However, research has demonstrated that it is possible to decode an anonymous or anonymized sample and combine it with other metadata (genealogical information, age, sex) to identify a surname. This latter method is currently complicated and cumbersome. While technology has not yet advanced to easily re-identify a previously anonymous sample, it is likely that in the foreseeable future an individual's identity may be determined with more ease and without reference samples.

The ethical, legal and social implications are significant. Several organizations have begun to address these issues and we would encourage the Data Sharing Policy Team to avail themselves of these resources including considerations raised in the October 2012 report from the Presidential Commission for the Study of Bioethical Issues, entitled "Privacy and Progress in Whole Genome Sequencing." One such recommendation of the Presidential Commission called for expanding the framework based around patient consent to include substantive standards and/or legal requirements around security, data access, confidentiality, and penalties for misuse. ASIP encourages additional exploration of these issues either

[5] Ibid, p. 5.

[6] Resources: Bohannon J "Genealogy databases enable naming of anonymous DNA donors," Science, (339), 18 Jan 2013, p. 262; Gymrek M, McGuire, AL, Golan D, Halperin E, Erlich Y "Identifying personal genomes by surname inference," Science (339) 18 Jan 2013, p. 321; Rodriguez LL, Brooks LD, Greenberg JH, Green ED "The complexities of genomic identifiability," Science (339) 18 Jan 2013, p. 275.

concurrently with the development of the Data Sharing Policy or in advance of the implementation of the policy. In this way, we believe that responsible data stewardship will be expressed and steps implemented to foster public trust in genomic research.

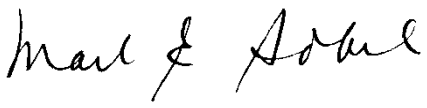
Extending HIPAA protections to data not previously subject to HIPAA creates further disharmony between The Common Rule and HIPAA, leading to confusion and increased administrative burdens. The Draft Policy indicates that data submitted should be coded to protect participant privacy and the standard that is chosen for de-identification is the standard set forth in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.[7] Research pathologists work with specimens that may be either obtained through a healthcare setting (in which both the HIPAA Privacy Rule and the Common Rule apply) or outside of the healthcare setting (in which only the Common Rule applies). Extending HIPAA protections to a population not currently under this regulatory framework is troubling. It is yet another example of the disconnect between these two regulations and the need for future harmonization.

It is unclear how archived pathology specimens should be treated if they are gathered at different points in time. Archived pathology specimens represent a unique challenge. For any given research study, specimens may have been gathered at various points of time, including those gathered after policy implementation, before policy implementation or from a biorepository with an unknown date of acquisition. It is unclear whether all data would be required to be put in a data repository or whether submission of partial data would be appropriate (for example, data gathered from specimens obtained after policy implementation but not information from older archived pathology specimens). Furthermore, patient consents on archived pathology specimens may differ as the standards have changed over time. If partial submission is mandated, an additional burden is placed on the researcher to obtain the acquisition date and then sort specimens by that date. Furthermore, the integrity of the data overall may not be maintained with partial data submission.

While ASIP supports data sharing efforts and its ability to advance science quickly, many aspects of this Draft Policy trouble us. We encourage NIH to carefully consider the points raised above, as well as the comments in the FASEB letter. We believe that without thoughtful deliberation, there will be significant confusion within the scientific community, increased administrative burdens and a decline in a number of individuals willing to participate in genomic research. We appreciate the opportunity to provide comments and are available to provide additional information should it be beneficial. Please contact Mark Sobel, M.D., Ph.D. at (301) 634-7130 or mesobel@asip.org.

Thank you for your consideration.

Sincerely,



Mark E. Sobel, M.D., Ph.D.
Executive Officer

[7] Request for Information: Input on the Draft NIH Genomic Data Sharing Policy, p. 12.

Commenter: Andrey P. Anokhin

Date of comment: 11/4/2013 12:55

Comment:

Section II. Scope and Applicability: I support the proposed policy on sharing genomic data

Section IV.A. Data Sharing Plans: I support the proposed policy

Section IV.B. Non-human and Model Organism Genomic Data: I support the proposed policy

Section IV.C. Human Genomic Data: I support the proposed policy

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: I support the proposed policy

Section VI. Intellectual Property: I support the proposed policy

Any other aspect of the draft GDS Policy: I support the proposed policy

Commenter: Bradley Aouizerat

Date of comment: 11/5/2013 13:04

Comment:

Section II. Scope and Applicability: Generally acceptable. However, there are populations at particular risk that are not adequately considered in the policy. For example, research participants from disadvantaged backgrounds or at risk for discrimination (e.g., HIV-positive, intravenous drug users) who have the right to participate in NIH-funded studies, but who may never be appropriate for participation under the Data Sharing Policy are not considered. This is particularly relevant given that the de-identification of samples does not adequately protect against undesired re-identification of participants.

Section IV.A. Data Sharing Plans: Criteria for a non-NIH-designated data repository in terms of acceptable data security, confidentiality, and privacy measures must be specifically delineated.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: The timeline for depositing data in NIH data repositories and release of data to approved investigators is not sufficiently long. Twelve months is generally appropriate.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Name of Individual(s) Submitting Comments:
Mary M. Langman

Email Address:
langman@mlahq.org

Affiliation:
Director, Information Issues and Policy, Medical Library Association

Commenting on behalf of:
Association of Academic Health Sciences Libraries (AAHSL) and Medical Library Association (MLA)

Academic Institution
Industry/Commercial Organization
Research Organization
Professional Society
Disease and/or Patient Advocacy Organization
General Public
Other

Name of Organization:
Association of Academic Health Sciences Libraries
Medical Library Association

City:
Seattle
Chicago

State:
WA
IL

As part of the process of developing the Genomic Data Sharing (GDS) Policy, NIH invites public comments on any aspect of the draft GDS Policy.

Comment 1:
Section II. Scope and Applicability

MLA and AAHSL members continue to support NIH policies that require timely sharing of NIH-sponsored research data, including regular review and updates to related policies. Like the NIH Public Access Policy, the Genomic Data Sharing Policy will enable researchers to realize new discoveries more quickly, accelerate the exchange of information among the research community, support scientific innovation, and preserve genomic data for years to come.

Comment 2:
Section IV.A. Data Sharing Plans

MLA and AAHSL members continue to support NIH in its promotion of data sharing plans for all NIH-funded genomic research. Health sciences librarians have the skills and expertise to provide researchers with assistance in the development of data sharing plans required in funding applications and proposals. In addition, librarians can help identify resources needed to support a proposed data sharing plan for inclusion in a project's budget.

Comment 3:
Section IV.B. Non-human and Model Organism Genomic Data

MLA and AAHSL fully support timely data release as described in the draft NIH Genomic Data Sharing Policy. To support successful implementation of this goal, and in the interest of streamlining the process for investigators, we recommend simplifying the timeline that is described in Appendix A. In particular, minimizing multiple data submissions and eliminating the need for additional consultations with NIH program staff whenever possible would enhance efficiency of the data sharing plan (Table "Expectations of Data Submission and Data Release" **Appendix A**). MLA and AAHSL believe that it would be helpful to include a link or reference to non-NIH-sponsored repositories within the Policy.

Comment 4:
Section IV.C. Human Genomic Data

MLA and AAHSL support submission of data to NIH-sponsored and non-NIH-sponsored data repositories with confirmed data security measures in place, as well as systems for managing both open and controlled access to human data, and support the use of Institutional Certification as part of the research data sharing planning process.

Comment 5:
Section V. Responsibilities of Investigators Accessing and Using Genomic Data

MLA and AAHSL members strongly support NIH–sponsored repositories and the pre-screened data security they offer, but recommend further clarification of the draft policy's options for submission to non-NIH-sponsored repositories.

We also recommend submission to repositories that allow open access to data where designated by the submitter as appropriate, as repositories permitting only controlled or closed access may unnecessarily restrict or delay access to data. Open access repositories allow for greater availability to its materials than closed repositories and align more closely with the objectives of this policy.

MLA and AAHSL support the underlying rules of informed consent as outlined in the draft NIH Data Sharing Policy, particularly where participant consent is addressed via a standardized process through an informed consent document or institutional IRB, Privacy Board or equivalent.

Comment 6:
Section VI. Intellectual Property

MLA and AAHSL strongly support the NIH position on Intellectual Property as it relates to research data, which furthers health research, innovation and development of new knowledge. MLA and AAHSL maintain that basic sequence data and related information are pre-competitive and such data made available through NIH-designated data repositories and all conclusions derived directly from them should remain freely available without any licensing requirements. Health sciences librarians actively support the principle that the copyright law was established to balance the rights of owners of with the rights of users. Accordingly, MLA and AAHSL support the NIH in encouraging broad use of NIH-funded genomic data that is consistent with a responsible approach to management of intellectual property derived from downstream discoveries as outlined in the NIH *Best Practices for the Licensing of Genomic Inventions and Research Tools Policy.*

Comment 7:
Any other aspect of the GDS Policy

We urge the NIH to extend the roll-out period for the NIH Genomic Data Sharing Plan beyond the currently proposed timeframe. We recommend allowing more time, considering that additional provisions for review of data submission plans by Institutional Review Boards and for Institutional Certifications are scheduled to be updated following final approval of the Policy (IV.C.4.). These organizational and structural changes will require additional months to plan and implement, and should be incorporated into a unified effective date.

Commenter: Association of American Medical Colleges (AAMC), Ann C. Bonham, Ph.D.

Date of comment: 11/20/2013 17:19

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: RE: Draft NIH Genomic Data Sharing Policy, 78 FR 57860-5, Sept. 20, 2013To the NIH Genomic Data Sharing Policy Team:The Association of American Medical Colleges (AAMC) is grateful for the opportunity to provide comments on the draft policy to share genomic data generated by NIH-funded research. The AAMC is a not-for-profit association representing all 141 accredited U.S. medical schools, nearly 400 major teaching hospitals and health systems, including 51 Department of Veterans Affairs medical centers, and 90 academic and scientific societies. Through these institutions, the AAMC represents 128,000 faculty members, 75,000 medical students, 110,000 resident physicians, and thousands of graduate students and post-doctoral trainees in the biomedical sciences. The AAMC fully supports the goals and intent of the proposed policy to promote sharing of genomic sequence data and related information among researchers. The goals are in line with and build upon prior policies, which the AAMC has also supported, including the policy for sharing of data obtained in NIH supported Genome-Wide Association studies and the best practices for sharing of genomic inventions, consistent with the sharing of research resources. (1,2) The AAMC has at every opportunity approved of objectives to increase sharing of research resources, data and information, and findings from federally sponsored research. Examples of such programs include the sharing of mouse models and the Public Access policy of the National Center for Biotechnology Information.(3,4) We also have supported the principles set forth in the Administration's recent directive for sharing of data, released by the White House Office of Science and Technology Policy in January, cited by the NIH draft policy.(5) In the AAMC's view, the imperative for sharing such data will, if anything, grow stronger with time. Increasingly, restricting access to data will not be perceived as advantaging certain investigators, but rather will be seen only as an effective means to marginalize their research. The AAMC commends NIH for developing a more comprehensive and effective policy for sharing genomic information widely, including non-human sequence data, and for sharing of complementary information on phenotypic expression of genetic information. The policy correctly emphasizes concerns for privacy and other dimensions in protection of human subjects, or participants, in research. The draft policy also updates and clarifies the disposition of intellectual property

protection on genetic sequences following from the 2013 U.S. Supreme Court decision in Association of Molecular Pathology v. Myriad. Recommendation 1: A final policy should state clearly that sharing genomic data resources should increase the urgency of including sufficient representation of all affected populations. If a study does not include such representation, the limitations of that study should be documented.Our primary concern is that the NIH and the research community's efforts to promote access to shared genomic data resources must also be accompanied by a redoubled commitment to ensure better representation of all affected populations in the studies for which genomic and phenotypic data are collected and from which these data are eventually shared. It has been documented for example that not all early GWAS studies effectively engaged all relevant populations within their scope of research, and as a result, conclusions based on such studies can be incomplete or misleading in application to patients.(6,7) Developing and expanding access to data drawn from limited populations or from narrow socioeconomic circumstances will risk exacerbating, not eliminating health inequities. To mitigate such risk requires researchers and institutions to practice improved outreach and community engagement, and for NIH and other agencies to promote better transparency in the research process and rules by which data are used and shared (which are elements that the draft policy currently addresses). Recommendation 2: A final policy should restate NIH's commitment to community engagement in the research process and complementary efforts to promote health equity in the application of genomic research.Above all, many prospective participants seek assurance that results and findings from studies will be applicable and actually applied to their communities. We understand that this is not an issue that can be entirely addressed by the current proposed policy, but urge the NIH and the greater research community to raise and take steps to address this concern at every opportunity. We also appreciate that the proposed genomic data sharing policy's aim to help promote transparency can strengthen the ability to ensure the data collected in these studies will be most widely applied. Recommendation 3: The final policy should recognize opportunities for innovative practices, such as dynamic consent, in promoting participant centered research collaborations.In the brief history of genomics, a persistent problem for government has been establishing policies that are sufficiently flexible to keep pace with and adapt to accelerating technological innovation. Traditionally, these technologies focus on the laboratory or clinical side of research, such as high-throughput sequencers or advanced computational facilities (as Dr. Collins has repeatedly noted, the efficiency of DNA sequencing technologies has improved at a rate faster than Moore's law for computer processors). However, there is another sense of "technology" to consider in the proposed data policy: patient advocacy and private voluntary organizations are developing systems that improve the process by which patients elect to participate in research studies, and provide informed consent to share information and samples. One innovation is through "dynamic consent" by which patients register an initial interest or disposition to participate in research projects, and can provide consent on an ongoing basis. These models have potential to improve rates of participation in research, while enhancing the autonomy of patients to make informed decisions with respect to the use of information or materials that they donate. We are aware that patient advocates are commenting on the potential to incorporate dynamic consent within the final proposal (currently, the draft policy anticipates only one-time, "static" consent in disposing identifiable data for subsequent research uses). Patients and other interested populations are a vital and increasingly active part of the research community, and are exploring new ways to participate in and enable research. This is one of the more exciting and notable developments in the evolution of genomic research.Recommendation 4: The final policy should

consider allowing the submission of data sharing plans on a “just-in-time” basis after award decisions are made, and the NIH should work with the research community in promoting best practices for sharing of genomic and other data among supported research projects.We understand that the new policy must necessarily increase to some extent the administrative burden on investigators and institutions, at a time that available resources are extraordinarily constrained. Colleagues at the Federation of American Societies for Experimental Biology (FASEB) have appropriately asked for more specificity in the requirements for investigators.(8) FASEB has also recommended that, rather than incorporate a detailed genomic data sharing plan in all applications, that NIH move to a “just in time” data sharing plan, similar to the just-in-time budgeting used for other NIH-funded research projects. The AAMC agrees that the just-in-time proposal has much merit and should be considered as a means to more economically use investigator time (too much of which may already be lost to writing applications). We understand that the NIH may see an advantage in reviewing genome data sharing plans within each research proposal prior to making an award; a better and more economical approach may be to help the community identify best practices in sharing genomic data among all awarded research projects.Recommendation 5: The policy should not be seen as enforcing standards stricter than that already required by regulation, and the NIH should provide more specification and clarification on the types of genomic data encompassed by the proposed policy. The AAMC notes that the policy represents a significant expansion over and above existing law and regulation on tracking and reporting of genomic data that are de-identified within HIPAA standards. The policy would require, for example, individuals’ consent for broad research use of de-identified information, and would require IRB review of proposed submissions for such data. The policy in effect removes some of the distinction between identifiable and de-identified information established under HIPAA, even requiring somewhat higher standards for handling de-identified information than HIPAA provides for identifiable personal health information (for example, 24-hour notification in the event of a breach in protection of de-identified data). The policy creates extensive review and notification requirements for investigators, institutions, and institutional review boards.Recommendation 6: Advances in health research will increasingly make use of genomic data that are integrated with information from many other sources, including from electronic imaging, PDAs and personal sensors, clinical and other databases; the policy should include a statement recognizing and encouraging such integration.As we consider future advances in genomic research, we hope to see better integration of sequence data with data from biomedical imaging and electronic sensors (including data from applications on patients’ cell phones), from electronic medical records, population level health data, and perhaps (or even probably) data from sources not directly encompassed by health systems (from workplaces, schools, grocery stores, etc.). It would be unfortunate if genome data sharing requirements served as a barrier to exploring the potential of these technologies, instead of a step forward in transparency. The point is not to stifle or shirk from encouraging such research, but to emphasize the need for more specificity in the types of data that would be covered by a proposed policy. The proposed data sharing policy lacks clarity or specificity on the scope of data to be shared, since data are generated from studies that include data on the phenotypic expression of particular gene sequences. In supporting calls to lessen the resulting administrative burden and provide better clarity on the scope and definition of any final data sharing policy, we must recognize that a comprehensive and effective policy for sharing of genomic data will in the end be more economical to the nation than a system that would erect obstacles to data sharing and encourage more duplication in effort. We believe, therefore, with FASEB and others that

identifying ways to reduce the administrative burden while promoting sharing of genomic data is the optimal approach.In summary, the AAMC recommends that NIH move expeditiously to revise the policy to allow for dynamic consent or other such innovations in patient participation, and to reduce further the administrative burden on investigators and institutions, and to reissue the revised policy for further public comment. The AAMC is again grateful for this opportunity to comment, and we look forward to working with NIH as it moves toward a final policy. Please contact Heather Pierce, J.D., M.P.H. at hpierce@aamc.org or Stephen Heinig at sheinig@aamc.org in my office with any questions about these comments. Sincerely,Ann C. Bonham, Ph.D.AAMC Chief Scientific OfficerReferences(1) AAMC, GWAS comments, Oct. 19, 2006.(2) AAMC, Comments on best practices on sharing genomic inventions, Jan. 4, 2005.(3) AAMC, Comments on “share my mouse”, July 31, 2003.(4) AAMC, Comments on enhanced public access to NIH research information, Nov.9, 2004.(5) AAMC, Memorandum on increasing access to the results of federally funded research, May 20, 2013. https://www.aamc.org/download/343538/data/ostpiomcomments.pdf, accessed Nov. 7, 2013.(6) Haga SB. Impact of limited population diversity of genome-wide association studies. Genetics in Medicine (2010);12:81-4.(7) Roman Isler M, Sutton K, Cadigan RJ, Corbie-Smith G. Community perceptions of genomic research: implications for addressing health disparities. North Carolina Medical Journal (2013); 74:470-6.(8) FASEB, Comment letter on the draft genomic data sharing policy, Nov. 6, 2013. http://tinyurl.com/p2c3a2e , accessed Nov. 7, 2013.

Commenter: Association for Molecular Pathology (AMP), Mary Steele Williams

Organization type or individual affiliation: Professional Society

Date of comment: 11/19/2013 15:09

Comment:

Section II. Scope and Applicability: Thank you for the opportunity to summit comments on the Draft NIH Genomic Data Sharing Policy. The Association for Molecular Pathology (AMP) is an international medical and professional association representing approximately 2,300 physicians, doctoral scientists, and medical technologists who perform laboratory testing based on knowledge derived from molecular biology, genetics and genomics. As a member of FASEB, AMP contributed to and concurs with its comments. In addition to the comments submitted by FASEB, AMP requests that NHGRI clarify that its proposed Genomic Data Sharing Policy does not apply to the submission of clinically-obtained genomic data to NIH-funded clinical databases.The Federal Register notice states that the policy applies to all NIH-funded research that involves large-scale human and nonhuman genomic data. Many of the public databases funded and curated by the NIH such as ClinVar include data collected from clinical testing. Because genetic testing laboratories typically lack direct interaction with the patient, these anonymized submissions may not have been specifically consented for genomic research. In these cases, such identification would entail violation of currently existing legal protections. Therefore, AMP requests that NHGRI clarify that its Genomic Data Sharing Policy specifically excludes submission of clinically obtained genomic data to NIH funded databases, unless the data itself is obtained through an NIH-funded research protocol.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Qasim Ayub

Date of comment: 11/20/2013 10:49

Comment:

Section II. Scope and Applicability: The Genomic Data Sharing (GDS) Policy is broad in scope and promotes sharing of data in a timely manner. I would advocate an even open access to de-identified data and do away with the two-tier system for open and controlled access.

Section IV.A. Data Sharing Plans: The data sharing plans are left a bit vague. Although there will be times when it would be difficult to meet the deadlines there should be a mechanism to ensure that this is not used as an excuse.

Section IV.B. Non-human and Model Organism Genomic Data: There should be no controlled access for such data.

Section IV.C. Human Genomic Data: With the burgeoning sequencing, transcriptomic and epigenomic data maintaining expensive layers of security for controlled data excess will become untenable. It would be simpler and cost effective to have open access with all users certifying that they will not attempt to identify or contact individual participants or use it for non-scientific pursuits and agreeing not to sell or profit from data thus accessed.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Regardless of whether the data is open or controlled access, all data should only be downloaded after researchers and institutes they work for agree to the NIH User Code of Conduct.

Section VI. Intellectual Property: Naturally occurring DNA sequences should never be patented.

Any other aspect of the draft GDS Policy: Sequencing of any genomic region in any human or non-human organisms or specimens should be covered by this policy.

Commenter: Joan Bailey-Wilson

Date of comment: 11/21/2013 0:08

Comment:

Section II. Scope and Applicability: The extremely broad scope of this policy will have many negative impacts. First, the expense of this program (to the NIH and to the funded investigators who must comply with these rules) will waste huge amounts of research funds that could be better spent on funding additional actual research. Linking the funding of science to the ability of researchers to obtain consent for very broad data sharing presents a model going forward where years of investment in sample collections and making good faith agreements with local communities could be destroyed. In addition, this connection to funding brings with it significant potential for financial conflicts of interest, as researchers are forced to balance the need for funds with the ethical considerations of the communities they study. With no ability to propose new genomics studies where the participants are allowed to enroll in the studies but REFUSE broad data sharing, many people will be prohibited from taking part in research without abrogating their rights to privacy and autonomy, which are major tenets of the Declaration of Helsinki. However, if an investigator allows participants to refuse dbGaP or other broad sharing of their genomics data, then that investigator will be punished under this policy with either not being funded at all or with only being allowed to perform genomics studies on a portion of their study participants (a biased subset of participants). This is simply bad for science and ethical considerations and is not outweighed by the benefits of broad data sharing.The policy hurts existing and future potential collaborations by reducing the incentive to form these large, collaborative groups in the first place. International collaborators gain little from forming formal collaborations with US groups, when they can simply wait and download the data after the short embargo has passed. But US researchers have no such access to data from other countries, which imbalances the negotiations for collaboration from the beginning.

Section IV.A. Data Sharing Plans: This policy discourages the collaborative nature of science that is essential to making significant advances against some of the most common diseases affecting Americans today. In addition, the resources required to support data sharing in many projects may make the study unfeasibly expensive, especially where extensive re-consenting of participants is required or where larger numbers of subjects must be approached for enrollment in order to identify the subset who will consent to broad data sharing. This sort of "sifting" of potential study participants to allow enrollment of only those willing to consent to broad data sharing will introduce biases into population-based studies and thus seriously impact the quality of the science. Also, acquiring re-consents from former participants to the study may pose a significant concern prohibiting further participation in biomedical research studies.The ethical implications for consenting study participants for such broad data sharing means that many studies may not be able to proceed because participants are not willing to consent to such broad data sharing, or that in the case of historical sample collections, participants may have died and it is therefore the decision of individual IRBs whether those samples can be used without consent. There are discrepancies in the IRB decision making processes among research sites in the United States; therefore, investigators in the institutions implementing rigorous IRB criteria will always be at a disadvantage compared to others. In many foreign countries, the use of samples from

deceased individuals is not permitted without previous consent before death. Requiring that ALL members of a family in a family-based study (quite important for modern genomics studies) means that the ability to get all members of complete, informative family to all enroll in a study will be even harder and more expensive than it currently is, because having only one or two critical members who will not enroll because of the proposed broad sharing requirements (with no ability to opt out of only the data sharing) will make that family useless. At present, the most expensive part of genomics studies is the enrollment and phenotyping of informative study participants. This policy is truly short-sighted in that it will cause large increases in study costs, especially since genomics costs are expected to continue to decrease whereas subject enrollment and phenotyping costs can only be expected to increase.

Section IV.B. Non-human and Model Organism Genomic Data: This is for Human subjects:3. Tiered System for the Distribution of Human DataMaking all NIH-funded research involving large-scale genomic data (but it is really small scale data - see above) available to third parties increases the risk of published results from these data being misinterpreted and/or misrepresented, especially in situations where the original data collection efforts are poorly documented. This is true even in controlled-access data situations, where secondary investigators only have to obtain initial approval from the NIH for data use, with no later assurance that analyses of these data without full knowledge of the data collection methods will lead to erroneous conclusions that are promulgated in the scientific literature.2. Data RepositoriesAlthough ostensibly the policy acknowledges that the NIH-designated data repositories need not be the exclusive source for facilitating the sharing of genomic data, investigators who elect to submit data to non-NIH-designated repositories are expected to confirm that appropriate data security, confidentiality and privacy measures are in place. It is unrealistic to expect that most labs will be able to set up their own repositories and so the only groups able to do this would be very large, commercially funded enterprises such as the Kaiser Foundation or 23andMe. Researchers working in universities across the country are unlikely in most cases to have the infrastructure required to support such a massive undertaking. For those labs, data submission to dbGaP "….no later than the time that data cleaning and quality control measures begin." will be an arduous task.

Section IV.C. Human Genomic Data: 1. Data Submission Expectations and TimelineThe unrealistically short time scales for data submission and embargo lifting will disproportionately impact small and moderately sized labs and research groups, that do not have the resources to analyze the data that they have invested significant time in funding and collecting, in the time allocated. Small labs will struggle to perform all of the necessary quality control and data analysis inside of six months whereas only the largest labs with more people and computational resources can publish their own papers in this timeline. This decision is unfair and creates more challenges for small research laboratories/investigators. It will especially impact junior investigators and will make it almost impossible for any but the largest, most well-funded institutions to compete for new data collection studies. This repression of the ability for junior investigators or investigators with excellent ideas but modest research staff sizes to be able to compete for genomics grants is bad for science and discovery.The GWAS policy repressed the ability for GWAS studies to be performed on many excellent well-phenotyped and unique datasets because adequate consent was not available for dbGaP submission and waivers could not be obtained (even though scientific review said these studies should be funded). However,

the GWAS policy did not apply to family studies and only applied to large-scale genotyping projects. The expansion of this policy to all types of genomic studies AND to quite small-scale studies will have serious implications and will affect scores of researchers in a negative manner. The new policy will apply to "Sequencing more than one gene or genesized region in more than 100 participants. OR More than 10,000 genes or regions from one participant (e.g., whole genome sequencing). OR More than 100,000 variant sites in more than 100 participants. These are TINY studies and thus will affect many junior investigators and investigators in small labs around the country. Trying to actually USE the shared data from thousands of such tiny studies will also have very little scientific validity.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: The concerns of the community about how the NIH would police the use of data granted under dbGaP have already been realized through some high-profile cases of breaking embargoes. In one particular case which has been covered extensively, a paper was retracted after the embargo breach was reported to the journal in question (PNAS) and the individual was sanctioned by suspending the investigators access to dbGaP and all work with the downloaded data was to be ceased. However, the breach did not seem to heavily impact the career of the researcher responsible, Dr Zhang is still employed by Yale and continues to receive NIH funds. And despite the retraction, it is still possible to find the paper online, albeit with a tag labeled “See Retraction Published September 9, 2009.”. Therefore, we believe that thepenalties for breaking data embargoes are poorly defined and clearly insufficient in the light of this case. It is essential that these policies be reviewed and strengthened in the new data sharing policy.

Section VI. Intellectual Property: 4. Informed ConsentI responded to the proposed data sharing policy for GWAS back in 2006 and a number of the concerns I raised then not only remain unaddressed but are in fact more pressing than ever as the depth of genomic data covered by the new proposed policy has been significantly increased. I am still concerned that deidentified data with such deep genotyping or sequencing is not truly deidentified since genotypes themselves could in fact be identifiers in association with some other identifiable information available publicly. In fact, a number of articles in the media addressed a paper published in Science in January 2013 that was able to identify a number of individuals from the Center for Study of Human Polymorphisms (CEPH) family collection whose genomes were sequenced as part of the 1000Genomes Project (Melissa Gymrek et al., Science 339, 321 (2013)). Deposition of data into U.S. government databases also carries the risk that U.S. Federal Law Enforcement agencies can legally search those databases without a court-ordered subpoena, whereas a subpoena is required for those agencies to obtain access to data stored in non-Federal databases. The policy still does not address the concerns about storing biometric identifiers of non-citizens in U.S. Federal databases, which may deleteriously affect international collaborations. Recent events have turned the spotlight on how U.S. Law Enforcement agencies have conducted their activities and this data sharing policy does nothing to assuage the concerns of researchers to whom it applies.I also believe that it may be impossible to obtain truly informed consent under this model of data sharing, as it is impossible to fully quantify the risks presented to participants if their data were to be deidentified. We cannot in good faith promise that their data will remain anonymous. Many participants may not be willing to accept these risks, and those that are willing to consent may not fully understand what they are consenting to as we ourselves cannot predict all the consequences of such broad data sharing. . In addition, a major portion of disadvantaged

population participating in biomedical research studies and a considerable portion of the general population may not have the necessary educational background to understand the informed consent to the extent where all consequences of such broad data sharing will be understood. Therefore, an ethical concern remains: to what extent we are using the term 'informed consent' administered to the subjects to understand such implications to participate in biomedical science research. Although a Certificate of Confidentiality has been mentioned, the Genetic Information Nondiscrimination Act (GINA) is only applicable to a business with 15 or more employees (http://www.eeoc.gov/laws/types/genetic.cfm).

Any other aspect of the draft GDS Policy: 7. Exceptions to Data Submission ExpectationsAlthough the NIH acknowledges that in some cases circumstances beyond the control of investigators may preclude submission of data to NIH-designated data repositories, the section of the policy describing exemptions seems both incoherent and potentially damaging to international collaborations. Section IV.C.2. suggests that investigators are not necessarily to be forced to submit their data to NIH-designated databases; however Section IV.C.7 seems to suggest that in fact investigators will not be able to simply "elect" to submit their data to non-NIH-designated data repositories but will have to justify this as an "exception". These two positions seem inherently contradictory. In addition, no recognition is made for situations where because of insufficient consent or legal requirements, data are not permitted to be shared at all. For researchers with long-standing data collections, this requirement to share essentially forbids them from being able to apply for federal funds to conduct their research. Cutting off researchers who cannot comply with broad data sharing policies has already caused many long-running epidemiological studies to refrain from certain types of research because they are unable to re-consent their participants. This new proposed policy will only make that problem worse. . Many collections with a large and esteemed track record, and at least partially NIH-funded in the past, will now be unable to move forward.These issues should be examined in more detail, and SUBSTANTIAL revisions ought to be made before adoption.

Commenter: Daryl E Beeson

Date of comment: 10/23/2013 13:34

Comment:

Section II. Scope and Applicability: I applaud the NIH for instituting this long needed policy. Though part of the general public but, with an extensive technical background in computing and storage, I have long wondered why the NIH hadn't instituted this policy for all NIH related research. Imagine how we could have accelerated cancer related genetic research by sharing information globally!

Having an avid scientific curiosity and Chief Technology Officer at a technology based company, I have wondered how I might be able to help in some regard to guiding the direction of these activities. I am confident that you feel you have the staff to make these decisions and guidelines but, if you ever decide to investigate public oversight, I would make myself available.

Section IV.A. Data Sharing Plans: It seems that all NIH related investigations, not just limited to genomic research would benefit from global data repositories.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: I was happy to see that you have a provision for stopping funding if the policy isn't followed!

Section VI. Intellectual Property: Organizations waving the IP flag should not be using NIH funds for research. The funding for this research comes from generic sources like taxes. If an orgainization wants to retain IP for testing than they should use private funds for their research.

Any other aspect of the draft GDS Policy:

Commenter: Michael J. Bell, MD

Date of comment: 11/8/2013 17:23

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: I don't have many concerns, but there are a few. I'm really only going to address the human issues and leave the non-human issues to others.First, the genomic material is a terribly finite resource. While we know that a few micrograms of DNA can be used by investigators to get the relevant material they need to identify the gene of interest, this will become a limitation as material is used again and again for studies. I'm not sure how this policy will handle the prioritization of release of DNA to investigators to do their work – and I cannot really envision how anyone could make such a "Solomon-like" decision.Second, related to above, I'm not certain how the material will "age" if it is freeze-thawed multiple times in the processing of the tissue.Third, while I recognize that genes cannot be patented based on recent US Supreme Court rulings, I think that the Intellectual Property of the data might be something that my University might be concerned over. If the NIH hasn't already reached out to large institutions for these comments, it probably should.Lastly, the IRB at the University of Pittsburgh basically refuses to allow for banking of samples unless a specific hypothesis is being asked and answered. As an example, if we wanted to do a serum biomarker study, our consent form and IRB application specify that these tests are being done and the sample size calculations. There can be stipulations that the extra material will be saved for other analyses, but I'm not sure that it would envision free sharing of material across the world like this. Maybe it would be fine with our IRB, but there would be concerns of DNA and children (since loss of confidentiality can affect a large portion of their lives) and central repositories. I'm not sure it couldn't be overcome, but there would need to be care in the presentation of the information to IRBs that deal with children and this issue.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Lyle G. Best

Date of comment: 11/11/2013 14:33

Comment:

Section II. Scope and Applicability: Dear Sirs:I began my Indian Health Service (IHS) career in 1977 with the Turtle Mountain Band of Chippewa (TMBC) community at Belcourt, ND. I am trained as a family physician; but have always had a particular interest in genetics and published a number of papers on genetic conditions found in this community during my career with IHS. When I retired from the IHS in 1998, I developed a self-supported, yearlong “sabbatical” at the DNA Diagnostic Lab and Clinical Genetics department in Winnipeg, Canada. Returning to our home in the Turtle Mountain area, I joined the adjunct faculty at Turtle Mountain Community College (TMCC) and taught the “Intro to Human Genetics” course from 1998 to 2012.I have participated in a number of NIH-funded clinical research efforts since 2000 (Strong Heart Study (SHS), Dakota Center PI) and in 2004 established an active genetics laboratory at TMCC, funded by an INBRE subcontract with the University of North Dakota (NIH, P20 RR016741). This project is investigating the possible effects of genetic variants on the risk of pre-eclampsia through a case/control study of 550 participants from the TMBC. Our lab facilities include 500 square feet instrumented with 2 Bio-Rad real-time PCR (48 sample) and 2 additional thermocyclers, a Bio-Rad chip-based capillary electrophoresis (Experion) and ABI 310 capillary electrophoresis instruments. Tribal college students at TMCC (over 25 thus far) are routinely employed on a part time basis, conducting real-time genotyping work that is the basis of this and other genetic epidemiology projects. More recently, I have initiated a case-control study of potential genetic effects on pediatric asthma at the Cheyenne River Sioux community, with genotyping conducted at TMCC and funded by an NIMHD grant to the Collaborative Research Center for American Indian Health (CRCAIH) at Sioux Fall, SD. TMCC has also conducted work related to pharmacogenetics in collaboration with the University of Nebraska. I am currently exploring (and receiving substantial interest in) the possibility of establishing a bio-repository that will be operated as a cooperative in two American Indian communities. This work has resulted in first authorship on 12 publications solely related to genetics in American Indian (AI) populations; and co-authorship with a TMBC tribal member who has nearly completed her PhD in Epidemiology at the University of Minnesota.

Section IV.A. Data Sharing Plans: Our participants and these communities have also been open to wider collaboration, allowing samples and data to be analyzed in Europe, as well as by many investigators in the US.I offer this information as evidence of:1) willingness on the part of AI communities to engage and support research, and in particular genetic research2) what I believe is the extremely powerful influence of a long-term, trusting relationship with investigators3) the efficacy and efficiency of utilizing and incorporating local clinical providers in the conduct of research in rural, and/or minority communities.Since at least 2004, my colleagues and I have attempted to inform tribal leaders and community members of the NIH emphasis, and more recently requirement for "data sharing". I enclose this term in parentheses since I have never experienced a refusal on the part of tribal entities to share their data and samples in traditional forms, such as publication and collaboration to further the work that they have consented to. Where I begin to notice indifference (if not hostility) is when explaining the

requirement for data submission to a public (or "controlled access") database accessed by unknown individuals with no direct control by the investigators they know and trust. I use the phrase "…attempted to inform…" because I believe it is my responsibility only to inform communities of the facts, without encouragement or bias in any particular direction. A number of our SHS tribes have formally requested a waiver of the NIH genomics data-sharing policy and been rather summarily denied without further explanation. There has been no attempt to utilize the government to government consultation process mandated by presidential directive to negotiate or resolve these differences.

Section IV.B. Non-human and Model Organism Genomic Data: Not applicable

Section IV.C. Human Genomic Data: Please see comment #7 below for a comprehensive response.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Please see comment #7 below for a comprehensive response.

Section VI. Intellectual Property: Not addressed.

Any other aspect of the draft GDS Policy: I will offer some suggestions and observations:1) The world has seen a surplus of ideologically driven crusades, such as that of the International Monetary Fund and World Bank to privatize what many countries long sought to provide on a communal basis. I realize and accept the common good that “data sharing” has provided throughout the history of science; and generally encourage newer, innovative forms, such as the human genome project, open publishing etc. I do not believe it is in society’s best interest to implement these new forms in an abrupt and heavy-handed fashion. 2) Minority populations have experienced and continue to experience discriminatory and inequitable treatment that generates considerable skepticism regarding majority mandates that are promulgated without any real consultation.3) The dominant institutions in our society may be perfectly confident in the correctness of their latest direction; but they should also be cognizant of the fact that minority communities possess unique genetic (and other) attributes that bring considerable value to research. In addition, if the dominant society has any genuine intention to address health disparities, it can’t be done without the cooperation of these communities.4) Those that are enthusiastically embracing these more encompassing and expansive methods of “data-sharing” would be wise to reflect a bit on some of the unanticipated twists and turns in this road thus far. Examples are the demonstrated ability to identify individual HapMap participants, recent need for accommodations with the family of Henrietta Lacks, not to mention the seeming impossibility of making any web-based data completely safe from hackers.5) Lastly, I sense a persistent under-tone in this policy driving those receiving funding to comply with these new modes of data-sharing; but a reluctance of NIH to share in the responsibility if anything goes wrong. Thus the investigators and their institutions will be required to “certify” that they have properly de-identified their data. Investigators are urged to obtain a “Certificate of Confidentiality” for especially sensitive data; and yet there have already been instances where these certificates have been abrogated and nothing in the law is “guaranteed”.In conclusion, I suggest that NIH move very cautiously in this area with much more consultation with minority populations and their leaders. This consultation should be instituted on a tribal government to

federal government basis as required by the constitution and NIH policy. It would also be very helpful if NIH appeared to be aware and more appreciative of the “data-sharing” and generosity of American Indian (and other minority communities) to date.Sincerely,Lyle Best, MDPrincipal Investigator, Strong Heart Study (Dakota Center)Principal Investigator, Genetics and Pre-eclampsia Study (TMCC)Principal Investigator, Factors Influencing Pediatric Asthma, (CRCAIH)

Commenter: BioMed Central, Amye Kenall

Date of comment: 11/20/2013 11:35

Comment:

Section II. Scope and Applicability: We would also like to suggest that not only should data be made publicly available but also the analysis workflows, software, and pipelines used to create the data. In addition, we would like to suggest that researchers receive more guidance on metadata standards and the resources needed to support and implement those standards for their data (see comment 2).

Section IV.A. Data Sharing Plans: At BioMed Central we have seen the incredible benefit of and work involved in making data reusable. We would like to recommend that in the NIH's data sharing plans they take into consideration the resources needed to invest in metadata to make data reusable. In this case, not only will guidance be needed as to which standards to use, but training will also be needed. The draft policy does recommend investigators contact their university for guidance, and though we do believe the time is right for genomic data to be more publicly shared, we would suggest the NIH work very closely to ensure universities are offering researchers the resources they need to make their data publicly available and re-usable. This infrastructure, especially for big data, is only starting to exist at many research centers.

For metadata standards, we would particularly recommend the NIH work with Biosharing (http://www.biosharing.org/) and ISA Tools (http://isa-tools.org/). We at BioMed Central also have much experience in integrating data, metadata, and scientific workflows into the published articles, especially through our journal GigaScience (for examples, please see our papers "Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species" (http://www.gigasciencejournal.com/content/2/1/10) as well as "Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification" (http://www.gigasciencejournal.com/content/2/1/4), about which we would be happy to talk with you further).

Section IV.B. Non-human and Model Organism Genomic Data: No comment.

Section IV.C. Human Genomic Data: We would also like to comment on the timeline for data release. In many cases this is 6 months after data submission to a repository, which can be up to 3 months after data generation. For some stages of data (for example, after initial round of analyses or computation), researchers are already releasing their data immediately. Nine months can be a very long time to delay research, especially when this affects human life. One example of the lives that can be saved in immediately opening up access to data at this level and further into analyses can be seen in the German E. coli outbreak of 2011 (http://blogs.nature.com/news/2011/06/the_german_e_coli_outbreak_40.html). Also see the wiki on GitHub for the crowdsourcing done here (https://github.com/ehec-outbreak-crowdsourced/BGI-data-analysis/wiki).

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: No comment.

Section VI. Intellectual Property: At BioMed Central we use a variable/combined license agreement for our journals, which places data in the public domain under CC0 and the remainder of paper under CC-BY. Although in the US data cannot be copyrighted, this is not true across the globe. We therefore would highly recommend the NIH at least recommend that all data generated under its funding be placed in the public domain under a CC0 waiver. We held a public consultation on this licensing change and published an editorial on this: "Licensing the future: report on BioMed Central's public consultation on open data in peer-reviewed journals." (http://www.biomedcentral.com/1756-0500/6/318)

Any other aspect of the draft GDS Policy: As an Open Access publisher, we are very pleased to see that the NIH is taking these progressive steps towards opening up genomic data and thus spurring research forward. Please do not hesitate to contact us at BioMed Central (amye.kenall@biomedcentral.com) if we can further support this endeavour. We will continue to work on strengthening the link between publications the various research outputs created by your funding, including data.

Commenter: Bryony Borneo

Date of comment: 11/19/2013 13:08

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: (1) The designated NIH-repositories are well described in section IV.C.2. It would be helpful to have more details on the non-designated NIH-repositories including examples/case studies and definition of a repository. For instance, it is not clear, if the data can be hosted on a FTP server (at the submitting institution or hosting provider) or on the cloud, i.e.as a public dataset collection via AMAZON cloud services (see for example the Annotated Human Genome Data provided by ENSEMBL (http://aws.amazon.com/datasets/2315?_encoding=UTF8&jiveRedirect=1)(2) In the GDS webinar you mentioned a 'Trusted Partner System' as part of contracts. Could you please provide more information about this partneringsystem including eligibility, security compliance, and other technical aspects (e.g. how the data has to be hosted/shared by the contracting institution?).

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Broad Institute, Stacey Donnelly

Date of comment: 11/20/2013 17:33

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Broad Institute Comment: Respect for the autonomy of individuals in relation to research participation is paramount in any scientific endeavor. One of the eight required elements of informed consent, as described in 45 CFR 46. 116, is key to applying this principle to genomic research: a statement describing [to the prospective research subject] the extent, if any, to which confidentiality of records identifying the subject will be maintained. A major challenge for the research community is to ensure that required consent elements are appropriately represented, while at the same time adhering to another mandate of 45 CFR 46. 116: "The information that is given to the subject or the representative shall be in language understandable to the subject or the representative." We therefore request that the NIH provide specific template language that succinctly describes the concepts of both open and controlled access in a way that research subjects can readily comprehend. IRBs are in almost unanimous agreement that consent forms are already too complex and lengthy; further guidance is needed from the NIH about how to adhere to the GDS policy in a way that does not further overtax an already formidable consent process. Provision of template language would also serve to ensure greater consistency across institutions and individual research protocols. Broad Institute Comment: Theoretically, because this mandate affects prospectively collected specimens and new cell lines, provision of information to patients regarding future research use and broad data sharing should be straightforward. However, we again request that further guidance be provided regarding how to implement these requirements. Specifically: Does NIH envision use of a blanket consent for all future research uses as the appropriate mechanism to fulfill this requirement? We strongly encourage NIH to provide template consent forms that accomplish the desired outcome. Will clinicians be expected to administer this consent, or will institutions be required to have "on-call" research administrators, who would be responsible for reviewing this information with prospective subjects? If the former, will clinicians with no research experience, little time available, and no specific information about potential future uses, be able to adequately respond to questions that might arise? Acknowledging that consent is not simply a form, but an ongoing process, who will be responsible for addressing questions that a subject may have after signing the consent –the surgeon who performed a tissue biopsy days, weeks, or months earlier? Broad Institute Comment: We appreciate the flexibility of grandfathering old collections and believe it will considerably expand access to valuable data generated from samples collected before the advent of the genomic technologies used today.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Draft Policy: Institutions should indicate in the certification whether aggregate genomic data from

datasets with data use limitations may be appropriate for general research use (i.e., use for any research question such as research to understand the biological mechanisms underlying disease, development of statistical research methods, the study of populations origins). If so, the aggregate genomic data will be made available through the controlled-access compilation of aggregate genomic data to facilitate secondary research. Broad Institute Comment – Because concerns regarding potential identifiability are considerably reduced when genomic data is pooled, the Broad Institute supports broad sharing of aggregated data. The draft GDS policy indicates that institutions will be responsible for determining whether this data should be made available for general research use, and that this determination be described in the Institutional Certification. We propose that rather than individual institutions being required to explicitly state this in their certifications (which burdens the institutions and may lead to inconsistent determinations), general research use of aggregate data be the default assumption underlying any Institutional Certification. If circumstances were such that general research use would not be appropriate for a specific aggregate data set, the Institutional Certification would then be required to explicitly indicate this restriction. In the absence of an explicitly stated restriction, all aggregate data certified by an Institution would be available for general research use.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Establishment of a Standardized Ontology for Data Use Limitation Letters Establishment of a standardized ontology for Data Use Limitation letters is not a point addressed in the draft policy, however we would like to take this opportunity to advocate for its inclusion in the final policy. It has been our experience that when IRBs are charged with interpreting informed consent forms to ascertain appropriate future use of a study's data, in the absence of a standardized menu of choices, interpretations run the gamut from clear and precise, to vague and internally inconsistent. While we appreciate that dbGaP's "Basic Study Information" (BSI) form does provide a menu of options from which researchers may select when entering data into dbGaP, in the absence of a corresponding list from which IRBs must also select acceptable uses, the burden falls upon individual research teams to interpret the IRB's interpretation of appropriate data uses. We believe there should be a one-to-one correspondence between the BSI form and an NIH data use limitations template. While "Points To Consider for IRBs and Institutions in Their Review of Data Submission Plans for Institutional Certifications" is useful and certainly well intended, we believe the current open-endedness of this guidance can lead to considerable downstream confusion.

Commenter: Laura Bull

Date of comment: 11/19/2013 13:05

Comment:

Section II. Scope and Applicability: 1. I wish there were more clarity r.e. the issue of sample sizes to which the policy applies; I think this is especially an issue for rare disorders. Maybe the policy is clear- i.e. no matter how small the sample set, the data must be posted- and I just somehow missed it? For example, if new NIH funds are used to do some sort of whole genome study of a single individual or small family with a very rare, and therefore potentially identifying disease diagnosis, does that data need to be deposited? What if a family is interested in participating in a study, but does not want their genetic data deposited on the web? Do we need to explain to them that there is no other option, and that they cannot participate in an NIH-funded study unless they agree to have their genomic data posted?

2. The issue of 'legacy samples'- i.e. samples which were consented prior to the establishment of this and similar NIH policies. If reconsent is not practical, can such samples be studied at the whole-genome level using NIH funds, perhaps if a data sharing/collaboration policy is established that the enrolling institution (i.e. University, etc) is comfortable with, although that policy might not allow deposit into an NIH database? For example, it might include a requirement that the applicant interested in data access apply to the IRB of the enrolling institution for access to the data. Or if someone wishes to study such samples, must non-NIH funds be sought for that study?

3. Is there any flexibility allowed when working with 'special populations' which may have cultural issues regarding wide sharing of whole-genome data, even if they are interested in participating in genetic studies of disease? Again, would the choice be that they must agree to having their genomic data available in dbGAP, etc or else they cannot participate in the NIH-funded studies?

Section IV.A. Data Sharing Plans: Are the proposed timelines really realistic? Especially the idea that the data must be released within 6 months of data cleaning, regardless of whether anything has been published from it or not, seems unreasonable to me. I understand that it is a good idea to encourage folks to publish and make their data available ASAP. However, it can certainly take longer than 6 months to go from having cleaned data to completing sometimes complex multi-tier analyses (which are often performed by specialized and extremely busy research teams with wait-lists of projects for which data is awaiting their analyses), performance of follow-up experiments sometimes required to facilitate publication, preparation and submission of a paper often involving contributions from a large number of folks, and actually getting that paper accepted, given that it is quite common to need to submit successively to more than one journal to find a home for a study. Of course, I suppose one solution to this is that

people will submit their data to lower-prestige journals, with no follow-up work, simply to get it out in time. Maybe that is what NIH wants?

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: It is unclear to me exactly how these proposed regulations differ from what is currently required.

Commenter: Linda Burhansstipanov and Lynne Bemis

Date of comment: 11/4/2013 13:01

Comment:

Section II. Scope and Applicability: We are concerned that potentially innovative grants can be excluded (withholding of funding) if the organization or community has ordinances prohibiting sharing of specimens. Although many tribal Nations have no formal genetic specimen sharing policy, other have formal or informal language restricting sharing of specimens until protection of privacy for both the individual and tribal nation is improved. A tribal program may be interested in collecting specimens with designated research partners, but have issues with the specimens being in a repository from which anyone can access the specimens after receiving overall approval from NIH. What if a local tribe desperately wants to help young tribal scientists investigate diabetes, albinism, cancer, or obesity (common topics of research interest to most tribal and urban Indian programs), but they want to limit research to a single study or to a single area of study (breast cancer)? If they do not want to share with anyone who likely knows little about local tribal cultural issues, they are likely to not be viewed as competitive for NIH funding. The language within this policy needs to be broader to respect such concerns and to not continue to discriminate against Natives who have concerns about sharing specimens.

Section IV.A. Data Sharing Plans: We have no issue with data sharing plans as long as they allow limitations of the sharing of specimens to others living anywhere in the world who may or is likely to know little or nothing about Indigenous health issues.

Section IV.B. Non-human and Model Organism Genomic Data: Tribal Nations did not appear to have issues with microbial data; a few tribes were concerned that this includes animals.

Section IV.C. Human Genomic Data:
No major issues with this section until section 3. "Tiered system for the distribution of human data", specifically the "controlled-access data in NIH-designated data repositories." The comments we received were"Tribes could go through controlled access, but requires negotiating with NIH on every requirement. So, NIH makes the decision about who gets the data? Does the tribe have a voice? This can be interpreted many different ways … how keep some tribal decision-making in control of specimens?? This sounds like there is no such role once the specimens are turned over to NIH. Will subsequent NIH program officers interpret this the same way as current NIH program officers?" So, it is a bit vague and the concern is that current NIH employees may understand what this is and is not allowing, but will "new hires" to NIH .... Is it feasible for tribal nations to be approved for NIH funding for a single study only? Or limited to a specific disease, such as breast repository, but not used for other diseases, such as what happened to Havasupai... and specimens used for purposes not approved by tribal Nation.We thought several of the paragraphs on informed consent were improvements over current practices. Thank you."While the NIH encourages broad access to genomic data, in some circumstances broad sharing may be inconsistent with the informed consent of the research participants whose data are included in the dataset" -- this is very relevant for some Indian

Nations 6. Data Withdrawal section. "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 ". If the specimen is really coded and phenotypes removed, how is this really feasible? What about tribes with ordinances prohibiting sharing?7. Exceptions to Data Submission Expectations. Could a researcher examine a specific set of markers and then challenge AI's identify as AI especially if the markers were drawn from a study that may or may not have comprised AI specimens. Markers linked to condition rather than to race. "In such cases, investigators should provide a justification for any exceptions requested in the application or proposal." If tribal nation keeps data within own system and security, how much access does NIH have? Can the tribe be awarded funding and NOT share with researchers outside of NIH? If tribal nation keeps data within own system and security, how much access does NIH have? Can the tribe be awarded funding and NOT share with researchers outside of NIH? Will the tribes or urban programs that request exceptions truly be considered "competitive" for NIH funding?

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: How many ethnic and racial minorities serve on the NIH Data Access Committee (DAC)? "• Not attempting to identify individual participants from whom the data were obtained" Does this include identification of race?

Section VI. Intellectual Property: Excellent section. Very nice. Thank you for conscientious effort, particularly in this section.

Any other aspect of the draft GDS Policy: For projects that involve ... and bullet list follows ... the question we have is, So what is actually excluded? Sounds like everything or almost everything is included???For bullet 2 in this same list, So if specimens were less than 100, they are excluded from the policy? What is the rationale for these numbers? Are more than 100 specimens considered within genomic research realm rather than genetic research?The table within "Expectations for Data Submission and Data Release" seems to be emphasizing 'genomic' research and eliminates researchers who are looking at smaller 'genetic' studies. Is that correct?Overall, it seems that about half of the tribes have no issues with sharing specimens. For those that do, we are concerned that if they want to start by having specimens only shared with a research institution they trust and no one else, that they may not be competitive for funding from NIH ... These tribes are basically not satisfied with the current level of safety and protection of privacy of the specimen for the individual and ramifications some research could have on the tribal nation. These communities are NOT STUPID, they have a different opinion and perspective about specimens. They are very interested in helping reduce diseases in their respective communities, but want active informed consent processes used. They don't want to be bullied into data sharing, nor do they want to be discriminated against. That is the Readers' Digest Version. Hope it helps.

Commenter: Wylie Burke MD PhD, Rosalina James PhD, Kenneth E Thummel PhD

Date of comment: 11/20/2013 21:35

Comment:

Section II. Scope and Applicability: We support the goal of the NIH data-sharing policy to expedite research by ensuring that genome-scale data generated by NIH funding are broadly available to the research community. However, we believe the policy is short sighted in its focus on submission of data to dbGAP or other federally approved data repositories. Rather, we believe that efforts should be undertaken to develop alternative approaches to data-sharing, to address concerns of tribal communities and other minority groups. Our perspectives are informed by our participation in university-community research partnerships with several tribal organizations in the Pacific Northwest.The policy proposal notes a strong commitment to the protection of research participant confidentiality, which we applaud. The approved repositories require that data be stripped of HIPAA identifiers. Access to individual-level data is restricted, requiring approval of a Data Access Committee, and any researcher who accesses data must affirm that s/he will not seek to re-identify participants. Data access is further limited to research projects that are consistent with the informed consent document that was used when participants were recruited. However, we are concerned that the policy provides insufficient protection for groups that have experienced research-related stigma or discrimination; fails to create opportunities for community participation in oversight of data use; and lacks adequate attention to researcher accountability. Group HarmThe policy states that a submitting IRB must affirm that "To the extent relevant and possible, risks to groups or populations associated with data submitted to NIH-designated data repositories were considered." However, the policy does not specify what actions would follow if the submitting institution's IRB determined that submission of the data posed potential group harm. Nor does the policy identify risk of groups harm as a justification for additional oversight or the choice not to submit data to a federally approved data repository, or any other protective measure. These issues are of concern to American Indian/Alaska Native and other minority groups because racial/ethnic identifiers are not removed in the de-identification process. Therefore, the data can be used to make comparisons between different racial/ethnic groups. There are many unfortunate examples of invidious comparisons or interpretations of research data that have been construed to stigmatize minority groups or justify discrimination. Tribal and other minority organizations have a responsibility to ensure that study designs using racial/ethnic identifiers are appropriate, and that publications deriving from such studies are free of discriminatory interpretations.

Section IV.A. Data Sharing Plans: Lack of Community Participation in Data Use DecisionsIn the federally approved repositories, researcher access to the individual level data requires approval of a Data Access Committee composed of federal employees. Once data are submitted to the repository there is no opportunity for a community, tribal leadership, or a local IRB to be involved in decisions regarding data use. The Data Access Committees may therefore lack adequate knowledge or expertise to ensure that data uses are responsible and in keeping with permissions given at the time of data collection. Further, this approach contradicts the practices of collaborative research based a partnership ethic between tribal governments and researchers. As a result, tribal and other minority communities are wary of delegating assessment of cultural

or other harms to federal employees assigned to Data Access Committees. Researcher MisconductThe policy does not specify any audit mechanisms to ensure that data use policies are upheld. As a result, researcher misconduct may go undetected. Nor or consequences for researcher mis-use of data are specified. While misconduct is likely to be rare, appropriate measures should be in place to identify and respond when it occurs. Tribal leaders are protective of rights to aspects of their property or culture. They are cautious of submitting genomic data to repositories that lack enforceable measures for researcher accountability.ExceptionsThe proposed policy notes the possibility of exceptions to the requirement to submit data to an approved repository such as circumstances in which state laws prohibit submission to a federal repository. Alternative data-sharing approaches are expected when exceptional circumstances preclude submission of data to approved repositories, but the policy does not sanction the creation of alternative data-sharing approaches that would include local tribal or community review of proposed data uses. The policy also encourages prospective use of broad consent – that is, a consent process in which participants are asked to give permission to have their data submitted to a federal repository for any research use approved by the Data Access Committee. These exceptions do not address the concerns of researchers engaged in partnership-based activities involving universities and tribal organizations.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: SummaryThe policy seeks to ensure broad availability to researchers of genomic data from NIH-funded studies, with no participation of communities in oversight of data access or publications deriving from data use. While the goal of data sharing is commendable, the lack of opportunity for local IRBs or organizational leadership to request review of study proposals is problematic, as is the lack of an alternative approach involving tribally supervised data repositories. We believe the approach of a single federally controlled approach to data sharing is unnecessary and harmful. It creates a strong disincentive for minority communities and their partners to pursue the types of genetic research covered by the policy. As proposed, the NIH data sharing policy is inconsistent with agreements we have with tribal partners, developed to support mutual trust. We believe NIH should allow diverse approaches to data-sharing, including opportunities to create data repositories that incorporate tribal or other community involvement in the review and approval of data access and use. We believe this approach safeguards against group harm, gains the benefit of community input into the research process, and offers a stronger platform for researcher accountability.

Commenter: Nicola Camp

Date of comment: 11/19/2013 13:22

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: For the data to be meaningful to a third part user, the process of calling and quality control steps is imperative to have access to -otherwise findings may merely be differences in workflows and anomalies. Such anomalies can over-power true signals -rendering the sharing of these data useless, or worse, misleading.The resources required by a funded investigator to ensure that the uploaded data (genomic and associated phenotypes) remain meaningful are substantial, and these tasks are time consuming. Appropriate funds, above and beyond those made for the project science, should be made available to investigators and their teams to carry this out. This will ensure that this data sharing effort does not unduly detract from the investigators primary line of study. Similarly, substantial infrastructure funding will needed on the NIH end to support the sites that house the data. If the data are uploaded, but cannot be searched or extracted easily... then the data won't be used.

Commenter: Neil E. Caporaso

Date of comment: 11/20/2013 17:22

Comment:

Section II. Scope and Applicability: The scope of the proposed changes are enormous. While we strongly favor data sharing, the proposed changes seem excessive. The policy includes human and nonhuman data and virtually all genomic categories of data. It also includes grants, contracts and intramural research.

Section IV.A. Data Sharing Plans: The data sharing will require an increasing proportion of budgets in a period when overall funding is dropping. We are steadily losing scientists who fail to gain support. These requirements place added unrealistic burdens on the survivors. Excessive and unrealistic sharing mandates will impede progress rather than promote it.

Section IV.B. Non-human and Model Organism Genomic Data: Assigning the date of publication for model organism data sharing is realistic and reasonable. Earlier dates could be considered in the event of unusual delays or other special circumstances. forcing an arbitrary early release is

Section IV.C. Human Genomic Data: It is unrealistic to register when 'data cleaning and QC begins'. The policy grossly underestimates the time and effort required to clean newer forms of genomic data. Releasing raw data in early stages is a disservice to science as the data is often interpretable, contains errors, or is otherwise unsuitable for use. Further, it is unclear if current data repositories are capable of accepting, storing newer types of data.It is much more realistic to provide data at the time of publication. This is realistically linked to the investigators own strong drive to publish and insure that the data have achieved an acceptable threshold of validity, annotation, etc. With rare exceptions, sharing at an earlier stage is a gross waste of effort.The policy indicates that essentially all forms of genomic data are to be included. These data have enormous differences in terms of applicability, size, error, rapidly of technological evolution, and many other factors. Some of these datasets are enormous in size. Often, they are difficult or impossible to interpret without extensive annotation. The policy seems to demand early submission followed by early release- potentially causing chaos as poorly cleaned or highly preliminary forms of data are forced into the public domain.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property: no comment

Any other aspect of the draft GDS Policy: The expectation that DNA sequence data be released 'within 3 months of generation' (i.e., level 2) is unrealistic, wasteful of time and effort, and will result in poor quality data being submitted. Also, it may result in duplicative and wasteful impositions on investigators, those who store the data, and those who eventually try to use it.The level 1 holding period of 6 months is too short. (de novo seq.)In general, the policy has the feel of an enormous unfunded mandate that uses the laudable end of enhanced data sharing to

impose a crushing new burden on already stressed geneticists. The requirement to submit and release early data prematurely are difficult. It is unclear what entity can reliable store and make available the extensive types and quantities of data expected to be generated by this policy.I advocate data-specific sharing guidelines, realistically linked to storage facilities optimized for the diverse types of data to be solicited. In this way some uniformity, standardization, appropriate annotation, and other quality thresholds can be specified insuring the ultimate product can be useful to the community. A shotgun approach (in terms of the types of data, the nonspecific submission requirements...) will not serve science.

November 20th, 2013

Submitted by email to: GDS@mail.nih.gov

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

RE: FR Document 2013-22941 [Draft NIH Genomic Data Sharing Policy Request for Public Comments]

Dear Members of the Genomic Data Sharing Policy Team,

We are writing as a multidisciplinary group of faculty and administrators from academic health science centers who work together through the CTSA Consortium's Biobanking Working Group of the Clinical Research Ethics Committee. We acknowledge the importance of broad data sharing to further our understanding of how genetics influences human health and disease, and request consideration of the following critical points relevant to the Draft NIH Genomic Data Sharing Policy (the "Draft Policy") (Federal Register 78:57860, 20 September 2013).

1. **Sponsor public education programs about the data sharing policy.**

We strongly encourage the NIH to assist researchers and participants by sponsoring a public education program to elevate understanding of the value of:

a. sharing research data broadly and widely, and

b. open access to research data generated with public funds.

Public education is essential to facilitate a publically acceptable approach.

2. **Provide sample informed consent language to facilitate meeting expectations.**

Given increasing evidence that research consent forms can and should be simplified to increase participants' comprehension, the practicality of implementing the following proposed expectation (in Section C.4 of the Draft Policy) needs careful attention.

> "The NIH expects the informed consent process and documents to state that a participant's genomic and phenotypic data may be shared broadly for future research purposes and also explain whether the data will be shared through open or controlled access."

If this or any similar expectation remains in the final Policy, it would be helpful for NIH to provide an example of concise language that would be acceptable to achieve the stated expectation.

3. **Aim for consistency and clarity across all policies governing use of humans in research.**

We also request that NIH be mindful that expanding the obligations of researchers, to explain

in detail the various ways that a participant's data might be shared, goes beyond what GWAS guidelines currently recommend IRBs consider when reviewing informed consent[1], and thereby goes further beyond the requirements of the Common Rule, which still permits use of de-identified samples without consent.

We are fully in favor of efforts to effectively increase a participant's understanding of the potential benefits and risks of contributing their specimens to broad data sharing. We request that new policies which seek to do so strive for consistency and clarity. Researchers and institutions will be best able to operationalize compliance with explicit, practical requirements.

4. Define expectation for deposition of phenotype/clinical data.

The Draft Policy articulates a requirement that phenotype or clinical data be submitted, but fails to define how much associated phenotype/clinical data should accompany the genomic data.

The intent may be to allow investigators and their institutions to determine what they wish to share broadly. However, if NIH officials have a pre-set notion regarding what they expect, the public should have an opportunity to review and comment on these expectations.

5. Establish consequences and penalties for those who intentionally misuse data.

Both the 2007 data sharing policy for data generated in GWA studies and the pending Draft Policy fail to articulate clear consequences and penalties for any investigator who intentionally breaches the privacy and confidentiality interests of the individual research participants whose data are being deposited and broadly shared. Since several publications have demonstrated that it is possible to identify individuals in databases of "de-identified data," it is essential that NIH make clear how it will sanction intentional abuse or misuse of the genomic and associated phenotype data covered by the Draft Policy.

6. Allow data from existing biorepositories, responsibly collected under existing consent requirements, to be deposited.

It is not clear whether data generated from biospecimens collected before the effective date of the Policy, according to an approved consent process not outlined by the Policy, will be acceptable for deposit into dbGaP or other NIH data repositories.

The Section C.4 of Draft Policy states:

> "in these cases, an assessment by an IRB, Privacy Board, or equivalent group is essential to ensure that data submission is not inconsistent with the informed consent provided by the research participant."

The phrase "not inconsistent with the informed consent provided" is open for interpretation.

IRBs, and other groups that are asked to make this determination, may take a highly conservative view and opt not to allow these data to be deposited. This may mean that data from a number of existing specimen repositories would be determined to be ineligible to be deposited into dbGaP and other NIH data repositories. In addition, the Policy suggests that existing biorepositories will need to modify their current consent documents to comply with the new Policy, raising the question as to whether institutions' with such biorepositories should proceed with a large scale re-consent process. This would require a

[1] "Does the consent form either allow or preclude…..submission of the participant's coded phenotype and genotype data to a government health research database for broad sharing to qualified investigators?" http://gds.nih.gov/pdf/PTC_for_IRBs_and_Institutions_revised5-31-11.pdf , at p.14.

huge investment in resources, but it may be the stance that some IRBs take based on the passage in the Draft Policy cited above.

7. **Include language that encourages sharing of aggregate findings**.

Finally, we encourage NIH to consider language about the value of researchers and institutions to share aggregate findings from studies with research participants.

We appreciate this opportunity to offer these suggestions from members of the research ethics community working in CTSA programs across the country.

Very truly yours,

Co-Chairs Biobank Working Group
CTSA Clinical Research Ethics Committee

John Lynch, Ph.D.
Associate Professor and Clinical Ethicist
University of Cincinnati's Center for Clinical and Translational Science and Training
University of Cincinnati. Cincinnati, Ohio
john.lynch@uc.edu

Karen L. Illuzzi Gallinari, J.D.
Director of Regulatory Affairs for Medical Research, Montefiore Medical Center
Einstein-Montefiore CTSA, Bronx, New York
kgallina@montefiore.org

Jennifer B McCormick, Ph.D., M.P.P.
Assistant Professor of Biomedical Ethics, College of Medicine
Divisions of General Internal Medicine and Health Care Policy & Research
Biomedical Ethics Program
Mayo Clinic**,** Rochester, Minnesota
mccormick.jb@mayo.edu

Alexander M. Capron, LLB
University Professor
Scott H. Bice Chair in Healthcare Law, Policy and Ethics
Director, Research Ethics, Southern California Clinical and Translational Science Institute
University of Southern California, Los Angeles, California
acapron@law.usc.edu

James M. DuBois, DSc, PhD
Professor of Medicine
Director, Center for Clinical Research Ethics
Washington University School of Medicine
St Louis, Missouri
jdubois@wustl.edu

Kelly Edwards, PhD
Senior Faculty, Institute for Translational Health Sciences (CTSA)
Associate Professor, Department of Bioethics and Humanities
University of Washington School of Medicine, Seattle, Washington
edwards@uw.edu

Kenneth W. Goodman, Ph.D., FACMI
Professor and Director, Bioethics Program
University of Miami, Miami, Florida
KGoodman@med.miami.edu

Elizabeth Heitman, Ph.D.
Chair, CTSA Clinical Research Ethics Key Function Committee
Associate Professor
Center for Biomedical Ethics and Society
Vanderbilt University Medical Center, Nashville, Tennessee
elizabeth.heitman@Vanderbilt.Edu

D. Micah Hester, Ph.D.
Chief and Professor, Division of Medical Humanities
University of Arkansas College of Medicine
Professor of Medical Humanities and Pediatrics, University of Arkansas for Medical Sciences
Clinical Ethicist
Arkansas Children's Hospital, Little Rock, Arkansas
DMHester@uams.edu

Crispin Noelle Goytia
Program Manager
Centers for Community and Academic Research Partnerships
Mount Sinai Institutes for Translational Sciences, NYC, New York
Crispin.Goytia@mountsinai.org

Robert M. Sade, M.D.
Distinguished University Professor (Surgery)
Director, Institute of Human Values in Health Care
Director, SCTR (CTSA) Clinical Research Ethics Core
Medical University of South Carolina, Charleston, South Carolina
sader@musc.edu

Nicholas H. Steneck, Ph.D.
Director, Research Ethics Program
University of Michigan Institute for Clinical and Health Research
Professor Emeritus of History, University of Michigan,
Ann Arbor, Michigan
nsteneck@umich.edu

Benjamin Wilfond, M.D.
Director | Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute
Professor | Department of Pediatrics, University of Washington School of Medicine
Director | Research Ethics Consultation Service, Institute of Translational Heath Sciences
Seattle, Washington
benjamin.wilfond@seattlechildrens.org

Mark Yarborough, Ph.D.
Dean's Professor of Bioethics
Bioethics Program
University of California Davis Medical School, Sacramento California
mark.yarborough@ucdmc.ucdavis.edu

AMERICAN UNIVERSITY

WASHINGTON, DC

Jorge L. Contreras
202-274-4424
contreras@wcl.american.edu

November 19, 2013

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

Re: Draft NIH Genomic Data Sharing Policy Request for Public Comments (Fed. Reg. Vol. 78, No. 183, p. 57,860, Sept. 20, 2013)

I appreciate the opportunity to share comments regarding the National Institutes of Health (NIH) Draft Genomic Data Sharing Policy ("GDS Policy"). I am a professor of law at American University Washington College of Law, prior to which I spent seventeen years as a practicing attorney representing major research institutions, R&D consortia and private enterprises. I have served as a member of NIH's National Advisory Council on Human Genome Research and currently serve as a member of the Advisory Board of the National Center for Advancing Translational Science (NCATS), the Cures Acceleration Network Board, and the National Conference of Lawyers and Scientists, of which I am co-chair. My current research focuses on the production and dissemination of scientific and technical information, and I have written extensively about data sharing in the context of genomic research. These comments represent my own views, and not those of American University Washington College of Law, or any of the other organizations mentioned above.

I commend the NIH on preparing a thoughtful draft policy on a topic that is of great importance to today's scientific research community. The following comments are offered in the spirit of supplementing this positive first step.

1. Legacy and Recognition of the Bermuda and Ft. Lauderdale Principles [Background]

As a preliminary matter, I find it surprising that the Draft GDS Policy makes no mention of the hugely important "Bermuda Principles"[1] that formed the basis for NIH's genomics data sharing policy in 1996.[2] The Bermuda Principles established, among other things, that human DNA sequence assemblies greater than one kilobase (Kb) in length should be released to the public *within twenty-four hours*, and that finished annotated sequences should be submitted

[1] Summary of Principles Agreed Upon at the First International Strategy Meeting on Human Genome Sequencing (Bermuda, 25–28 February 1996), http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1.
[2] NATIONAL HUMAN GENOME RESEARCH INSTITUTE, NHGRI POLICY REGARDING INTELLECTUAL PROPERTY OF HUMAN GENOMIC SEQUENCE (April 9, 1996), http://www.genome.gov/10000926.

immediately to a public database. Likewise, the Draft GDS Policy fails to mention subsequent data sharing policies adopted by NHGRI in 1997,[3] 1999[4] and 2000.[5]

Instead, the "Background" section of the Draft GDS Policy implies that NIH's genomic data sharing policy began with NIH's February 2003 policy relating to the sharing of research data.[6] While this NIH-wide policy was a significant milestone, the Draft GDS Policy also omits mention of NHGRI's concurrent 2003 data sharing policy that arose from a key stakeholder meeting in Ft. Lauderdale,[7] and which reaffirmed the importance of the Bermuda Principles to Federal data sharing policy.[8] Both the Bermuda Principles and Ft. Lauderdale Principles have been widely recognized by researchers, commentators and NIH officials as cornerstones of U.S. genomics data sharing policy.[9]

Accordingly, omitting any mention of the well-known Bermuda Principles and Ft. Lauderdale Principles, and the NIH/NHGRI policies that embodied them, in a document establishing new Federal policy in the area of genomic data release could be viewed as an attempt by NIH to distance itself, and its current policies, from these earlier statements. While it is certainly NIH's prerogative to alter its policies in accordance with current needs, institutional goals and administration priorities, such changes, and shifts away from existing policy priorities, should be disclosed to, and discussed with, the public.

This point is of more than academic interest. As I observe below, several of the proposed provisions of the Draft GDS Policy move away from the standard of rapid, pre-publication data release that have been in effect for the past two decades, but without any explanation of why. Discarding these earlier policies might be justified by current conditions. But in order to enable an informed public assessment of, and debate over, the Draft GDS Policy, it is important that

---

[3] National Human Genome Research Institute, Current NHGRI Policy for Release and Database Deposition of Sequence Data (Mar. 7, 1997), http://www.genome.gov/page.cfm?pageID=10000910.
[4] National Human Genome Research Institute, Policy on Availability of Genomic DNA Sequence Funded by NHGRI (Jul. 1, 1999),
[5] National Human Genome Research Institute, NHGRI Policy for Release and Database Deposition of Sequence Data (Dec. 21, 2000), www.genome.gov/page.cfm?pageID=10000910.
[6] Final NIH Statement on Sharing Research Data (Feb. 26, 2003),
[7] Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility (Ft. Lauderdale, Jan. 14-15, 2003), http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf.
[8] NAT'L HUMAN GENOME RESEARCH INST., Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-Scale Sequencing and Other Community Resource Projects (Feb. 2003), http://www.genome.gov/10506537.
[9] *See, e.g.* Barbara R. Jasny, *Realities of Data Sharing Using the Genome Wars as Case Study – An Historical Perspective and Commentary*, 2 EJP Data Science (2013); Jorge L. Contreras, *Bermuda's Legacy: Patents, Policy and the Design of the Genome Commons*, 12 MINN. J.L. SCI. & TECH. 61 (2011), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1667659; Jorge L. Contreras, *Prepublication Data Release, Latency, and Genome Commons*, 329 SCIENCE 393 (2010); Toronto Int'l Data Release Workshop Authors, *Pre-Publication Data Sharing*, 461 NATURE 168, 169–70 (2009); Jane Kaye, et al., *Data Sharing in Genomics – Re-shaping Scientific Practice*, 10 NATURE REV. GENETICS 331, 332 box 1 (2009); Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S.Guyer, *A Vision for the Future of Genomics Research: A Blueprint for the Genomic Era*, 422 NATURE 835 (2003).

NIH disclose and openly discuss the policy considerations that motivated any meaningful shift away from its prior stance.

2. *Timing of Data Release* [Appendix A – Expectations for Data Submission and Data Release]

The most significant change that would be introduced by the Draft GDR Policy is a lengthening of the period between the generation and public release of human DNA sequence data. Under previous NIH/NHGRI genomic data release policies (1996, 1997, 1999, 2000 and 2003), genomic data was required to be released in publicly-accessible databases with 24 hours, or similarly short time periods. This rapid release of data, introduced by the Bermuda Principles, was once a cornerstone of NIH's approach to genomic research. As stated by the leaders of the Human Genome Project in 2001, "We believed that scientific progress would be most rapidly advanced by immediate and free availability of the human genome sequence. The explosion of scientific work based on the publicly available sequence data in both academia and industry has confirmed this judgment."[10] More recently, the demonstrable benefits of rapid data release have been studied and documented by scholars such as Heidi Williams.[11]

Under more recent "second generation" data release policies issued by NIH/NHGRI (GWAS, GAIN, ENCODE, TCGA),[12] rapid release of data to the public was still required, but users of this public data were generally obliged to refrain from publishing or presenting analyses of that data during "embargo" periods of 6-12 months. In all of these cases, data was released almost immediately after generation or validation, so that it would be of the greatest use to the broader scientific community and so that scientific discovery and medical advancement would proceed as rapidly as possible.

In contrast, the Draft GDR Policy would introduce substantial time delays to the public release of genomic data. As described in Appendix A, DNA sequencing reads would not be publicly released at all. DNA data that has undergone an initial round of analysis and "cleaning" would be submitted to an NIH repository within three months after completion of this analysis, and then released to the public up to six months later. In total, there would be a theoretical minimum of nine additional months before data is released to the public, but in all likelihood a period closer to one year or longer (given the time necessary to conduct initial analysis and data cleaning).

---

[10] Int'l Hum. Genome Sequencing Consortium, *Initial Sequencing and Analysis of the Human Genome*, 409 Nature 860, 864 (2001). *See also* Collins, et al., *supra* note 9, at 846 ("Scientific progress and public benefit will be maximized by early, open and continuing access to large data sets."); Robert Mullan Cook-Deegan & Stephen J. McCormack, *A Brief Summary of Some Policies to Encourage Open Access to DNA Sequence Data*, 293 Science 217 (2001) ("[W]ithout [the Bermuda Principles] the wait for information sufficient to meet patent criteria from high throughput sequencing programs would lead to long delays, and thus be a serious drag on science, undermining the publicly funded sequencing programs' very purpose").

[11] Heidi L. Williams, *Intellectual Property Rights and Innovation: Evidence from the Human Genome*, 121 J. Political Econ. (2013).

[12] *See* Contreras (2011), *supra* note 9, at 97-104 and Table 1 (summarizing these policies and their data release requirements).

The introduction of this time lag, which I have previously termed "knowledge latency",[13] is not necessarily a bad thing. It may be necessary to ensure that higher-quality data enters the public sphere and to give data-generating researchers a "head start" in analyzing the data that they generated. These considerations have been discussed at length in the literature, beginning as early as the Bermuda Principles themselves. Private sector genomic research groups have also adopted time-delayed data release programs, and have contributed substantial data to the public domain in so doing.[14] However, what is lacking in the Draft GDR Policy is any reasoned explanation for the proposed lengthening of the data release period, and any acknowledgement that a significant change is being made. It is incumbent on NIH to offer the public its rationale for the introduction of this new latency period, and to justify the inevitable and corresponding reduction in scientific output, knowledge dissemination and biomedical discovery.

3. *Penalties for Non-Compliance* [II. Scope and Applicability]

In order for the data release and other provisions of the Draft GDR Policy to be effective, there must be a realistic expectation that they will be enforced, as well as meaningful penalties for non-compliance. Past NIH/NHGRI data release policies have been roundly criticized for their lack of penalties and enforcement mechanisms.[15] This lack of enforcement is believed to have resulted in widespread noncompliance with rules such as embargo restrictions on the use of publicly-available data. The most publicized incident of this nature occurred in 2009 with respect to an article appearing in the *Proceedings of the National Academies of Science* (*PNAS*).[16] Interestingly, it was the journal, not NIH, that took action against the alleged violator.

In an apparent attempt to address this criticism, Section II of the Draft GDR Policy outlines an enforcement framework for noncompliance. It states that "Compliance with this Policy will become a special term and condition in the Notice of Award or the Contract Award. Failure to comply with the terms and conditions of the funding agreement could lead to enforcement actions, including the withholding of funding, consistent with 45 CFR 74.62 and/or other authorities, as appropriate."

While this language is a good start, it may not go far enough to deter noncompliance with the policy. In particular, the reference to "withholding of funding" implies that future funds may not be disbursed under a grant. In many cases, this penalty will be no deterrent at all, as projects

---

[13] *See* Jorge L. Contreras, *Confronting the Crisis in Scientific Publishing: Latency, Licensing and Access*, 53 Santa Clara L. Rev. 491 (2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2015885; Contreras (2010), *supra* note 9, and Jorge L. Contreras, *Data Sharing, Latency Variables and Science Commons*, 25 Berkeley Tech. L.J. 1601 (2010), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1501280.

[14] *See, e.g.*, Jorge L. Contreras, Aris Floratos & Arthur Holden, *The International Serious Adverse Events Consortium's Data Sharing Model*, 31 NATURE BIOTECH. 17 (2013).

[15] *See, e.g.*, Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 Law & Contemp. Probs. 289, 293-94 (2003) (offering a critique of NIH's "hortatory" approach to issues of patenting and biomedical research).

[16] Randy Schekman, *Editorial – PNAS Takes Action Regarding Breach of NIH Embargo Policy on a PNAS Paper*, 106 Proc. Natl. Acad. Sci. 16893 (2009).

may already be completed at the time that data release is required or noncompliance is detected. While the Draft GDR Policy also makes reference to other potential penalties under 45 CFR 74.62, these are not specified and most researchers are probably unaware of the range of penalties available under that regulation. Thus, I would recommend making explicit the range of penalties that may be incurred as a result of noncompliance with the GDR Policy. Under 45 CFR 74.62, such penalties could include withholding *further* awards (74.62(a)(4)), and debarment and suspension of the entire institution from receipt of Federal funding (74.62(d)). It is important to emphasize to NIH-funded researchers that the consequences of noncompliance with the GDR Policy may be substantial, if not catastrophic, not only to their individual research programs, but to their institutions, as the prospect of such dire consequences could significantly improve compliance with policy requirements.

4. *Intellectual Property* [Section VI]

As in previous NIH/NHGRI data release policies, the provisions relating to intellectual property have little, if any, legal effect. I commend NIH for recognizing the landmark Supreme Court decision in *Association for Molecular Pathology v. Myriad Genetics*[17] and its holding that naturally-occurring DNA is not patentable in the U.S. However, this is merely a restatement of existing law. The following statement that basic DNA sequences and other forms of data are "pre-competitive" is not compelled by the Court's holding, and is not discussed in the case. In fact, the term "pre-competitive" is a poorly-defined term that has little legal meaning. This being said, the Draft GDR Policy makes a valuable contribution by stating that such data should be made available without restriction. The contribution would be greater, however, if NIH more clearly stated that such data "must" be made publicly available without licensing restrictions (especially royalties), as the term "should" is notoriously ambiguous and could be interpreted to mean that such an approach is recommended, but not required (i.e., there is no penalty for noncompliance).

Even less definitive are further statements in Section VI of the Draft GDR Policy regarding the use of patents to block access to data developed with NIH support. NIH "discourages" such blocking tactics, and "encourages" uses of data that are consistent with its 2005 "best practices" for licensing genomic inventions and research tools.[18] While these sentiments are welcome, it has long been recognized that policies stated in terms of encouragement rather than requirement have little to no real impact on behavior. This seems particularly true with regard to NIH's licensing "best practices", which have been in place for nearly a decade, but which have never, to my knowledge, been asserted against any entity that has failed to abide by them.[19] In fact, it does not appear that any mechanism for policing or monitoring compliance with such policies even exists. Thus, while the sentiments expressed in Section VI of the Draft GDR Policy are admirable and welcome, they are not likely to have a

[17] 569 U.S. __ (2013)
[18] Natl. Inst. Health, Best Practices for the Licensing of Genomic Inventions, 70 Fed. Reg. 18413 (Apr. 11, 2005), http://www.gpo.gov/fdsys/pkg/FR-2005-04-11/pdf/05-7247.pdf.
[19] *See* Rai & Eisenberg, *supra* note 15.

meaningful effect on researcher behavior unless they are made more definitive and binding and become subject to real penalties for noncompliance (see Paragraph 3 above).[20]

Thank you again for the opportunity to offer these comments in response to your inquiries. Please do not hesitate to let me know if there is any additional information that I can provide in support of these important matters.

Respectfully yours,

Jorge L. Contreras

[20] I have been told informally that NIH is loathe to strengthen the patent deterrent measures contained in its policies for fear of violating the Federal Bayh-Dole Act of 1980, 35 U.S.C. §§ 200-12 (2006). This fear does not appear to be warranted, and a more complete discussion of NIH's legal position in this regard would be useful to the public consideration and debate of the Draft GDR Policy.

COGR
an organization of research universities

COUNCIL ON GOVERNMENTAL RELATIONS
1200 New York Avenue, N.W., Suite 750, Washington, D.C. 20005
(202) 289-6655/(202) 289-6698 (FAX)

BOARD OF DIRECTORS
SUSAN CAMBER, Chair
University of Washington

JAMES BARBRET
Wayne State University

SARA BIBLE
Stanford University

LOIS BRAKO
University of Michigan

PAMELA CAUDILL
Harvard University

W. MARK CROWELL
University of Virginia

KELVIN DROEGEMEIER
University of Oklahoma

CYNTHIA HOPE
University of Alabama

MICHAEL LUDWIG
Purdue University

JAMES LUTHER
Duke University

JAMES R. MAPLES
University of Tennessee

DENISE MC CARTNEY
Washington University in St. Louis

ALEXANDRA MC KEOWN
The Johns Hopkins University

KIM MORELAND
University of Wisconsin

CORDELL OVERBY
University of Delaware

PATRICK SCHLESINGER
University of California, Berkeley

SUSAN SEDWICK
University of Texas, Austin

JAMES TRACY
University of Kentucky

PAMELA WEBB
University of Minnesota

DAVID WINWOOD
University of Alabama
at Birmingham

KEVIN WOZNIAK
Georgia Institute of Technology

ANTHONY DE CRAPPEO
President

November 19, 2013 By email to: GDS@mail.nih.gov

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

SUBJECT: Draft NIH Genomic Data Sharing Policy

The Council on Governmental Relations (COGR) is an association of 190 research universities and their affiliated academic medical centers and research institutes. COGR concerns itself with the influence of federal regulations, policies, and practices on the performance of research conducted at its member institutions. We appreciate the opportunity to offer comment on the Draft National Institutes of Health (NIH) Genomic Data Sharing Policy.

We recognize and appreciate that the approach taken to design this policy is consistent with the Genome Wide Association Studies (GWAS) policy implemented in 2007. The use of the GWAS model raises the question of whether the policies and procedures for GWAS and the associated database (dbGaP) will be subsumed under this draft policy. While the policies are similar, they are not identical notably in some of the assurances in the Institutional Certification. If the policies can be aligned, the institutional processes and procedures can be streamlined.

In a similar effort to decrease the burden on investigators and the grantee institutions and streamline processes, we urge NIH to consider a "just-in-time" approach to the policy. The implementation of this policy will increase the responsibilities of the Institutional Review Board both for the retrospective reviews and the likely increased volume of studies that require a data sharing review as well. Investigators will be required to design detailed data sharing plans for proposals that have a declining likelihood for success. To mitigate some of the burden, NIH could require the complete plan and Institutional Certification as a part of the "just-in-time" process for proposals recommended for funding. This process has been successful with other NIH requirements and should be considered for implementation with this policy as well.

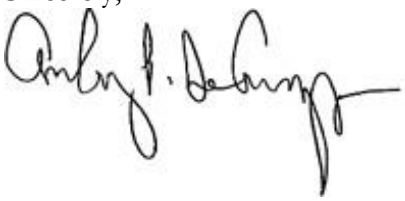
Some questions have been raised concerning the submission of data prior to submission for publication. We appreciate that there can be some lag time between data generation and the organization of data for publication but some investigators expressed concern with submission within three months particularly for Level 2 data.

The need for a retrospective review of human genomic data collected prior to the effective date of the policy is clear. Clarification concerning the use of non-human genomic data and the need for retrospective submission of all data classified as large-scale and collected before the effective date of the policy would help

investigators in the preparation of applications. In a similar manner, Appendix A offers some examples of research covered by the policy but the addition of definitions or explanations will help investigators in meeting the data sharing requirements.

We appreciate the opportunity to offer these comments. We recommend the more detailed comments prepared by the Federation of American Societies for Experimental Biology (FASEB) in response to this Request for Information.

Sincerely,



Anthony P. DeCrappeo
President

Commenter: Dana-Farber Cancer Institute, Michele Russell-Einhorn

Date of comment: 11/20/2013 15:46

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

The Dana-Farber Cancer Institute (DFCI) is the IRB of record for the institutions conducting cancer research involving human subjects under the umbrella of the Dana-Farber/Harvard Cancer Center. These institutions include: DFCI, Brigham and Womens' Hospital, The Childrens' Hospital Boston, Beth Israel Deaconess Medical Center and the Massachusetts General Hospital.We fully support the goals and mission of the NIH in "promoting the sharing of genomic research data, which maximizes the knowledge gained." Our member institutions have already been significant contributors to the NIH databases covered by this data sharing policy and hope to expand their participation going forward. We have reviewed the draft GDS policy. DFCI would request that NIH define the term "compelling scientific reasons" in Section II (C) 4. It would be helpful to investigators, institutions and IRBs to know when a potential submission meets this criterion. Specifically, it would be helpful to know:1. What is considered compelling2. Who/what committee would make that decisionDFCI also requests that IRBs should be able to continue to evaluate secondary research projects using de-identified samples destined to be submitted to a NIH database under the same regulatory framework as other types of research that are determined to be minimal risk and for which a waiver of consent is deemed appropriate. In discussion with our investigators, collaborating institutions, and our IRB chairs, the consensus we have reached and would assert as meeting the goals and spirit of this policy, is that secondary research with somatic sequencing data presents a minimal risk of re-identification and/or a privacy breach. DFCI has already taken steps to include appropriate language in the consent forms that involve research with tissue samples, even if the submission of data to public databases has not yet been determined. The requirement for explicit consent has the full support of the DFCI community and that commitment predates this policy. However, DFCI can foresee situations where valuable secondary research is proposed that involves samples obtained after the date of this policy from patients who are deceased or were not consented because collection of specimens was originally undertaken for a non-research purpose. We therefore request that the

policy specific state that an IRB may use the tool of a waiver of informed consent so long as the appropriate regulatory criteria have been met.

Commenter: Priya Duggal

Date of comment: 11/19/2013 13:09

Comment:

Section II. Scope and Applicability: Data sharing is critical and important for the advancement of all science. The scope of this data sharing plan may be harder to implement and enforce since it includes sequencing of one or more genes in more than 100 participants. Will this type of data sharing be as fruitful or manageable as compared to release of sequence array or exome or whole-genome data?

Section IV.A. Data Sharing Plans: The proposed 6 month plan for release of data will likely disadvantage smaller groups and individual scientists who will not have the means to complete QC on large datasets and then perform analysis and submit for publication. Since QC pipelines and data analysis are still being established for sequence data it is especially important that investigators have the time necessary to properly analyze and interpret their data before contributing to the literature. I suggest extending this date to 12 months. After 12 months data can be released for data sharing without an additional embargo period.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: The current proposal accurately represents what likely will happen if a participant decides to "opt out" of a study. Specifically, it will be difficult to retrieve data that has already been sent out and/or published.

However, since the proposal requires a yearly renewal for investigators, it should also allow for samples to be removed from further analysis at that yearly renewal. This will insure that participants truly have a voice to "opt out". Of course, for new submissions, the specific participants should be removed prior to data release.

As an epidemiologist, there is considerable concern that all future studies will require the broad sharing of data and no individuals without this consent will be considered for genetic research. How will this affect participants who are already consented and enrolled and how will this impact or bias participation ?Clear guidelines for marginalized or special populations, families, foreign nationals, etc that may need and want to restrict data due to specific circumstances or concern that they will be identified should be evaluated. These are still important populations that may have large contributions and benefit to genetic research. Ideally, these consent issues should be considered on a global NIH scale and not just per individual institute.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Eric Engels

Date of comment: 11/19/2013 12:08

Comment:

Section II. Scope and Applicability: It is unclear why it is necessary for all researchers to upload all genomic data from all samples in all studies, including sequences for genes, exosomes, and microbes. It seems like the greatest value comes from the publication of the research results themselves, in peer reviewed journals, which is already happening. If other researchers would like access to some or part of these data, then they already can contact the researchers who produced them.

Section IV.A. Data Sharing Plans: It is a huge effort for researchers to determine how the data they have can be formatted and uploaded to a central site. Also, given the enormous amount of such data, it is entirely unclear how the infrastructure can be set up to store all of it, and how the system would be funded. This appears to be an unfunded mandate that will unfairly burden NIH researchers.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: I worry about the loss of confidentiality that uploading this large amount of genomic data will expose research subjects to. There are already reports of investigators being able to deduce the identities of subjects from genome sequence results coupled with publicly available information.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: The policy as written appears too broad and will create a substantial burden to researchers. I believe the costs to researchers and the risks to subjects (in terms of loss of privacy) outweigh any potential benefits from such a broad demand for data sharing.

WHITEHEAD INSTITUTE

Yaniv Erlich
Fellow/Principal Investigator
Whitehead Institute for Biomedical Research

Nine Cambridge Center
Cambridge, Massachusetts 02142-1479
617.324.6256 / 617.258.5578
yaniv@wi.mit.edu

November 19, 2013

To NIH GDS Committee

Subject: Comments on the Genomic Data Sharing Policy (GDS Policy) Draft

My background is computational human genetics, I am a member of MIT IRB committee, and my group at the Whitehead Institute has been working on several issues regarding genetic privacy, secondary usage of genomic data, and mining Web 2.0 information for human genetics.

I would like to suggest the following comments for the GDS Policy:

1. **Open Data Sharing Plans**

The new GDS proposal states: *"Investigators and their institutions are expected to address plans for following this Policy in the data sharing section of funding applications and proposals…[T]he NIH expects the informed consent process and documents to state that a participant's genomic and phenotypic data may be shared broadly for future research purposes and also explain whether the data will be shared through open or controlled access.*"

My suggestion is that **the data-sharing plan sections** of awards will be publicly accessible in RePOTER (or an equivalent system) once the award is granted. Such transparency has several advantages: first, it will allow broader oversight about the adherence of the awardees with their data sharing plans. Second, this will show the level of commitment of the NIH and its grant recipients for data sharing. Third, making the data sharing plans available will enable empirical ELSI research regarding data sharing trends. Fourth, data sharing plans do not contain any sensitive or proprietary information. To the best of my knowledge based on the NIH FOIA Office guidelines, data sharing plans are accessible via FOIA requests. Therefore, there is no particular reason for not publicly posting them in advance.

2. **Data Submission Expectations are not Clear**

In section C1, Data Submission Expectations and Timeline, the new GDS proposal says: "*Human data that are submitted to NIH-designated data repositories should be de-identified according to … the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*"

Specifically, when HIPAA Privacy Rule, the GDS proposal refers to 45 CFR 164.514(b)(2), the HIPAA Safe Harbor section.

WHITEHEAD INSTITUTE

Yaniv Erlich
Fellow/Principal Investigator
Whitehead Institute for Biomedical Research

Nine Cambridge Center
Cambridge, Massachusetts 02142-1479
617.324.6256 / 617.258.5578
yaniv@wi.mit.edu

The current wording is confusing. The HIPAA Safe Harbor lays out two possible tactics to de-identify data. The first tactic requires the removal of 18 different identifiers, including "*Biometric identifiers*" (Identifier #16) and "*Any other unique identifying number, characteristic, or code*" (Identifier #18). The second tactic requires that "*[t]he covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information."*

Multiple lines of studies from our lab and other groups have shown that genomic and other types of omics data confer sufficient information to identify individuals in various scenarios (Lin et al, *Science*, 2004; Homer et al., *PLoS Genetics*, 2008; Shadt et al., *Nature Genetics*, 2012; Gymrek et al., *Science*, 2013. For a detailed review: see our preprint in arXiv 1310.3197). Omics datasets inherently contain the biometric identifiers and unique characteristics that are described by HIPAA Safe Harbor. In addition, there is actual knowledge that this information could be used to identify individuals. Therefore, it is **impossible** to de-identify omics datasets according to the HIPAA Safe Harbor without destroying the actual data. In other words, the GDS proposal asks investigators to submit empty datasets.

I am sure that this was not the intention but HIPAA Safe Harbor tactics are not an adequate mechanism. As an alternative, I suggest that you will specify a subset of identifiers from HIPAA Safe Harbor that should not be included.

Another issue is that the GDS proposal states that "*Human data that are submitted to NIH-designated data repositories **should** be de-identified*". While still a minority, certain study participants are not interested in keeping their data private and might decide to allow sharing with explicit identifiers to facilitate future research (Goolsby et al., *IOM Roundtable*, 2010). For example, genetic studies of facial dysmorphologies sometimes ask the permission of study participants to share their photos. In other cases, such as the PGP-10, the participants decide to publicly identify themselves. It is impossible to de-identify these datasets. The current GDS proposal seems to prevent the sharing of these studies in NIH-repositories.

My suggestion is to revise this sentence to something similar to "the NIH-designated repositories accept studies with and without explicit identifiers (such as name, photos, or contact information) of participants. The final decision whether to release these explicit identifiers should be addressed in the informed consent*".*

3. **Including unmapped reads to BAM files**

WHITEHEAD INSTITUTE

Yaniv Erlich
Fellow/Principal Investigator
Whitehead Institute for Biomedical Research

Nine Cambridge Center
Cambridge, Massachusetts 02142-1479
617.324.6256 / 617.258.5578
yaniv@wi.mit.edu

Appendix A of the GDS proposal sets five levels of data based on the amount of processing. Level 2 is the rawest format that is expected to be commonly shared for human studies "[It] *would be a file (e.g., binary alignment matrix (BAM) files)* ***usually containing the unmapped reads*** *as well.*"

It is highly advisable **to include the unmapped reads as a standard** in these BAM files to maximize consistency, serendipity, and secondary data usage.

Alignment programs considerably vary in their performance, algorithms to deal with repetitive elements, and gapped alignment methods (Treangen and Salzberg, *Nature Reviews Genetics*, 2011). These variations can translate to certain error patterns in downstream algorithms. Data fusion from multiple studies with different alignment techniques might be error prone and introduce systematic biases that can create false positive. Including the unaligned reads as a default will enable to recover the original FASTQ files and neutralize most the systematic biases.

In addition, a wide range of software including from our group and others developed specialized algorithms to call "exotic" types of variations and biological events in DNA and RNA data (Hormozdiari et al., *Bioinformatics*, 2010; Gymrek et al., *Genome Research*, 2012; Highnam et al., *Nucleic Acid Research*, 2012; Dobin et al., *Bioinformatics,* 2013). These algorithms usually require access to the unmapped reads as they employ specialized alignment strategies. Including the unaligned reads as default will enable this type of research.

Summary
I hope that you will find my comments useful. Please do not hesitate to contact me with any questions by phone (617-913-1318) or via email (Yaniv@wi.mit.edu).

Commenter: Peggy Farnham

Date of comment: 11/14/2013 14:13

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: There are some points in the document that are not clear. 1) According to the document, ChIP-seq data becomes level 2 as soon as it is aligned to the human genome. To me, this suggests that all ChIP-seq data, whether it is good or bad, one replicate or two, has to be submitted to a public repository. This makes no sense. A ChIP-seq sequenced sample could be of good "quality" in the sense of sequencing metrics, but have few peaks or high background (due to antibody or other experimental issues) and yet according to the standards this data should be submitted. ChIP-seq is not the same as simply sequencing the genome of a particular cell type, where there is a correct answer. In the real world of ChIP-seq, experiments sometimes simply don't work. 2) What about people working out new technologies, new library methods, etc. Do even the failed genomic experiments go into a repository? If a sample goes onto a sequencing machine, is there an obligation to deposit the results no matter what the outcome or utility of the data produced? 3) There seems to be no mention of reproducibility anywhere in the document. For example, for a ChIP-seq, will the requirement be that a single replicate must be submitted? How would anyone know if it was good data without a second ChIP-seq replicate of the same factor in the same cell type? This new policy may result in a lot of bad analyses by future investigators.4) The document does not make clear what constitutes a genomic experiment. If a sample needs more reads, would the investigator submit after the second lane or after the first lane, or both? What about a Miseq run? If a small number of reads are obtained for a ChIP-seq sample to see if it worked, does this have to be submitted or does the investigator wait for the Hiseq run? 5) What kind of information will the investigator be required to give? For a ChIP-seq, would they have to document the antibody used and lot number? For RNA-seq, do they have to provide cell growth conditions, treatments, etc? Without this metadata, is the genomic information usable? If it isn't usable, why collect it?6) Finally, in my opinion we should stick with the current policies:a) For individual investigators, have the requirement that all genomic data used in a publication be deposited prior to acceptance of the manuscript.b) For large consortia, have the requirement that all genomic data be deposited immediately after VERIFICATION of the quality of the data using standards developed by the consortium for each specific data type. Verification could, for example, require the comparison of replicate datasets, a determination of whether the appropriate sequencing depth has been achieved, etc.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: W. Andrew Faucett & David H. Ledbetter

Date of comment: 11/20/2013 17:57

Comment:

Section II. Scope and Applicability: The policy should clearly indicate that it only applies to data generated as part of research. It should not apply to data collected by clinical labs, even if the collection and deposition in databases of clinical data is funded as part of research. Special provisions should be made for data collected before this policy is enacted to allow that data to be shared and palced in NIH databases.

Section IV.A. Data Sharing Plans: Most inidivduals particpate in research due to a motivation to help others and/or increase knowledge about a particular disease that they have or that runs in their family. In respect of their original motivations to participate in research, broad sharing of data should be required of all NIH funded researchers. All new grant proposals should be required to include a data sharing plan and only be funded if that plan allows broad data sharing.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Safeguards should be put in place to protect research participant's confidentiality, while also allowing broad data sharing. Significant presonal and institutional penalties should be put in place for the misuse of Human Genomic Data.Grant reviewers should be required to access the data sharing plan and the plans to allow the return of medically actionable results.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Anyone accessing data from public databases should be required to pledge to protect research participant confidentiality and to NOT attempt to re-identify research participants except when they are returning medically actionable results. Systems should be developed to allow the tracking of all uses of NIH genomic data both controlled-access data and open-access data.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: 1. The policy needs to allow for broad consent policies where those policies are approved by the local IRB and institution. NIH policies should not be more restrictive than the local policy where the data originated.

2. Exceptions to the policy need to be developed to encourage the sharing of data collected before the enactment of this policy, even if the data was not collected with full consent. Notification of participants and/or public posting that data was deposited should be considered.

FASEB
Federation of American Societies for Experimental Biology

Representing Over 110,000 Researchers

301.634.7000
www.faseb.org

9650 Rockville Pike
Bethesda, MD 20814

November 6, 2013

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

Dear NIH Genomic Data Sharing Policy Team:

The Federation of American Societies for Experimental Biology (FASEB) appreciates the opportunity to comment on the National Institutes of Health's (NIH) draft Genomic Data Sharing Policy (NOT-OD-13-119). FASEB is comprised of 27 scientific societies, collectively representing over 110,000 biological and biomedical researchers. FASEB recognizes the importance of promoting the sharing of data resulting from genomic research studies and commends NIH for developing a draft policy to encourage such behavior among NIH-funded investigators.

FASEB agrees with the general principles underlying the draft Policy, as the exchange of research findings serves to increase the efficiency of scientific research and accelerate discoveries that could improve human health. We are, however, concerned about potential unintended impacts; specifically, large increases in administrative burdens for investigators and institutions engaged in genomics research and a decrease in participation of human subjects in clinical research with genomics components due to concerns regarding misuse of shared data. Therefore, FASEB strongly recommends that NIH revise the draft Policy to address the concerns detailed below prior to its implementation.

Increased Administrative Burdens for Investigators and Institutions
While FASEB appreciates the role data sharing plays in advancing scientific discovery, we are concerned that the Policy may introduce unintended administrative burdens at every stage of the research process, from development of a data sharing plan and timeline in the grant proposal to managing the submission of data to databases, including de-identification and development of coding schemes and maintaining this information after the conclusion of the grant award. Below, we highlight some key areas that NIH should address to improve the clarity of the draft Policy.

Lack of Clarity Regarding Types of Research Covered by Policy
FASEB's greatest concern is the lack of clarity regarding the types of research that would be covered by this Policy. Appendix A contains only four examples of the types of research to be covered, which vary greatly in terms of sample size, species of research subject, and overall detail of the volume of sequence data that would trigger coverage by the Policy. The table describing expectations for data submission and data release is similarly vague. While the intent may have been to produce a policy with flexibility to adapt to the rapidly changing field of genomics research, this lack of clarity will likely result in confusion and increase administrative burden as investigators and institutions struggle to determine what constitutes compliance. In response, some institutions may treat all research utilizing genetic methods as the "large-scale" genomic research addressed by this Policy, extending administrative burdens far beyond NIH's intention.

To reduce confusion, FASEB strongly recommends that NIH supplement the final Policy with a workflow diagram or chart to help investigators and institutions navigate and implement the Policy as intended. For example, the supplemental grant application instructions for the PHS 398 and SF 424 forms include scenarios to help guide researchers in determining whether the proposed research does not include human subjects research, includes non-exempt human subjects research, or includes exempt human subjects research. The supplemental instructions also link to the Department of Health and Human Services Office of Human Research Protections where additional information regarding each category can be obtained. A similar set of detailed scenarios for the draft Genomic Data Sharing Policy could serve to enhance implementation and compliance. This would also decrease the risk of over-regulation by institutions, a phenomenon that FASEB's recent survey of federally funded researchers found to be a major source of unnecessary administrative burden.

Option of "Just-In-Time" Data Sharing Plans for Funded Research Proposals
While FASEB agrees that investigators should consider plans for data sharing in the early stages of their research project, we are concerned that the draft Policy requires fairly detailed data sharing plans at the research proposal stage. In fiscal year 2012, the success rate for NIH grant applications was 17.6 percent; NIH Director Francis Collins stated that the success rate for the current year will be 15 percent. The time devoted to the development of a detailed data sharing plan for a proposal with a high likelihood of not being funded is wasteful, and thus we urge NIH to consider adopting a "just-in-time" process that allows investigators to submit basic data sharing plans at the time of proposal submission. Institutional certification and associated documentation should only be required for those proposals recommended for funding. This is commensurate with current policy regarding development and submission of detailed project budgets and has also proven to be successful with human subjects and animal research protocols.

Lack of Clarity Regarding Activities that Constitute Data Sharing
Throughout the draft Policy, "data sharing" is described as deposition of sequence data into "any widely used data repository, whether NIH-funded or not." Per these guidelines and NIH's current policy regarding open-access to publications resulting from federal support, it is unclear whether publication of research results in scientific journals would be accepted as compliance with this draft Policy. To alleviate confusion and potential non-compliance, FASEB urges NIH to clarify the role of publication activity in compliance with the Genomic Data Sharing Policy.

Additional Barriers to Human Subject Research and Research Participation
Genetics and genomics investigators have been leaders within the scientific community in their recognition of the ethical, social, and legal implications (ELSI) associated with their research. FASEB has identified three aspects of the draft Genomic Data Sharing Policy that may be problematic in light of current ELSI research.

Lack of Guidance Regarding Data for which De-identification Cannot Be Guaranteed in Perpetuity
The draft Policy does not address the fact that de-identification **cannot** be guaranteed for certain types of data, including whole genomic sequences. As more data become publically available in both biomedical and non-biomedical databases (such as those for personal ancestry research), and as genotype-phenotype correlations become more predictive, the possibility increases that a genomic sequence could allow for re-identification of research participants. Therefore, NIH should consider alternative models to protect human research subjects. Shifting from a paradigm centered on an institution's responsibility to ensure

data privacy to one that provides research subjects with substantive legal protections against the misuse of or inappropriate access to their data may be a more effective way to minimize risks to research participants.

Exemptions to Data Sharing for Individual Research Participants
The draft Policy has striking implications for the recruitment of patients for clinical trials. Can individuals who, at the time of consent, restrict use of their genomic data to only the original research team still be allowed to enroll in the study? In certain situations, allowing enrollment of these individuals may be critical to the integrity of the research project. The following are a few possible challenges an investigator may face: (1) a higher frequency of non-consent to data sharing among some populations and groups, which would reduce the representativeness of the study population; (2) for rare diseases, loss of only a few individuals from a study, which could be detrimental to achieving a sufficient sample size; and (3) increases in time and costs of study recruitment greatly beyond what their source of funding can provide.

Exemptions to Data Sharing for Select Research Projects
Finally, FASEB is concerned that the draft Policy does not specifically allow investigators to seek exemptions for an entire study when the research is associated with "at risk" or vulnerable populations, including children or specific racial, ethnic, or tribal groups. Some subjects may be more distrustful due to historical abuses and may wish to limit use of their data to the original research team. Also, specific restrictions regarding data sharing may be requested during community consultations, and it is important that researchers are able to respect these requests. Limited technological and biomedical literacy within some populations could also pose a barrier to ensuring informed consent for genetic data sharing. FASEB is concerned this Policy could have an unintended chilling effect on research designed to address health issues or ameliorate health disparities found among these populations.

In conclusion, FASEB commends NIH for its leadership in the development of policies to guide the rapidly changing field of genomics research and appreciates the opportunity to provide comments on the draft Genomic Data Sharing Policy. While we recognize the important role data sharing plays in furthering scientific discovery, if the draft Policy is implemented as currently written, we are concerned it will cause a large increase in administrative burdens for investigators and institutions engaged in genomics research and a decrease in participation of human subjects in clinical research with genomics components. Therefore, we urge NIH to make significant revisions to the areas of the draft Genomic Data Sharing Policy noted above and provide the public with an opportunity to comment on the revised version prior to its finalization and implementation.

Sincerely,

Margaret K. Offermann, MD, PhD
FASEB President

Commenter: Andrew Feinberg

Date of comment: 11/20/2013 17:44

Comment:

I thank the committee of their hard work on this important problem, and I am enthusiastic supporter of the whole idea of data release and open access. That said, I want to raise two issues/concerns with the policy and implementation that follow from my own experience as director of the nation's first Genome Center for Epigenetics. I also apologize for the brevity of these comments but hopefully their parsimony reflects clarity. First, it has now become quite clear to members of the epigenetics community that the generation of epigenetic data and its relationship to phenotype adds additional levels of complexity that are still being worked out, much as it was for GWAS in its early days. These issues include correction for cell type, which often requires additional measurements, and replication, the sample sets for which are not necessarily clear until the primary data are obtained. Note the recent Nature Methods paper by a consortium of statisticians drawing attention to these very two issues. Therefore it will be often or usually necessary to perform additional experiments to obtain sufficient confidence of biological truth. My concern is that if data are publicly available for analysis before these measures are completed, it could lead to misleading interpretations or public perceptions of biological findings from NIH research. I have read the guidelines several times, and discussed them with Program Staff, and I still do not understand what they say regarding this issue. I apologize for that, but I thought it important to register this issue before the deadline is closed. The second issue is that despite the best of intentions, my colleagues and I have found it extremely difficult to access dbGAP data. It is hard to say what the issues are, but it seems to take us something like a half a year, with many confusing rejections, before access is finally granted. I am grateful to the program and its staff, and if I had to guess, the program is woefully underfunded and understaffed. But if new data sets are going to be made available, then these issues need to be addressed as well. Again thank you very much for your kind consideration of these sentiments.
Andrew Feinberg, MD, MPH
Gilman Scholar, Professor of Medicine and Molecular Biology & Genetics
Johns Hopkins University

4301 Connecticut Avenue, NW
Suite 404
Washington, DC 20008
202.966.5557
info@geneticalliance.org
http://www.geneticalliance.org

Dear NIH Genomic Data Sharing Policy Team,

The Genomic Data Sharing Policy provides an excellent opportunity for the NIH to establish methods to engage research participants in a truly participant-centered approach to genomic medicine. The participant consent process provides an opportunity to engage individuals and their families, to explain the intentions of research, and to initiate dialogue about the participant's role in the research process.

First, we believe more emphasis should be put on Fair Information Practice Principles, so that the burden of engagement is not placed upon informed consent alone, and particularly not upon a form, rather than a process.

Second, we believe that the proposal to adopt a "broad consent" approach undermines the NIH's focus on ensuring that participants are appropriately informed about the research to which they are contributing. NIH wishes to engage participants and the public in a much broader understanding of biomedical research, and those who 'raise their hand' to participate in biobanks, registries and clinical trials are prime stakeholders in this engagement.

Instead, we propose that NIH adopts at least a dynamic consent approach, and perhaps a granular (allowing sharing of specific subsets of information) and dynamic consent process. Using dynamic consent will empower participants to understand the potential of the proposed research, improve their level of engagement, and provide input in the process. The recent Institute of Medicine recommendations for the Centers of Translational Science Awards (CTSA) highlighted community engagement as an essential element of the research enterprise. NIH welcomed those recommendations, and it would be inconsistent for the agency to use broad consent instead of dynamic and participatory consent.

The deficiencies of broad consent are considerable and well articulated. Participants cannot make genuinely informed decisions when sharing and decisions about secondary use of their data is beyond their reach and control. Broad consent is effectively consenting to have all the important decisions made by other people—its primary effect in practice is to marginalize and trivialize the trust and involvement of donors in research. Dynamic consent will provide an opportunity for researchers to gain participant input as the research field develops and progresses, and will enable participants to receive timely information about the research that is being undertaken.

The primary argument for broad consent—that it relieves researchers from having to engage in expensive, time-consuming recontact and re-consent of participants—is limited by the fact that broad consent does not protect research from changes in law and regulation, from innovative

new technologies that permit novel and un-anticipated uses of data, or from changing demands from publics or policy makers. Further, it diminishes the power of the connection between individuals and their data and samples. Only by integrating the whole of the individual, their family and their community into the research enterprise will researchers have the data they need to understand stratified medicine, and the contribution of the environment and microbiome.

In an era of participant-centered innovation and increased public engagement in science, research that treats participants as 'subjects' rather than participants, and static paper-based consent models, are becoming increasingly out-dated and unfit for the purpose of patient consent.

Dynamic consent is an alternative to broad consent that addresses the changing nature of biomedical research. Dynamic consent maintains and upholds participant respect by actively producing research as an ongoing partnership between participants and researchers. To achieve this, dynamic consent uses information technology to place patients and research participants at the center of decision–making. These technologies are ubiquitous in other sectors, but new to biomedical research. It makes what seemed onerous and impossible in the past, possible and simple.

There are advantages in employing a dynamic consent system.

1. This participant-centered paradigm of consent recognizes user autonomy and tailors the experience to meet individual needs.
2. Engaging participants promotes scientific literacy, transparency, and trust in research as participants become more informed about the research carried out on their samples and information.
3. An engaged and dynamic consent process creates an online, responsive, and highly engaged cohort of participants for researchers to contact regarding further studies or further collection of information.
4. A dynamic participatory process allows research governance to respond to changes in law and regulation, new scientific techniques and capabilities, and changing social perspectives by engaging with participants to discuss the changes rather than making assumptions about what patients 'would probably be comfortable with.'
5. It makes the consent process meaningful and allows for nuanced consent choices that avoid the 'all or nothing' flaw of broad consent.
6. In this age of abundant information, an engaged dynamic consent process meets the highest international ethical and legal standards for consent in a world where data protection laws are changing.

In 2007, the NIH conducted a public consultation to gather comments relating to the policy for sharing of data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS). The Notice outlining the result of this exercise stated that the *'NIH recognizes that the ethical considerations relevant to GWAS data sharing are complex and dynamic*.' Consent was a specific area of concern for respondents, with the Notice stating that efforts to address the complex nature of these issues would include *'discussion of the optimal methods for communicating with participants about relevant issues through the informed consent process for prospective studies.'* It also conceded that *'[t]he NIH anticipates that a number of GWAS proposing to include pre-existing data or samples may require additional consent of the research*

participants,' providing a clear example of the difficulty involved in setting up a system of broad consent that adequately caters to future research developments.

While this previous exercise specifically focused on GWAS studies, many of the concerns raised are directly applicable to the data sharing issues discussed in the Genomic Data Sharing Policy. The white paper produced by the global alliance states: "Within research, there are a number of participant-centric initiatives (PCIs) that use social media tools, offering new ways to engage with research participants. These can enable on-going communication, allowing individuals to give consent to research, specify personal privacy levels and to become partners in the research process in ways that have not been possible before. By enabling control over personal information and the potential to give on-going consent in real time, these initiatives meet international legal standards for the protection of privacy. Active engagement with the public and relevant governmental and regulatory officials will be needed to encourage the use of PCI and promote beneficial research while providing adequate privacy protections. In the long term, there needs to be greater transparency in data handling, commensurate punishment for mishandling of data, and governance procedures that include public input...". Renowned experts produced this white paper after much deliberation.

Genetic Alliance has developed the Platform for Engaging Everyone Responsibly (PEER), which uses cutting edge technology to give individuals to share their genomes, and other data, as they wish. They chose, after guidance from one of three members of their community, their own sharing, privacy and data access preferences. Open access repositories are part of the choices, and varying levels of de-identified to identified sharing are available. We believe that we can and must make the governance mechanisms and processes even more transparent and participant-centric given the personal nature of decisions and preferences in context[1]. As characterized by members of our Ethics Team and elsewhere, the shift to participant-centric models involves several important elements [2-7]. One is recognition that respect for persons requires asking permission and providing guidance to make meaningful discreet choices[8]. As many public opinion surveys have shown, there will not be a consensus regarding how comfortable people are with sharing personal health information, and with whom[9; 10]. The premise of the granular consent approach built into PEER is that we can move beyond "one size fits all" to tailored access preference management. Second, by integrating participants into every layer of decision-making and implementation, biomedical research will allow for an adaptive and responsive governance approach that makes context-specific decisions as needed, recognizing that the needs of the individual, the investigator and the biomedical research enterprise evolve over time[3]. With participants involved throughout the governance processes, we can move closer to trustworthy systems, and avoid making decisions on behalf of someone's interests unless deputized to do so[11].

Finally, in addition to the inherent participant-centric focus and design, PEER also follows best practices as outlined by disease advocacy-run biobanks and research registries[12]. These include keeping all decisions mission-focused with the mission defined as those actions and activities that advance positive impacts on human health. Other practices include being flexible and creative with partnerships and forging collaborative efforts with a full range of public-private partners who share and value the common purpose of advancing health. Combined with the adaptive approaches utilized by participant-centric designs, PEER works creatively to respond to

opportunities and find solutions when barriers are met. Our tool is ready to be put to use, and has already been deployed in a series of campaigns and activities:
https://www.reg4all.org
https://www.trialsfinder.org
http://www.free-the-data.org
and 8 sites for a patient focused drug development initiative to support FDA in PDUFA V activities.

This is an opportunity for NIH to lead on behalf of participant-centric solutions.

Sincerely,

Sharon F. Terry, MA
President and CEO
For Genetic Alliance Council and Staff

1. Meslin, E.M. (2010). The value of using top-down and bottom-up approaches for building trust and transparency in biobanking. Public Health Genomics 13, 207-214.
2. Kaye, J., Curren, L., Anderson, N., Edwards, K., Fullerton, S.M., Kanellopoulou, N., Lund, D., Macarthur, D.G., Mascalzoni, D., Shepherd, J., et al. (2012). From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 13, 371-376.
3. O'Doherty, K.C., Burgess, M.M., Edwards, K., Gallagher, R.P., Hawkins, A.K., Kaye, J., McCaffrey, V., and Winickoff, D.E. (2011). From consent to institutions: designing adaptive governance for genomic biobanks. Soc Sci Med 73, 367-374.
4. Norman, T.C., Bountra, C., Edwards, A.M., Yamamoto, K.R., and Friend, S.H. (2011). Leveraging Crowdsourcing to Facilitate the Discovery of New Medicines. Science Translational Medicine 3, 88mr81.
5. Horn, E.J., Edwards, K., and Terry, S.F. (2011). Engaging research participants and building trust. Genet Test Mol Biomarkers 15, 839-840.
6. Fullerton, S.M., Anderson, N.R., Guzauskas, G., Freeman, D., and Fryer-Edwards, K. (2010). Meeting the governance challenges of next-generation biorepository research. Sci Transl Med 2, 15cm13.
7. Anderson, N., and Edwards, K. (2010). Building a chain of trust: using policy and practice to enhance trustworthy clinical data discovery and sharing. In Proceedings of the 2010 Workshop on Governance of Technology, Information and Policies. (Austin, Texas, ACM), pp 15-20.
8. Greely, H.T. (2007). The uneasy ethical and legal underpinnings of large-scale genomic biobanks. Annu Rev Genomics Hum Genet 8, 343-364.
9. Kaufman, D.J., Murphy-Bollinger, J., Scott, J., and Hudson, K.L. (2009). Public opinion about the importance of privacy in biobank research. American journal of human genetics 85, 643-654.

10. Kaufman, D., Murphy, J., Erby, L., Hudson, K., and Scott, J. (2009). Veterans' attitudes regarding a database for genomic research. Genet Med 11, 329-337.
11. Yarborough, M., Fryer-Edwards, K., Geller, G., and Sharp, R.R. (2009). Transforming the culture of biomedical research from compliance to trustworthiness: insights from nonmedical sectors. Acad Med 84, 472-477.
12. Edwards, K. (2013). Governance of Registries and BioBanks.

Commenter: Tyler Gibson

Date of comment: 9/24/2013 13:06

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: There is very little here on the participant's right to access his/her own genomic data after participating in the study. Typically, studies that fall under the guidelines laid out in this proposal refuse to share the data gathered with participants. This is not due to any intrinsic challenge in transferring the data to the participant as requested, which is trivial in the digital age. The end result is that only investigators with direct access to sequencing equipment get their genomic information provided back to them, while participants are denied access. Investigators should be encouraged to allow participants to access their own freely-donated genomic data in a format that they can either view or share with other parties (e.g., a health care professional) in a manner that said participant chooses.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: David M. Gilbert

Date of comment: 10/2/2013 22:38

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Section 2, Data Repositories: First and second sentence seem contradictory. First sentence categorically states all data should be registered with dbGaP. Second sentence says it may be submitted to any of several data repositories, including dbGaP. Meaning should be clarified. Section 4. Informed Consent. This section implies that some sort of formal change will be made in consent forms and that all samples collected after s certain date will require this. What will the consent form look like? Will it be created by NIH and distributed to all NIH-funded investigators prior to the data of this policy?

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Lynn R. Goldin

Date of comment: 11/19/2013, 11:26

Comment:

Section II. Scope and Applicability: 1. Proposal for sharing omics data is extremely broad. 2. Additional time is needed for public comment on such a broad plan with so many consequences to researchers.

Section IV.A. Data Sharing Plans: 1. These requirements put a substantial burden on investigators because of the volume and complexity of the data.

2. The timeline for data submission is too quick given comment 3. The data cleaning period is inadequate.

3. The requirement to immediately post bam files for DNA sequencing data is problematic because alignment programs are variable among users and change frequently. No standards have been given for these files and no process for determining standards is included.

4. It would be preferable to share the variant call format files with details on the methods used to generate the data.

5. Will NIH pay centrally for the level of management and storage of the data that will be required by this policy? DNA sequence (and other omics data files) are extremely large and will require substantial storage facilities. A much increased level of data management will be required and this will greatly increase the cost of studies. Current resources and databases are inadequate.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Alisa Goldstein

Date of comment: 11/15/2013 14:58

Comment:

Section II. Scope and Applicability: The draft does not appear to recognize the effort and resources required to comply with the scope of the policy. The data that will be posted will swamp dbGAP. The amount of storage required for all of the genomics data is also daunting and it is not clear how/where all of the data will be stored.

Section IV.A. Data Sharing Plans: The draft does not appear to recognize the complexity of analyzing genomic data, particularly next generation sequencing data, and the amount of time required to analyze the data. There is also concern about the timing for posting data given the complexity and time required to analyze the data. Sequencing data is often re-annotated and there is no mention about altering posted data if new annotation, alignment, etc. is done resulting in changes to the variants reported. Realignment and re-annotation is a regular ongoing activity that will likely mean there are major discrepancies between what is posted and what should be used for analysis. There is no mechanism about updating results and about how to indicate whether the BAM files are and/or should be altered.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: The draft does not appear to recognize the complexity of analyzing genomic data, particularly next generation sequencing data, and the amount of time required to analyze the data. There is also concern about the timing for posting data given the complexity and time required to analyze the data.

Sequencing data is often re-annotated and there is no mention about altering posted data if new annotation, alignment, etc. is done resulting in changes to the variants reported. Realignment and re-annotation is a regular ongoing activity that will likely mean there are major discrepancies between what is posted and what should be used for analysis. There is no mechanism about updating results and about how to indicate whether the BAM files are and/or should be altered. If updating results is added to the policy, it will add another level of complexity and cost to an already overwhelming situation.

There is concern that potential study participants will choose not to participate in genomics based research if they are required to allow their data to be made available to the larger scientific community. For rare familial disorders, this is a major potential problem since the potential study sample is very limited and without the critically informative research subjects, the research cannot be effectively conducted.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: H. Lee Moffitt Cancer Center and Research Institute, Thomas A. Sellers, PhD

Date of comment: 11/19/2013 19:53

Comment:

Dear NIH officers:

Thank you very much for the opportunity to comment on the Draft Genomic Data Sharing Policy published in the Federal Register on 09/20/2013. While at its core, the philosophy behind this policy is laudable, the leaders at the H. Lee Moffitt Cancer Center & Research Institute have reviewed the draft policy and have several key concerns. This is particularly germane to our work because, in addition to our genomics work, we have one of the largest prospective longitudinal annotated tissue/data repositories in the nation.

Name of Individual(s) Submitting Comments: Thomas A. Sellers, PhD, on behalf of more than 400 Moffitt Cancer Center Research Members and Clinicians

Affiliation: Commenting on behalf of: A research, healthcare, and academic organization

Name of Organization: H. Lee Moffitt Cancer Center and Research Institute

City: Tampa

State: Florida (FL)

Comment 1:

For human data, the value is in the associated information, such as therapy, time to recurrence, overall survival, etc. Without these data, the genomic information is not meaningful. We have concerns over the source of funding for the capture of this additional data. If the expectation is that federally funded grants support this effort, it will not have been included in the original budgets, which are also being cut across the board. While section I.V.A. discusses the need to include this in grant budgets, the limitation of the budgets will impact the ability or potentially the quality of the submission. The time, effort and funding necessary to validate the data coming from a wide variety of sources (even under these guidelines) is underestimated and could undermine the value such a repository if quality becomes questionable by the community.

Comment 2:

Policies like this have the potential to move into other areas such as "radiomics." It would be very burdensome to upload all of the images, and all of the radiomic and medical data.

Comment 3:

We have concerns about confidentiality regarding the ability of a Principal Investigator (PI) to protect his/her ideas for future use of the data. An investigator might have designed a study with many Aims in mind, even if there were only enough money in the grant application to address some of the Aims. If the investigator's strategy were to fund the recruitment, sample collection, and lab testing through an initial grant, and had only enough funds to request statistical analysis of one set of genes, but had plans to write additional grants to fund more statistical analyses, how could he/she protect her ideas before someone else grabbed the data off the website to analyze him/herself? This could be particularly important with earlier data release (as mentioned in section IV.B.1. or IV.C.2.). In IV.C.3 , while a tiered system with NIH providing the approval may protect the patients, it does not sufficiently safeguard the investigator's research related to the data they are required to post.

Comment 4:

In section V.B., there is guidance on acknowledgement in publications of future work, but there is not a mention of co-authorship or significant recognition of scientific contributions to future studies.

Comment 5:

Nationally, individual PIs may have limited experience de-identifying and transmitting genomic data sets. This could be an important point, particularly for junior investigators.

Comment 6:

In Scope and Applicability (Section II), the policy references NIH-funded research. While this seems simple at face value, this would need to be defined. Because of the trend towards shared funding, does this mean funding of any part of the genomic data produced, or if the infrastructure used to produce it has any funding, or if it is carried out under a grant where the actual genomic data is collected under another source? This distinction has profound implications. We suggest that this also be clarified to refer to all federally-funded research.

Comment 7:

There should be some window of time explicitly stated when data requests will be reviewed (section V.A.) and will deliver a response perhaps similar to the FDA, where submissions after a 30-day window carries the assumption of approval. This is particularly critical if data is necessary for a grant or project application (e.g., IV.C.3.). While the role of the Data Access Committees is well defined, details of how they will operate is lacking and the challenges they will face is underestimated. Care will have to be taken to ensure the responsiveness and integrity of the DACs.

Comment 8:

As in comment #7, this will be critical if annual approvals are required, where waiting for approval could stifle scientific progress on an active project (section V.A.). We suggest consideration of having approval for the project period, rather than annual renewals.

Comment 9:

In section C.4., the NIH should present model consent language in addition to expectations, to ensure that NIH specifications are met.

Comment 10:

In section V.A., the expectation that data can only be shared with individuals other than those listed in the data access request should have some stipulation for exemptions (e.g., junior investigators working with senior investigators, trainees, etc.) and/or changes (approval allowed at the local level in the middle of a year, with inclusion on the subsequent request). It is not reasonable in terms of institutional or NIH workload to address changes, unless these are handled as in the FDA example mentioned above (similar to 1572 form changes).

Comment 11:

In Data Submission Levels 2 and 3 (p16), we suggest greater reliance on a published date versus time of acceptance of publication. Publications can sit in press up to a year and it would not be reasonable to have the data made available prior to publication.

Sincerely,

Tom Sellers, PhD

Thomas A. Sellers, PhD

Director, Cancer Center & Research Institute

Executive Vice President

Moffitt Cancer Center

12902 Magnolia Drive Tampa, FL 33612 | **tel:** 813-745-1315 | **fax:** 813-449-8126 | **email:** Thomas.Sellers@Moffitt.Org http://labpages.moffitt.org/sellerst/

Commenter: Jonathan Haines, Anthony Wynshaw-Boris, Aaron Goldenberg, Stuart Younger, Stanton Gerson, Sudha Iyengar

Date of comment: 11/20/2013 16:17

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: "Any resources needed to support a proposed data sharing plan should be included in the project's budget." We agree that such funds are essential and must be provided since preparation and deposition of genomic datasets often requires substantial time and effort. However, and particularly in times of tight budgets, automatic budget cuts, and budget caps, including this in the initial budget can significantly reduce the funds available to perform the actual research. In addition, it is often difficult to estimate such costs before a study is completed. An alternative solution is to provide administrative supplemental funding awarded after the primary grant is funded and when a better estimate of cost can be made. Such costs will be highly dependent on the details of submission requirements, which will likely vary from dataset to dataset. Datasets with retrospective consent may not contain all samples, and obtaining this type of consent may require additional resources.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Section C.1: Data submission expectations and timelines. "The NIH will release data…no later than six months after the initial data submission to an NIH-designated data repository…" Given the very large size and complexity of many, if not most, of the datasets, providing only a six month window for the primary data generators to explore, analyze, and interpret their datasets is unreasonable. While some extremely large labs or centers may have the bandwidth to accommodate such a tight timeline, most researchers, labs, and institutions are not so privileged. We suggest a more reasonable time frame of one year from submission of the final dataset, or publication of the first paper resulting from a primary analysis of the bulk of the dataset, whichever comes first."The de-identified data should be assigned a random, unique code, and the key held by the submitting institution." This assumes that all data is generated and hosted by a single institution. In the case of collaborative efforts that combine data across institutions (now a common occurrence), this requirement could either violate the confidentiality of participants or generate a highly complex and easily broken web of multiple random identifiers for data from a single sample or dataset.Section C.2: Data Repositories. "Applicable studies…should be registered…no later than the time that data cleaning and quality control measures begin." We have two concerns with this statement. First, datasets often change in size and structure due to quality control measures and/or addition or subtraction of subsets of the data. Thus revision or removal (neither of which is a simple process) is not only possible, but likely. This represents a significant burden on both the repositories and the investigators. We propose that registration should occur when a dataset has been finalized (e.g. a data freeze has been declared) and data cleaning/QC has been completed.Second, it is common practice to combine datasets across multiple investigators (e.g. through networks or consortia), to improve power and advance discovery. Some, if not the

majority, of such datasets may have been contributed individually or as part of other projects. Tracking such submissions will be logistically difficult, if not nearly impossible. Without such tracking, however, the same samples/datasets may be inadvertently used to both discover and replicate a result. Aggregated datasets may need a specific and separate requirement.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: 1). The policy is generally silent on what non-genomic (e.g. associated) data should also be submitted. We would argue that the primary outcome (e.g. disease status, quantitative trait measure) and any necessary co-variates (e.g. typically age, gender) be the maximum required for submission. Inclusion of additional data raises the specter of intentional or unintentional re-identification and increases concerns about privacy and confidentiality.2). Requesting broad consent for prospective collections will require investigators to modify their explanations during the consent process. NIH should, through a broad, deliberate, and consultative process, develop and provide patient educational material and tools to help guide this process and provide a more consistent understanding of the purpose of such consent. Use of such materials and tools would be optional.

Commenter: Jonathan Haines, Anthony Wynshaw-Boris, Aaron Goldenberg, Stuart Younger, Stanton Gerson, Sudha Iyengar

Date of comment: 11/20/2013 16:18

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Section C.4: Informed Consent. There is an expectation of obtaining from the participants the broadest possible consent for current and future use of the data. For recruitment into a study, willingness to participate may be impacted by the scope of who has access to their data and how that information is used. For example, communities who have lower levels of trust of either in a medical institution or government agencies may also have increased level of concerns about a policy for data sharing. In these cases, an expectation of obtaining from the participants the broadest possible consent for current and future use of the data may be a hindrance for recruiting participants from already marginalized communities or groups of participants who tend to have lower levels of participation in genetic research more generally. Additional guidance and model (but not required) language would be very helpful to address the concerns of these communities.Section C.7: Exceptions to data submission expectations. Some groups and communities have potential concerns about the collection and use of genetic information. The data sharing policy is silent or vague on this issue, which may have substantial impact on subsequent recruitment from these groups. Thus more explanation and guidance on granting exemptions for this policy would be helpful. There are many additional circumstances in which exceptions should be granted, particularly to respect the wishes of the participants. One such example is the need to consider and respect the cultural or religious beliefs of individuals or communities and vulnerability of the populations (e.g. indigenous or isolated populations such as Native Americans and the Amish, or children). Recognition of these additional potential exceptions should be provided in the policy so that there is more consistency in how and why exceptions are granted.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Harvard University, Ara Tahmassian

Date of comment: 11/20/2013 16:05

Comment:

Section II. Scope and Applicability: I am writing on behalf of Harvard University to offer comments on the draft Genomic Data Sharing Policy (GDS), which was issued on September 20, 2013. In principal, we agree with the philosophical underpinnings of the draft policy, that the greater availability of research findings can increase the efficiency of scientific research, "maximize… knowledge gained," and "enhance… public benefit by helping to speed discoveries that increase the understanding of biological processes that affect human health and the development of better ways to diagnose, treat, and prevent disease."

At the same time, the draft policy raises a number of concerns that have been well articulated in the comments and the recommendations submitted on November 6, 2013, by the Federation of American Societies for Experimental Biology (FASEB) and in the comments submitted on November 20, 2013 by the Council on Governmental Relations (COGR). Harvard endorses and supports the concerns raised, and recommendations made in both comments.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

THE HUMAN VARIOME PROJECT
an NGO official partner of UNESCO
sharing data • reducing disease

The Human Variome Project International Limited
Level 5, 234 Queensberry Street, The University of Melbourne, Victoria 3010, Australia
Phone +61 3 8344 1893 | Fax +61 3 9347 6842 | info@variome.org | www.humanvariomeproject.org
ABN 20 145 251 485

Interim Board
David Abraham (Chair)
Chris Arnold
Richard Cotton
Xitao Li

Interim Scientific Advisory Committee
Garry Cutting (Chair)
Arleen D Auerbach
John Burn
Mireille Claustres
Richard Cotton
Aida Falcón de Vargas
Johan den Dunnen
Marc Greenblatt
Stephen Lam
Finlay Macrae
Yoichi Matsubara
Mona El-Ruby
Gert-Jan B van Ommen
Mauno Vihinen

Non-Voting Members
Richard Gibbs
Ming Qi

19 November, 2013

Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892
GDS@mail.nih.gov.au

Submission regarding the Draft NIH Genomic Data Sharing Policy

1. Background to Submission

The Human Variome Project is an international consortium of scientists and health-care professionals who are working towards a significant reduction in the burden of genetic disease on the world's populations. The aim of the Human Variome Project is to ensure that all information on genetic variation can be collected, curated, interpreted and shared freely and openly. This will lead to speedier, better and cheaper diagnosis and treatment of genetic disorders, and better insight into the causes, severity and effect of common disease. The Human Variome Project achieves its aims by establishing and maintaining the necessary standards, systems and infrastructure, by providing education and training to scientists, clinicians, genetic counsellors, other healthcare professionals and the general public and by assisting nations build their capacity in medical genetics and genomics. The Human Variome Project acts as an umbrella organisation and works to encourage communication and collaboration around its central vision.

The importance of the Human Variome Project was recognised in 2011 by the United Nations Educational, Scientific and Cultural Organisation in the Project's admittance to Consultative Partner status.

2. General Comments on the Draft Policy

We are very pleased to see that the NIH is taking an active interest in data sharing and will be, through this policy, mandating the submission of data generated by the array-based and high-throughput research projects they fund. The Human Variome Project exists to facilitate the free and open sharing of just this type of information and we applaud the NIH on taking the first steps down the road to full data sharing.

However, we feel that this policy does not yet go far enough in terms of scope or vision, to fully enable and promote the routine sharing of genetic variation

information by activities funded by the NIH. We have outlined our concerns in detail in the sections below.

3. Comments on Specific Sections of the Draft Policy

3.1 Scope and Applicability (Section II)

The Policy as drafted makes mandatory the submission of data generated by "all NIH-funded research that involves large-scale human and nonhuman genomic data produced by array-based or high-throughput genomic technologies, such as GWAS, SNP, whole-genome, transcriptomic, epigenomic, and gene expression data, irrespective of funding level and funding mechanism."

This is an admirable start, but we must ask why limit this policy to "large-scale" genomic data? There is undoubtedly much data generated by NIH-funded research projects that would not be considered "large-scale" that has potential value, both in terms of re-use in future research projects and also clinically, when diagnosing and treating patients afflicted with rare, genetic disorders. The technical and administrative effort required to submit such data, to an NIH-funded repository (or otherwise) is not burdensome and could be easily integrated into existing research reporting workflows.

The NIH, in drafting this Policy, obviously acknowledges the benefits of wide-scale data sharing. It seems unnecessarily restrictive and potentially confusing, to limit the scope of this policy to certain research projects that are categorized only under the ambiguous phrase "involves large-scale…genomic data."

3.2 Human Genomic Data (Section IV.C)

We are pleased to see that the draft policy does not seek to restrict the submission of data by research projects within the scope of the Policy solely to NIH-funded repositories. Although NIH-funded repositories play a key role in the emerging data-sharing infrastructure that is operating globally, they are by no means the only places that such data could be stored and shared. Indeed, in quite a large number of cases, future research projects and clinical practice would be better served if the data from NIH-funded research projects were shared much more broadly than between NIH-funded repositories. Numerous disease-specific research and clinical communities are well served, and have been well served for over a decade, by international gene/disease specific databases (GDSDBs) or locus specific databases (LSDBs). While it could be argued that individual researchers working in these disease areas would be aware of these extant resources and include submission to them in their data sharing plans, and this might very well be the case. However, we still believe that it would be wise for the NIH to include a more direct acknowledgement of data sharing options beyond those funded by the NIH. The desired end result of this Policy is obviously to better human health through improved research and clinical delivery: encouraging researchers to think more carefully about where the submission of their data would be most useful would go a long way to reaching this end result.

Finally, we are a little concerned over the implied insistence of submission to an NIH-funded repository as the minimum requirement of this policy, as indicated by the sentence: "NIH-designated data repositories need not be the exclusive source for facilitating the sharing of genomic data." Our concerns are related to how this policy would be applied to those

research projects that generate data, via overseas partners and collaborators, from samples derived from non-US citizens and within a different ethical, legislative or social context. In these situations, researchers may have difficulty, or be prohibited from sharing data, either completely, or in part, to a repository that is operated within a different country and funded by a separate sovereign power. While we are sure that such a circumstance would be taken into consideration if articulated in the required data sharing plans, the draft Policy itself is silent on the matter and would benefit from this matter being addressed.

4. Concluding Remarks

The Human Variome Project fully endorses the proposed draft Policy. While we believe that the Policy could be extended to further promote and encourage the free and open sharing of data to enable more expansive use within both research and clinical contexts, it is a good first step.

Yours Sincerely,

Richard G.H. Cotton AM PhD DSc FRCPA (Hon.)

Scientific Director
Human Variome Project

for the Human Variome Project International Scientific Advisory Committee

Commenter: HVTN | HCRI-Uganda, Rachael McClennen

Date of comment: 11/19/2013 1:32

Comment:

1. It is not clear how this policy impacts the HIV Vaccine Trials Network (HVTN) given that we do what we consider to be limited genetic testing (e.g. HLA typing). Would this policy apply to HLA typing?

2. In a Network situation, it's not clear who is the submitting institution when we share our specimens for protocol-related work or even ancillary studies. In some cases, we do fund outside investigators to do this work. In other cases, we do not. Who is responsible for assuring that the data will be submitted in a timely way? The institution where the work is being conducted? Or the institution that provided the specimens and NIH-funding through a subaward?

3. As a federally funded repository, what responsibilities does the HVTN have to enforce the data sharing policy for external investigators to whom we supply specimens? Does this apply to all data from specimens collected (through a clinical trial) under NIH-funded research or does it apply to NIH-funded research on these specimens?

4. Section III A. Data Sharing Plans. The guidance indicates that budget should be included. However, awards have had substantial budget cuts lately. We question whether it is reasonable to flatly hold institutions accountable to this requirement under the current budget situation. Could there be a provision to waive this requirement if we are not sufficiently funded to do the activities outlined in the data sharing plan? This is of particular importance for grantees that submitted their budgets prior to this policy being issued (as is the case for the HVTN), such that they would not have included budget provisions for this activity. This is of particular concern for multi-year grants, such as Networks like the HVTN.

5. We need clarification about the definition of nonhuman and model organism genomic data. Is this referring to viral genomic work that Networks may do on specimens collected from humans? There seem to be no provisions for human subjects in the nonhuman process, which would be problematic if the samples are collected in the context of a clinical trial.

6. Appendix A. We object to the timeline for data submission "within 3 months after data generation" and the notion that data could be held for up to 6 months in a limited access area. Given the size of HVTN studies, we may not have data cleaned/locked by that time. The guidance indicates that we could indicate the timeline in our data sharing plan. However, for the HVTN, this is a very broad data sharing plan. Would we need to develop project-specific data sharing plans?

7. C.7. Exceptions. We're confused about how this would work in our studies. We may have countries, sites and/or individual participants that may restrict genomic testing and subsequent uploading. At what point do we need to request an exception? At the time of the grant application? This would not be feasible given that our Network funding is a 7-year cycle. In addition, obtaining exceptions for specific participants seems onerous. And to note, that there may not be "other mechanisms" to share data in some countries or it may not be appropriate (if the participant did not give permission to share).

Thanks,
Rachael

Rachael McClennen Regulatory Affairs Associate HVTN | HCRI-Uganda Office: 206-667-2124 Mobile: 206-391-0993	Office Hours: Please note that I may be out of the office on Mondays and Wednesdays starting at 3 PM. If you need to reach me urgently, please don't hesitate to call my mobile.	Mailing Address: Fred Hutchinson Cancer Research Center 1100 Fairview Ave. N., E3-300 PO Box 19024 Seattle, WA 98109-1024 UW Courier: Box 358080

Confidentiality Notice: This e-mail message and any attachments may be confidential and privileged. If you received this message in error, please destroy it and notify the sender. Thank you.

Commenter: Hae Kyung Im

Date of comment: 11/19/2013 18:07

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: The access to level 3 and level 4 data will be extremely useful for secondary analysis and data integration across multiple datasets. However, it may be useful to clarify that the full set of results from the analysis of the data, not just the top significant ones, should be made available for secondary use. This would allow the secondary analyst to choose from different thresholds for significance.

This comment has not been reviewed by The University of Chicago, and as such does not reflect the opinion of the institution.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: In dbGaP, for datasets that have been consented for general research use, it would be useful to have a more streamlined application process where the full set of datasets can be requested with a single application. The user would describe the use of the data that have been downloaded in an annual report and he/she would acknowledge the datasets used when sharing/publishing results.

This comment has not been reviewed by The University of Chicago, and as such does not reflect the opinion of the institution.

INDIANA UNIVERSITY
INDIANA CLINICAL AND TRANSLATIONAL SCIENCES INSTITUTE
School of Medicine

November 18, 2013

Submitted electronically at www.regulations.gov

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

RE: FR Document 2013-22941 [Draft NIH Genomic Data Sharing Policy Request for Public Comments]

Dear Members of the Genomic Data Sharing Policy Team,

We are writing on behalf of the Indiana Clinical and Translational Science Institute to provide our comments on the Draft NIH Genomic Data Sharing Policy ("Draft Policy") (Federal Register 78:57860, 20 September 2013). We congratulate the NIH on preparing and circulating this Draft Policy for public and community comment. Getting clear on these issues are central to the goals of translational science and it is hoped that the policy will advance the goals of high quality science undertaken in an environment of ethical integrity.

While our comments were very much informed by the excellent work of the Biobanking Working Group of the national CTSA Consortium's Clinical Research Ethics Key Function Committee, and whose written submission our comments directly resemble, we felt it was important to provide a separate set of comments indicating the importance of these issues. A unifying thread in the comments is the need for more specificity and guidance to IRBs and equivalent groups, which we think is necessary to encourage more uniformity in data sharing.

1. It is not clear whether data generated from biospecimens collected before the effective date of the policy according to a consent process not outlined by the policy will be acceptable for deposit into dbGaP or other NIH data repositories.

   The Section C.4 of Draft Policy states:

   "in these cases, an assessment by an IRB, Privacy Board, or equivalent group is essential to ensure that data submission is not inconsistent with the informed consent provided by the research participant."

The phrase “not inconsistent with the informed consent provided” is open for interpretation.
Indeed, it is possible that IRBs, and other groups with the authority to exercise discretion when making this decision, may take a more restrictive view and decline a proposal that seeks to deposit these data. The result is that data from a number of existing specimen repositories would not be eligible for deposit into dbGaP or other NIH data repositories.

In addition, the draft policy suggests that existing biobanks and biorepositories may need to modify their current consent documents to be in compliance. For those institutions who consent policies and procedures may not be in compliance it is possible that they will need to undertake a comprehensive review of their consent process, leading to reconsent procedures for all biobanks and biorepositories. This would require a huge investment in resources, but it may be the stance that some IRBs take based on the passage in the draft policy cited above. Criteria or benchmarks for assessing whether institutions already are in compliance may be a useful component to be added to the policy.

2. The Draft Policy articulates a requirement that phenotype or clinical data should be submitted but fails to define how much associated phenotype/clinical data should accompany the genomic data.

   While the intent may be to allow the investigators and institutions themselves to determine what they are comfortable broadly sharing, it remains unclear how much discretion each institution or investigator has. As with the consent requirement above, it is unclear if the NIH has an expectation regarding the type and volume of data to be shared. Conversely, if it is truly the intention of the policy that institutions retain this discretion, then that too should be clarified in the policy. Either way, the public should have an opportunity to review and comment on this aspect of the policy.

   Both the 2007 data sharing policy for data generated in genome wide association studies and the current Draft Policy fail to articulate the consequences and penalties for any investigator who intentionally breaches the privacy and confidentiality interests of the individual research participants whose data are being deposited and broadly shared. Since several publications have demonstrated that it is possible to identify individuals in databases of “de-identified data,” (including studies that can re-identify individuals whose data had been de-identified) the NIH has an ethical obligation to be transparent regarding sanctions for intentional abuse or misuse of the genomic and associated phenotype data covered by the Draft Policy. Not only will this aid investigators, but it will go some distance in giving the public the confidence they need to trust science.

3. Given increasing evidence that research consent forms should be simplified to increase participants’ comprehension, the practicality of implementing the following proposed expectation (also in Section C.4 of the Draft Policy) needs careful attention.

   “The NIH expects the informed consent process and documents to state a participant’s genomic and phenotypic data may be shared broadly for future research purposes and also explain whether the data will be shared through open or controlled access.”

   If this or any similar expectation remains in the final policy, it would be helpful for NIH to provide an example of concise language that would be acceptable to achieve the stated expectation. At the same time, if it is the intention of the policy that institutions, through their IRBs, still retain broad discretion to approve language in consent forms, then we would encourage the policy be explicit on this matter as well.

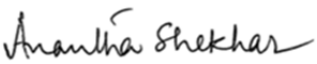
4. Finally, we encourage the NIH to undertake three further activities

    a. Sponsor a public education program to elevate understanding of the value of broad data sharing, and the importance of open access to genomics data generated with public funds. Such a program could be developed through an RFP to ensure high quality, peer review
    b. Undertake an evaluation of the impact of the policy after a suitable period of time to assess whether it is accomplishing the goals it set out to achieve.
    c. Support ongoing studies of the ethical, legal, and social implications of data sharing more generally

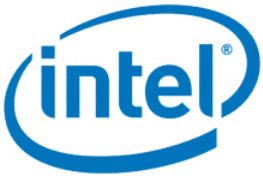
We appreciate this opportunity to offer these suggestions from members of the research ethics community working in CTSA programs across the country.

Sincerely,

Eric Meslin, PhD
Director, Indiana University Center for Bioethics
Associate Dean for Bioethics
Indiana University School of Medicine



Anantha Shekhar, MD, PhD
Director, Indiana CTSI
Raymond E. Houk Professor of Psychiatry
Associate Dean for Translational Research
Indiana University School of Medicine



November 20th, 2013

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

Re: NOT-OD-13-119

Intel would first like to express appreciation to NIH for proposing an updated Genomic Data Sharing Policy. Given the acceleration of genome sequencing and the technology available to build algorithms that can draw upon big data from research to personalized diagnosis and treatment, it is ever more urgent that this data be shared among NIH entities, academic research centers and international centers of excellence. We are particularly pleased that these standards will be a condition in the contract award for all NIH grantees.

Intel and Health Care

Intel is known as a world leader in silicon innovation, but our company is also active in the healthcare and life sciences arenas both directly and indirectly.[1] Our technologies help to power the Internet, the broadband connected world, and many biomedical and life sciences institutions globally. Our objective is to provide innovative technologies and solutions that connect patients, families, providers, and healthcare researchers with one another. For more than a decade Intel has also focused a portion of its research and development efforts specifically on healthcare to better understand how to connect all of the major players through

[1] Additional information about Intel is available at www.intel.com/pressroom and http://blogs.intel.com/policy

this wide array of health information technologies. Intel social scientists, medical informaticists, clinicians, and engineers have studied more than 1,000 patient homes and 250 hospitals and clinics in more than 20 countries to inform the development of products and solutions that can help bring forth a connected world for healthcare.

Our technologists and architects have advised and led Health IT standards efforts creating what we call solution blueprints for healthcare entities in many parts of the world. For example, Intel architects helped the UK with its National Health Information Backbone (Spine) and N3 Architecture. We have worked with similar regional and national health exchange efforts in Canada (with Health Infoway) and China (with its Regional Health Information Network or RHIN requirements) among many others. And more recently, we have partnered with organizations globally to accelerate whole genome processing - reducing the compute analytics time from months to hours. As one example, Intel has worked with Schrödinger to test a cancer drug concept, developing a configuration that ran 16 million molecular simulations in an hour and developed a 1000 molecule list in eight hours. The innovations in I/O acceleration, compute storage, network integrated fabrics and security are contributing to a robust ecosystem by incorporating tools like Hadoop and Lustre into the genomics processing analytics and pipeline optimization.

Data Sharing Essential to "finding the cure"

Biomedical informatics combined with data-sharing has already provided new insights into autism, diabetes, depression, arthritis as well as leading to improvement in quality and safety. But we are only scratching the surface of our complex biology, health maintenance as well as the environmental determinants of health and disease. We have entered an era where 'big data' can offer big answers, for example, by uncovering uncommon complications of new drugs through access to large patient datasets.

To capture the value presented in datasets and to encourage researchers worldwide to advance the global research of personalized medicine, we would suggest that not only should research data be appropriately consented and contributed to NIH designated data repositories or other widely used data repositories as a requirement of funding, but should also be a

requirement of scientific publication. We strongly concur that awardees should agree to share the data through the open or controlled access provisions and, in addition, to submit a data management plan, similar to the requirements for NASA's EARTH Science Division[2], thereby, creating a culture of data sharing to advance future research.

We concur with the informed consent parameters requirements and are pleased to see that past studies with a variation of informed consent from patients will be considered for data sharing based upon assessments made by an IRB, Privacy Board or equivalent groups. The ability to use longitudinal data from important research of 5, 10, or even 20 years prior to the new data sharing rule is paramount in providing scientists with the tools needed for determining treatments based upon genomic and environmental data. However, we would suggest that the typical IRB review process could be streamlined to expedite the use of data, while providing sufficient security and privacy controls. Data Use Limitation (DUL) and the corresponding consent should be "standardized" for studies for current and future use. Guidance on crafting consent that encourages the broadest data use, while educating participants at the time of consent is key to maximizing the value of data sharing.[3]

Additionally, we recommend that participants retrospectively have access to the studies that have used their data. Giving such visibility may encourage sharing and certainly better informed patients.

Privacy and Security

To keep privacy and security risk manageable, enable much broader sharing of data, and support research that requires more than fully de-identified data, the best practice of a multi-layered approach to security should be used. De-identification is combined with other safeguards including encryption, tokenization and access controls which must be usable,

---

[2] http://science.nasa.gov/earth-science/earth-science-data/data-management-plan-guidance/.

[3] http://gds.nih.gov/pdf/NIH_PTC_in_Drafting_DUL_Statements_3-13-12.pdf.

performant and robust to effectively mitigate risk and avoid end user workarounds when security gets "in the way." In addition, Intel urges adoption of IT architecture designed to store this data meet federal government security standards as established by the National Institute for Standards and Technology (NIST).

Many healthcare, academic research and life sciences organizations still take a traditional perimeter approach to privacy and security, where there is over-reliance on perimeter controls such as firewalls in the logical sense and buildings in the physical sense. End user technologies enable anytime, anywhere access to anonymized data inside this perimeter. Cloud moves the data out of this perimeter and into the cloud provider's data center. Malware infections routinely occur inside security perimeters of health care organizations. With this in mind, Intel strongly recommends protecting healthcare data directly, wherever it is at rest or in transit, including the use of encryption on research clients, servers, databases and backup systems.

Intel strongly advocates a holistic approach to privacy and security including the use of policy, risk assessments, procedures, training, and technology. Risks resulting from implementation vulnerabilities and operational aspects of the research should be analyzed, including use cases for security key management. Highest priority risks identified should be mitigated through a combination of administrative, physical, and technical controls.

Interoperability:

As NIH creates a sharable secure database from diverse research studies, we recommend consideration be given that will:

- Create and ensure the interoperability of technical standards for managing and sharing sequenced data in research and clinical samples.
- Develop the technology platform with open standards designed to enable secure storage; with a computational architecture and application programming interface (API) supporting apps and services. These standards need to provide global interoperability, scalability, stability and resiliency; serving as building blocks for further

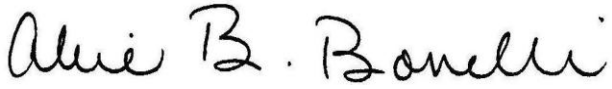
innovation; and contributions to a global community of data for disease cures and treatments.

- Develop standards to deal with the new dynamics of data – where to store, access and make use of millions of variants for each individual.

Making genomic data and tools interoperable in a secure and trusted manner will generate a powerful network effect: the more data exchange on common platforms, the more value to patients, researchers and healthcare professionals.

We commend NIH for taking this important step towards creating a shareable data base that will impact the future of the discovery of treatment options targeted to each and every individual's health.

Sincerely,



Director, Global Healthcare Policy
Intel Corporation
1155 F Street, NW
Washington, DC 20004
202-320-0963

Commenter: International Collaboration for Clinical Genomics, Erin Rooney Riggs

Date of comment: 11/20/2013 14:00

Comment:

Section II. Scope and Applicability: The draft states the policy would apply to many different types of research, examples of which were provided in Appendix A. The policy does not specify, however, its stance on sharing of genomic data from categories not specified. ICCG is an organization that has received NIH-funding to collect and share genomic data primarily from clinical genetic testing laboratories as part of NHGRI's ClinGen Resource Program. Through this initiative, we will be submitting data of the types outlined in Appendix A (such as the results of cytogenomic microarray testing and, eventually, whole exome and whole genome sequencing) to NCBI, but we also intend to submit data that is not specifically addressed in Appendix A. Examples of such data include results from SINGLE gene sequencing tests, likely in more than 100 individuals for any given gene. Is the list in Appendix A simply meant to represent a list of things that should DEFINITELY be submitted to public data repositories under the GDS policy, with anything falling outside of this list being considered optional? Will NIH data repositories accept data that is not of a category defined in this list? If a data type, such as the results of single gene sequencing tests, is not included on this list, does that mean that submission of this type of data is not subject to the rules and regulations outlined in this document? Additionally, the draft references consequences for those not abiding by the provisions set forth in the policy. The language used implies that compliance is only a condition for those receiving Notices of Award or Contract Awards, and specified withholding of funding as a potential consequence. We agree that there should be serious consequences for those that do not abide by the provisions agreed upon in the final policy. We suggest that perhaps this section be broadened to include those requesting controlled access to the data; though subsequent sections outline the various terms data requestors must agree to, they do not elaborate the repercussions of violating these agreements, nor how far any potential repercussions may extend. Based on discussions taking place on the public webinar outlining this draft policy, it was suggested that violating the terms of the data access agreement may result in loss of data use privileges for both the individual and his/her collaborators, but this is not clear in the draft as it is currently written. The discussion of these consequences in this document should reflect the serious nature of such violations.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: This proposal seems to be constructed with traditional research in mind (subjects go through a consent process to participate in specific research). What is unclear is whether this policy applies to clinical practice, or research data collected through clinical practice. As stated in the "Purpose" section of this policy, data sharing is "essential to…improve human health." It is our opinion that data sharing not be limited to data obtained in the traditional research setting. Clinical laboratories are the largest generators of genomic data; hundreds of thousands of genetic tests are done each year in clinical labs, and the NIH has funded projects, such as the ClinGen Resource, dedicated to harnessing this data and

using it to elucidate ties between genomic variation and human health. ICCG and ClinGen are charged with supporting data submission from clinical labs to NCBI databases. The patient data collected here is not generated in the research context. Clinical labs often have no relationship with the patient or access to contact information, and do not have the same opportunities as researchers to obtain consent. Consent for clinical testing falls to the ordering physician. While consent is encouraged, it has not historically been required, particularly as laws largely do not require consent for diagnostic testing. Many clinical labs do not have the resources to track the thousands of samples they receive each year to ensure the clinician has obtained consent. ICCG supports the sentiment from the Policy Overview that "protection of research participant privacy … is paramount." While we feel the policy is clear on procedures appropriate for traditional research settings, we feel this is difficult for clinical labs to uphold. We fear that, should clinical labs be held to the same regulations as traditional research, their ability to share data will be significantly reduced, compromising a valuable genomic resource. We request that the policy A) be amended to outline specific provisions for maintaining the privacy of samples obtained clinically or B) be amended to state that this policy is not meant to cover the sharing of data from clinical samples, and to provide separate guidance for clinical data. This policy will guide IRBs reviewing genomics studies, and we want to ensure they evaluate those involving deidentified clinical data sharing in the proper context. We ask that the "opt-out" notification process be considered as a mechanism for deidentified clinical data sharing. This has been used by the Collaboration Education and Test Translation (CETT) Program and the International Standards for Cytogenomic Arrays (ISCA) consortium (predecessor to ICCG) successfully in the past. It allows the clinical lab to share their deidentified data sharing practices in places seen by the patient (the lab's website and educational materials, test requisition, the patient's clinical test result). Contact information for a lab representative is provided. The patient can contact them to learn more or opt out of having data shared. We ask that the policymakers consider alternatives such as these, and the idea that different methods of consent/notification may be appropriate for different types of data. For example, opt-out could be viewed as sufficient for "lower risk" data, such as variants identified in disease-targeted tests, with full consent being the gold standard for tests with the potential to return thousands of variants per individual, such as clinical whole exome or genome sequencing.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Just as with the proposed procedures for submitting genomic data, it appears that the procedures for accessing and using data are also geared toward traditional research applications. Although we feel these proposed procedures are appropriate for research applications, to achieve the stated goal of "facilitate[ing] knowledge…to improve human health," we also believe the data must be accessible to the clinical community. Clinical users (physicians, genetic counselors, clinical genetic testing laboratories, etc.) may want to use the data differently than researchers, particularly the data coming from clinical laboratories as supported by the new ClinGen Program. These users may only want to look at information on a particular variant for use in a single clinical case. There may not be a specific research question, and there may not be the need to have full access to an entire genomic data set. Procedures such as having to have an IRB-approved protocol, having to apply for access through a DAC, requiring high-level institutional sign-off, etc., would effectively prohibit individuals involved in clinical care from being able to use small, discrete pieces of data to aid in clinical decision-making.We request that the policymakers consider alternate access for intermittent clinical use of the data, one with

protections and requirements in place that are more in line with this type of limited scope usage. ICCG has previously submitted suggestions to NHGRI and NCBI staff regarding ideas for this new level of access. If this clinical access level is seriously being considered, we ask that this document specify that provisions for clinical use are forthcoming and may differ from those for research use as outlined in the document.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY

Website: http://geneticepi.org
email: iges@geneticepi.org

Position Statement of the International Genetic Epidemiology Society

in response to

"Draft NIH Genomic Data Sharing Policy Request for Public Comments"

The International Genetic Epidemiology Society (IGES) has long been an advocate of the sharing of scientific knowledge, and the sharing of data from large genetic epidemiological studies where this is appropriate. IGES also recognizes that the potential advantages of such data sharing must be balanced against various scientific concerns, as well as the critical need to protect the confidentiality of the participants in the studies for which data are shared.

Here we address the sections of the policy that give rise to specific concerns about the proposed policy.

Comment 1: Section II Scope and Applicability

It is the position of the Society that the extremely broad scope and applicability of the proposed policy presents a number of problems which will negatively impact genetic epidemiological studies in the US and international collaborations. Linking the funding of science to the ability of researchers to obtain consent for very broad data sharing presents a model going forward where years of investment in sample collections and making good faith agreements with local communities could be destroyed. In addition, this connection to funding brings with it significant potential for financial conflicts of interest, as researchers are forced to balance the need for funds with the ethical considerations of the communities they study.

The policy hurts existing and future potential collaborations by reducing the incentive to form these large, collaborative groups in the first place. International collaborators gain little from forming formal collaborations with US groups, when they can simply wait and download the data after the short embargo has passed. But US researchers have no such access to data from other countries, which imbalances the negotiations for collaboration from the beginning.

Comment 2: Section IV.A. Data Sharing Plans

The Society is highly supportive of the concept of data sharing and believes that collaborations between groups of scientists is the most effective way to move genetic epidemiological studies forward, in this era of increasingly large sample size requirements for adequate statistical power. However, we are concerned that this policy discourages the collaborative nature of science that is essential to making significant advances against some of the most common diseases affecting Americans today. In addition, the resources required to support data sharing in many projects may make the study unfeasibly expensive, especially where extensive re-consenting of participants is required. Also, acquiring re-consents from former participants to the study may pose a significant concern prohibiting further participation in biomedical research studies.

The ethical implications for consenting study participants for such broad data sharing means that many studies may not be able to proceed because participants are not willing to consent to such broad data sharing, or that in the case of historical sample collections, participants may have died and it is therefore the decision of individual IRBs whether those samples can be used without consent. There are discrepancies in the IRB decision making processes among research sites in the United States; therefore, investigators in the institutions implementing rigorous IRB criteria will always be at a disadvantage compared to others. In many foreign countries, the use of samples from deceased individuals is not permitted without previous consent before death.

Comment 4: Section IV.C. Human Genomic Data

1. Data Submission Expectations and Timeline

The unrealistically short time scales for data submission and embargo lifting disproportionately impact small and moderately sized labs and research groups, that do not have the resources to analyze the data that they have invested significant time in funding and collecting, in the time allocated. Small labs will struggle to perform all of the necessary quality control and data analysis inside of six months (which we note is a maximum time rather than a fixed limit) whereas only the largest labs with more people and computational resources can easily turn around data downloaded from dbGaP and "scoop" the researchers who invested all the work in sample collection and data generation. This decision is unfair and creates more challenges for small research laboratories/investigators.

2. Data Repositories

Although ostensibly the policy acknowledges that the NIH-designated data repositories need not be the exclusive source for facilitating the sharing of genomic data, investigators who elect to submit data to non-NIH-designated repositories are expected to confirm that appropriate data security, confidentiality and privacy measures are in place. It is unrealistic to expect that most labs will be able to set up their own repositories and so the only groups able to do this would be very large, commercially funded enterprises such as the Kaiser Foundation or 23andMe. Researchers working in universities across the country are unlikely in most cases to have the infrastructure required to support such a massive undertaking. For those labs, data submission to dbGaP "….no later than the time that data cleaning and quality control measures begin." will be an arduous task.

3. Tiered System for the Distribution of Human Data

Making all NIH-funded research involving large-scale genomic data available to third parties increases the risk of published results from these data being misinterpreted and/or misrepresented, especially in situations where the original data collection efforts are poorly documented. This is true even in controlled-access data situations, where secondary investigators only have to obtain initial approval from the NIH for data use, with no later assurance that analyses of these data without full knowledge of the data collection methods will lead to erroneous conclusions that are promulgated in the scientific literature.

4. Informed Consent

We responded to the proposed data sharing policy for GWAS back in 2006 and a number of the concerns we raised then not only remain unaddressed but are in fact more pressing than ever as the depth of genomic data covered by the new proposed policy has been significantly increased. In 2006 we expressed the concern that deidentified data with such deep genotyping was not truly deidentified since genotypes themselves could in fact be identifiers identifiers in association with some other identifiable information available publicly. In fact, a number of articles in the media addressed a paper published in Science in January 2013 that was able to identify a number of individuals from the Center for Study of Human Polymorphisms (CEPH) family collection whose genomes were sequenced as part of the 1000

Genomes Project (Melissa Gymrek et al., Science 339, 321 (2013)). Deposition of data into U.S. government databases also carries the risk that U.S. Federal Law Enforcement agencies can legally search those databases without a court-ordered subpoena, whereas a subpoena is required for those agencies to obtain access to data stored in non-Federal databases. The policy still does not address the concerns about storing biometric identifiers of non-citizens in U.S. Federal databases, which may deleteriously affect international collaborations. Recent events have turned the spotlight on how U.S. Law Enforcement agencies have conducted their activities and this data sharing policy does nothing to assuage the concerns of researchers to whom it applies.

We also believe that it may be impossible to obtain truly informed consent under this model of data sharing, as it is impossible to fully quantify the risks presented to participants if their data were to be deidentified. We cannot in good faith promise that their data will remain anonymous. Many participants may not be willing to accept these risks, and those that are willing to consent may not fully understand what they are consenting to as we ourselves cannot predict all the consequences of such broad data sharing. . In addition, a major portion of disadvantaged population participating in biomedical research studies and a considerable portion of the general population may not have the necessary educational background to understand the informed consent to the extent where all consequences of such broad data sharing will be understood. Therefore, an ethical concern remains: to what extent we are using the term 'informed consent' administered to the subjects to understand such implications to participate in biomedical science research. Although a *Certificate of Confidentiality* has been mentioned, the Genetic Information Nondiscrimination Act (GINA) is only applicable to a business with 15 or more employees (http://www.eeoc.gov/laws/types/genetic.cfm).

7. Exceptions to Data Submission Expectations

Although the NIH acknowledges that in some cases circumstances beyond the control of investigators may preclude submission of data to NIH-designated data repositories, the section of the policy describing exemptions seems both incoherent and potentially damaging to international collaborations. Section IV.C.2. suggests that investigators are not necessarily to be forced to submit their data to NIH-designated databases; however Section IV.C.7 seems to suggest that in fact investigators will not be able to simply "elect" to submit their data to non-NIH-designated data repositories but will have to justify this as an "exception". These two positions seem inherently contradictory. In addition, no recognition is made for situations where because of insufficient consent or legal requirements, data are not permitted to be shared at all. For researchers with long-standing data collections, this requirement to share essentially forbids them from being able to apply for federal funds to conduct their research. The NIH holds a unique position in US research in its ability to fund research that is too expensive or too high risk for private enterprises to be able to fund. Cutting off researchers who cannot comply with broad data sharing policies has already caused many long-running epidemiological studies to refrain from certain types of research because they are unable to re-consent their participants. This new proposed policy will only make that problem worse. There is a considerable danger of wasting the investment in past resources and existing long-term follow-up studies. Many collections with a large and esteemed track record, and at least partially NIH-funded in the past, will now be unable to move forward.

Comment 5: Section V. Responsibilities of Investigators Accessing and Using Genomic Data

The concerns of the community about how the NIH would police the use of data granted under dbGaP have already been realized through some high-profile cases of breaking embargoes. In one particular case which has been covered extensively, a paper was retracted after the embargo breach was reported to the journal in question (PNAS) and the individual was sanctioned by suspending the investigators access to dbGaP and all work with the downloaded data was to be ceased. However, the breach did not seem to heavily impact the career of the researcher responsible, Dr Zhang is still employed by Yale and continues to receive NIH funds. And despite the retraction, it is still possible to find the paper online, albeit with a tag labeled "See Retraction Published September 9, 2009.". Therefore, we believe that the

penalties for breaking data embargoes are poorly defined and clearly insufficient in the light of this case. It is essential that these policies be reviewed and strengthened in the new data sharing policy.

In summary, it is the position of IGES that the proposed policy presents a number of problems and challenges for researchers and the structure of the policy (in particular the very short embargo limits) disadvantages smaller research groups in favor of the very largest institutions. It undermines the formation of national and international collaborations and fails to adequately protect participants or prior research investments. It is our recommendation that these issues be examined in more detail, and that substantial revisions ought to be made before adoption.

Claire L. Simpson, Ph. D.
Chair of the IGES ELSI Committee
Dr. Simpson is serving in her own capacity

Prof. Dr. Andreas Ziegler
IGES President

2013 IGES Officers

President:
Andreas Ziegler
University of Lübeck
Lübeck, Germany

Past-President:
Suzanne Leal
Baylor College of Medicine
Houston, TX, USA

President-Elect:
Alexander F. Wilson
NIH/NHGRI
Baltimore, MD, USA
Dr. Wilson is serving in his own capacity

Treasurer:
Mariza de Andrade
Mayo Clinic
Rochester, MN, USA

Secretary:
Heike Bickeböller
University of Göttingen
Göttingen, Germany

Editor-in-Chief, Genetic Epidemiology
Sanjay Shete
U Texas/MD Anderson Cancer Center
Houston, TX, USA

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

DEPARTMENT OF HEALTH & HUMAN SERVICES Public Health Service

National Institutes of Health
National Institute of
Environmental Health Sciences
P. O. Box 12233
Research Triangle Park, NC 27709

November 20, 2013

Dear Committee on Genomic Data Sharing Policy:

I am writing to affirm that the attached comments regarding the draft Genomic Data Sharing (GDS) policy are reflective of the concerns and needs of intramural investigators at the National Institute of Environmental Health Sciences, National Institutes of Health (NIEHS/NIH) including the authors of the comments (who have expertise in bioinformatics, epigenomics and cancer mutagenesis/genomics) and the broader NIEHS research community.

A GDS policy that balances the goals of expeditious data sharing particularly for data generated by large consortia with the needs for robust data curation, metadata association, and the usual practices of investigator driven science should be the primary goal. The enclosed comments recommend such a balanced data sharing vision and associated policy.

Thank you for the opportunity to provide comments on this draft policy.

Sincerely,

Darryl C. Zeldin MD

Darryl C. Zeldin, M.D.
Scientific Director

NIEHS Division of Intramural Research draft comments to be submitted in response to "Draft NIH Genomic Data Sharing Policy Request for Public Comments" *Federal Register* Notice.

Developed by David Fargo[1], Dmitry Gordenin[2], and Paul Wade[3],

[1]Director of NIEHS Integrative Bioinformatics
[2]Senior Associate Scientist, NIEHS
[3]Director of Epigenomics Core Facility (all DIR NextGen Sequencing activities) and Senior Investigator, Eukaryotic Transcriptional Regulation Group, Laboratory of Molecular Carcinogenesis.

Note: in reference to our comments, please note that there could be minor changes in developing final version.

Comment 1:

Section II, Scope and Applicability.

This section contains a definition of what types of research should be covered by the policy. It would benefit the major goals of this policy to make this definition more precise and less inclusive.

The section stipulates that "This policy applies to all NIH-funded research that involves large-scale human and nonhuman genomic data..." The critical term here is large-scale. The document then refers to Appendix A wherein examples of the types of research to which the policy may be applicable are given.

In **Appendix A, Examples of Types of Research Covered Under the GDS Policy**, the following examples are provided:

- Sequence data from tens of isolates from infectious organisms
- Sequencing more than one gene or gene-sized region in more than 100 participants
- More than 10,000 genes or regions from one participant (e.g. whole genome sequencing)
- More than 100,000 variant sites in more than 100 participants

It is our considered opinion that the definition of large-scale, and the examples given, may benefit from additional detail. In particular, bullet three will likely result in inclusion of an exceedingly large number of data types – some of which apparently go beyond the intent of this policy. For example, the hg19 build of the human genome contains approximately 500,000 exons – each of which could (should) be considered a 'genomic region'. Using the example provided for 'large-scale' – *more than 10,000 genes or regions from one participant* – a single RNA-seq experiment from a human cell line (which would be expected to contain sequence data from 100's of thousands of exons – *genomic regions*) would be considered 'large-scale' and subject to all the criteria outlined in the GDS Policy. Likewise, an overwhelming majority of single experiments utilizing NextGeneration Sequencing to analyze human nucleic acids selected in any manner (e.g. by chromatin immunoprecipitation, nuclease accessibility, RNA immunoprecipitation, total RNA, miRNA, etc.) would fall within the definition of 'large-scale'. We

respectfully suggest that this definition may be overly inclusive and may not accurately reflect the intent of the policy. While we appreciate need to balance public access to data with investigator interests, the definition proposed here would likely benefit from careful consideration. As written, literally hundreds to thousands of simple experimental datasets on limited numbers of cell lines or subjects assembled by individual investigators would fall within the definition of large-scale.

Based on above we suggest the following addition to the paragraph **"Examples of Types of Research Covered Under the GDS Policy"**

Examples of Types of Research NOT Covered Under GDS Policy

> Exploratory experiments utilizing human cell lines, or fewer than x participants (we suggest 10 participants)
>
> Preliminary experiments utilizing fewer than x participants (we suggest 10 participants)
>
> Experiments involving biochemical fractionation of the human genome (for example by chromatin immunoprecipitation, nuclease digestion, and other related techniques) in fewer than x participants (we suggest 10 participants)
>
> Experiments sequencing total or fractionated human RNA from fewer than x participants (we suggest 10 participants)

Comment 2:

Section IV.A. Data Sharing Plans

none

Comment 3:

Section IV.B. Non-human and Model Organism Genomic Data

none

Comment 4:

Section IV.C. Human Genomic Data

The data release policies, as outlined in Section IV, Subsection C, Part 1 Data Submission Expectations and Timeline, formulate a very aggressive timeline for full public access. It is our considered opinion

that the timelines outlined in this policy draft would likely have significant negative impact on individual investigators.

The GDS Policy Draft stipulates that "NIH will release data submitted to NIH-designated data repositories without restrictions on publication or dissemination no later than six months after the initial data submission to an NIH-designated data repository, or at the time of acceptance of the first publication, whichever occurs first." Appendix A in Expectations for Data Submission and Data Release outline an expectation for data submission for Level 2 data as "Project specific, generally within 3 months after data generation." Level 2 data (defined as "Data after an initial round of analysis or computation to clean the data and assess basic quality measures") are often generated very early in the data pipeline for most investigators. Generation of a BAM file from raw NextGeneration sequencing data would seem to fit this definition. The draft policy, as currently written, stipulates that a single experiment generating a single lane or portion of a lane of NextGeneration sequencing data must be submitted within 90 days of generation of a BAM file, it will be publicly released within 6 months of submission regardless of whether a manuscript is submitted. This time frame seems quite aggressive, as applied to individual investigators performing what are now routine studies. These time constraints seem appropriate for large, publicly funded data collection studies (e.g. ENCODE, Roadmap Epigenome, TCGA, etc.). In turn, they seem to limit the available time for an individual investigator to analyze their data fully, to carefully consider the biological implications of their data, to formulate a manuscript describing their findings, and to shepherd such a manuscript through the publication process prior to data release.

Based on the above we suggest that the expected timeframes for all levels of the data for the NIH funded research covered by the GDS policy would be modified as follows:

Data submission: "No later than the end of the period of the project funding or at the time of publication submission to the Journal, whatever comes first"

Data release: "No later than one year after the end of the project funding or at the time of publication by the Journal, whatever comes first"

Suggested additional note relating to large scale projects such as TCGA, ENCODE etc.: "Large scale project defined as such by NIH panels and staff will have special timeframes for submission of each data level established by NIH including a case-by-case process."

Data release policy should be designed to maximize general scientific utility including high or well defined quality standards and appropriate and useful metadata descriptors. Aggressive or inflexible release policy may contaminate data quality and diminish overall utility.

Comment 5:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data
Comment to this section also relate to Level 3 and Level 4 description Appendix A

This section lacks description of important responsibilities for properly reporting the results of analyses based on the use of large genomic datasets. The current model in many publications is to only refer to the location of the raw data (Level 1, e.g. SRA files) or to the data resulting from the first level of analysis (Level 2, e.g., BAM files). However, it is the next step of analysis, i.e., Level 3 and/or Level 4 data that are used to generate conclusions and display items in manuscripts. Level 3 and 4 data should be made available to other investigators and to manuscript peer reviewers. Such data should be formatted in a manner, and should include necessary content, to empower (1) independent verification of conclusions and display items, (2) exploration of hypotheses proposed within the published research and (3) exploration of new hypotheses. If such data contains elements requiring controlled access the content should be described in a publication, e.g. by the annotated listing of column names in the dataset organized in a table. This will enable others to understand whether it is appropriate and useful to apply for controlled access as well as reduce the workload on depositary (e.g. dbGaP) personnel. In investigations that do not require controlled access, all data used for generating display items should be included in supplementary information and submitted to public. These Level 3 and 4 data should include appropriate and useful metadata.

A challenge arises for sharing when Level 3 and or Level 4 data that must be retained under controlled access are produced not by investigator(s) whom also produced underlying Level 1 and/or Level 2 data. Such data often represent explorations of novel hypotheses and/or global analyses of very large datasets, (such as TCGA) with a potential for re-use by many scientists. Until recently there has been no way to either directly share such data or to submit them back to NIH databases with controlled access. Currently dbGaP allowed a pilot submission of such "secondary" Level 3 and/or Level 4 datasets (dbGaP study phs000677.v1.p1). Alternatively, there could be a simple mechanism of direct sharing such Level 3 and or Level 4 data with other investigators having controlled access to those Level 1 and/or Level 2 that were used to generate secondary Level 3 and Level 4 data. As of now, the difficulty in authorized sharing of secondary Level 3 and Level 4 is prohibitive.

We recommend developing a Subsection C. "Responsibilities of Proper Reporting" with the following text:

If the datasets from NIH-designated data repositories were used for producing conclusions or display items in a publication, there should be clear identifiers in a special section of Material and Methods and/or in other appropriate places of the manuscript to all data levels (see Appendix A) that were directly used for generating each conclusions and/or display item. These pointers should refer to a specific supplementary item or to a location in a database, including specific filename or other identifier allowing unambiguous identification for reproduction and/or re-use and should be provided at the time of manuscript submission or other data release.

Note that the suggested requirement closely follows the requirement put forward by Nature Publishing Group (http://www.nature.com/authors/policies/availability.html)- quote:

"Datasets must be made freely available to readers from the date of publication, and must be provided to editors and peer-reviewers at submission, for the purposes of evaluating the manuscript."

Comment 6:

Section VI. Intellectual Property

none

Comment 7:

Any other aspect of the draft GDS Policy

1. Comment to: Overview of the policy:

The first paragraph of this Section starting with "The draft GDS Policy describes..." misses important parts of actual content of the document and leaves an impression of very narrow Policy goals. Overview of the policy should also include a statement about description of NIH resources (such as repositories described in IV-A-2 and IV-C-2) and setting the stage to promote, facilitate, and assist in sharing being the goal of the Policy.

2. Comments to Appendix A:

Comments to the section Expectations for Data Submission and Data Release

Comment 1

Table in Appendix A, Level 1.
The Draft states that it is "Not expected for human data if reads are included in Level 2 aligned sequence file (e.g., BAM)."

Suggested addition:
"...BAM files should include all generated reads and not be limited to reads with successful alignment".

Rationale:
Importantly, in many cases alignment is made against the fraction of the genome defined by a study-specific hypothesis, so unaligned reads can be of great value for exploring other hypotheses. For example: RNA-seq data could be aligned against non-repeated genes, while the sequence data will also contain reads matching polyA-transcripts from repeated elements and integrated viruses.

This comment also relates to below suggestions to Level 2 in the Table and to descriptions of Level 1 and Level 2 in the text below the Table

Table in Appendix A, Level 2
Supplement words "DNA sequence alignments to a reference sequence..." with words "... including non-aligned reads" (See comment to Level 1 above).

Level 0 and level 1 – text below the Table:
Add "only if unaligned reads are listed in the BAM files" on line 2.

Level 2 – text below the Table:
Modify the sentence "A submission would be a file (e.g., binary alignment/map (BAM) files) USUALLY containing the unmapped reads as well," to replace USUALLY with WHICH SHOLD ALSO CONTAIN

3. Comment to "References:"

The web link in reference 49 "Research Tools Policy. See
http://www.ott.nih.gov/policy/research_tool.aspx
does not work. Another location not mentioned in the Draft might be relevant:
http://www.gpo.gov/fdsys/pkg/FR-1999-12-23/pdf/99-33292.pdf
- a link to 1999 NIH policy for SHARING BIOMEDICAL RESEARCH RESOURCES

4. General comments:

Science faces a growing challenge to develop ethics, rules and policies for sharing and dissemination of unprecedented amounts of data. This is especially true for datasets produced using modern genomics technologies. Modern science has already developed traditions, requirements and ethics for sharing unique materials and reagents underlying a publication. These have evolved to their current state over at least a century. In many cases genomics datasets can be likened to unique materials and reagents produced in traditional experimental research in that the access to the data is required for reproduction of conclusions of a published study as well as for conducting a new study addressing a different hypothesis. While applying the patterns developed for sharing of experimental materials appears to be a productive and time saving approach, sharing of large data will require careful thinking and trial-and-error approach in policy development. Developing helpful and balanced policy may not require another century of evolution; however the 60 day period for public comment on the current policy draft is clearly insufficient to acquaint the scientific community as well as the general public with the policy and to generate high quality insightful comments. Considering this, it is reasonable to state in the policy text the period over which comments to the acting version of the policy would be collected to be considered for its periodic update. It would be reasonable if the existing GDS website (http://gds.nih.gov/) would provide the forum for collecting such comments. Such a forum could help to refine the areas and mechanisms covered by this draft, as well as to bring up questions that would be natural for inclusion into NIH Policy, but currently underdeveloped in the Draft.

One group of such questions includes the interaction between NIH and scientific journals in setting standards and mechanisms to facilitate all levels of genomic data sharing. Such mechanisms could develop using PubMed Central as a model. Similar to a standard Data Sharing Plan, NIH could develop a standard Genomics Data Sharing Checklist, which journals may include in their publishing policy as a separate checklist or as a part of more general checklist (see e.g., Nature checklist and policy at
http://www.nature.com/authors/policies/checklist.pdf
and

http://www.nature.com/authors/policies/availability.html
If the Journal does not develop an NIH compliant checklist, the expectation for the NIH funded research could be to include a standard NIH checklist into supplemental section or if the Journal does not accept such an addendum to attach the checklist to PubMed Central version of the paper.

Another problem that should be addressed in future development of genomics data sharing standards is the access of journal editors and reviewers to the data protected by controlled access, when such data are used to produce conclusions and/or display items in a submitted manuscript. Currently there is no mechanism for such access within the short timeframes of editorial consideration and peer review. This lack of detailed peer review could lead to decrease in quality of published research and ultimately to erosion of public trust in research based on controlled access human genomic data. Altogether, inclusion of Journals in facilitation of data sharing policy would mobilize the self-regulating capacity of the scientific community through peer review and post peer review follow up.

Kaiser Foundation Health Plan
Program Offices

November 20, 2013

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

Submitted via email to GDS@mail.nih.gov and fax (301-496-9839

RE: *Draft NIH Genomic Data Sharing Policy Request for Comments*

Dear Sir or Madam:

Kaiser Permanente offers the following responses to the Request for Comment on the *Draft NIH Genomic Data Sharing Policy Request for Comments* ("Draft Policy") published on September 20, 2013 in the Federal Register.[1] We appreciate the opportunity to provide our feedback for your consideration.

The Kaiser Permanente Medical Care Program is the largest integrated healthcare delivery system in the U.S. and is committed to delivering the highest quality health care to over 9 million members in nine states and the District of Columbia.[2]

Research is a hallmark of Kaiser Permanente and one of the ways we demonstrate our commitment to care transformation within communities across our seven regions. Kaiser Permanente conducts a broad agenda of both health services and clinical trials research throughout its healthcare system. A number of Kaiser Permanente research centers conduct genomics research. For example, the Northern California Division of Research ("DOR"), leads

[1] 78 Fed.Reg. 57860

[2] Kaiser Permanente comprises Kaiser Foundation Health Plan, Inc., the nation's largest not-for-profit health plan, and its health plan subsidiaries outside California and Hawaii; the not-for-profit Kaiser Foundation Hospitals, which operates 37 hospitals and over 600 other clinical facilities; and the Permanente Medical Groups, independent physician group practices that contract with Kaiser Foundation Health Plan to meet the health needs of Kaiser Permanente's members. Kaiser Permanente also includes the Permanente Dental Associates, a multispecialty dental group, in the Northwest.

One Kaiser Plaza, 27th Floor
Oakland, CA 94612

the *Research Program on Genes, Environment and Health* ("RPGEH"), KP's largest study involving genomics, which identifies genetic and environmental factors that affect health

We support the NIH's goal to advance genome-wide association studies (GWAS) that identify common genetic factors that influence health and disease, and we also support the principle to make available to scientific investigators the genotype and phenotype datasets to achieve this goal.

Kaiser Permanente is an institution that would consider submitting member genotype and phenotype information data. We are very mindful of our responsibility to protect our members who choose to participate and our accountability, under the Draft Policy, to minimize risks to their privacy and confidentiality.

We acknowledge the need to assure the protection of research participants and to provide for clear, transparent informed consent and data access processes that will both encourage participation in GWAS and assure participants that the NIH and submitting institutions intend to protect the privacy and confidentiality of this sensitive information. The submitting institutions, would bear the primary responsibility for: the informed consent process; reconsenting, where applicable; institutional written certification that the data has been deidentified according to all applicable laws and policies; IRB/Privacy Board review of all submissions of data; and a summary of limitations on data use based on the informed consents.

Our members trust that Kaiser Permanente will be good stewards of their data. Our responsibility as stewards is to ascertain that there are appropriate oversight and review processes to maintain that trust. As a submitting institution, Kaiser Permanente proposes the following comments and recommendations on the Draft Policy that will enhance our ability to protect our members who choose to participate and will establish a process for the submitting institutions' review of requests for access to their entity's data.

I. Human Subjects Protection

Under the Draft Policy, the institution, IRB, investigator and individual participants all bear significant responsibilities for assuring protections, to the degree possible, of the individual participant's privacy and confidentiality. The Draft Policy requires the institution, IRB, and investigator to assure that all submitted data are de-identified and that the informed consent describes all potential risks and delineates all possible future uses and disclosures of the GWAS data. The consent language would control data access by public or private entities. However, due to the complex and evolving environment of genomics research; the sensitive nature of the data; the potential risks to an individual's privacy and confidentiality; and negligible immediate benefit to the individual participant, the protection of human subjects should be paramount. This protection can best be achieved if the informed consent process is clear, concise, and consistent across submitting institutions.

Because of these requirements, data submission to NIH may increase burdens on institutions, investigators and IRBs, as well as on the participant who is expected to fully understand the potential uses and disclosures of her data and provide consent.

Recommendation:

To assure that the informed consent process is adequate, clear, and transparent for participants, Kaiser Permanente recommends that NIH provide guidance and specific language that is required for all informed consent documents. NIH should ascertain that the required informed consent language has been appropriately vetted and tested with representative population groups to assure that the consent uses clear, simple language and that individual participants have a high likelihood of understanding both the societal benefits of participation and the potential risks.

II. Access to Data

Under the Draft Policy, access to the GWAS data base will be either through the public/open portal or through a controlled-access portal. There would be no charge for access, even if the recipient entity is private or for-profit and has not submitted data. If the entity requesting data access is not a submitting entity, it bears limited responsibility for protecting participant data and bears no costs associated with data submission requirements.

The NIH's Data Access Committee ("DAC") reviews and approves all requests for access in conformance with submitting data use limitations established by Institutional Certification and informed consent. However, the submitting institution does not participate on the DAC nor does it have an opportunity to know and anticipate in advance all possible requests for uses and disclosures of data. Conceivably, the DAC could unintentionally provide access to an entity that does not meet the submitting institution's data use limitations because of the complex nature of this research, rapidly evolving landscape, and inconsistent or variable understanding of the science.

Although the submitting institution is responsible for meeting all the requirements for data submission, under the Draft Policy it cannot review requests for data access. The Data Use Certification, signed by the recipient of the data, is an agreement solely between the recipient institution and NIH. This arrangement does not recognize the responsibility of a submitting institution to protect its interests and patients' data. However, in the event of violations of the GDS policy that resulted in an unapproved use or disclosure, it is likely that the submitting institution could bear responsibility for the violation, in the eyes of its members, and thus risk its members' trust and willingness to participate in future research through GWAS or other similar national research repository efforts.

We believe that access by entities, particularly for-profit institutions who are not submitting data, could be a major deterrent to participants in consenting to submit their data to GWAS. Consenting to future, unspecified research may also be a significant deterrent if participants think that unidentified, third-party entities might use their data for unknown purposes outside their control or knowledge. While there can be undeniable benefits in allowing for-profit entities to use data for innovative research, individual participants and organizations may determine that the risks and downsides outweigh any potential, long-term benefits.

Recommendation:

We recommend that NIH establish a process that allows the submitting institution's Institutional Official to review and approve all access requests by entities who are not themselves submitting individuals' phenotype or genotype data to GWAS. As the steward of the institution's data, the Institutional Official must assure that all proposed research uses of the institution's data are appropriate, allowed and ethical. If the access request raises ethical concerns or the proposed research project poses risks to the entity's data that outweigh the proposed benefits, then the entity's Institutional Official has the opportunity to deny access to the entity's data, unless the request is modified to address the concerns.

To provide for greater transparency, a quarterly record of all access requests and their disposition should be made available upon request to submitting institutions and members.

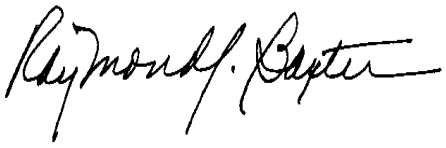
Submitting institutions should be immediately notified of any policy violations reported by the investigators and/or entities receiving the GWAS data.

Given the additional direct costs to submitting institutions, all NIH GWAS awards should recognize and cover the additional direct costs of the institution's compliance with the submission requirements. One method for covering those costs would be to charge a reasonable and appropriate fee for database access (to the controlled-access GWAS portal) by entities that have not borne any costs of data submission.

CONCLUSION

We appreciate your willingness to consider our comments on this RFC. Please feel free to contact VP of Research Karen Emmons with any questions or concerns: (510-625-4724); email: Karen.M.Emmons@kp.org.

Sincerely,



Raymond J. Baxter, PhD
Senior Vice President
Community Benefit, Research and Health Policy

Commenter: Dr Jane Kaye

Date of comment: 11/13/2013 5:46

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Dynamic Consent The Genomic Data Sharing Policy provides an excellent opportunity for the NIH to establish methods to engage with research participants, and to take a lead in engendering a truly patient-centred approach to genomic medicine. The consent process will provide a vital opportunity for engagement, to explain the intentions for research and to initiate a dialogue about the research process, and the participant's role.We believe that the proposal to adopt a "broad consent" approach undermines the NIH's focus on ensuring that participants are appropriately informed about the research to which they are contributing. Instead, we propose that NIH adopts a "dynamic consent" approach, which will enable participants to choose for themselves how engaged they would like to be in the project, will provide the opportunity for researchers to gain participant input as the research field develops, and will enable participants to receive information about the research that is being undertaken.The deficiencies of broad consent are considerable and well known: participants cannot be said to be making a genuinely informed decision particularly when increased sharing and secondary use of data make it impossible to inform participants about what research will be done using their genomic data in the future. Broad consent is effectively consent to have all the important decisions made by other people - its primary effect in practice is to marginalise and trivialise the involvement of donors in research. The primary argument for broad consent – that it protects researchers from having to engage in expensive, time-consuming re-contact and re-consent of participants - is limited by the fact that broad consent does not protect research from changes in law and regulation, from innovative new technologies that permit novel and un-anticipated uses of data or from changing demands from publics or policy makers. In an era of patient-centred innovation and increased public engagement with science, research that treats participants as 'subjects' rather than participants, and static paper-based consent models are seeming increasingly out-dated and looking unfit for purpose.Dynamic Consent (DC) is an alternative to broad consent which addresses the changing nature of biomedical research in ways that respect participants by actively producing research as an ongoing partnership between participants and researchers. To achieve this, dynamic consent uses information technology to place patients and research participants at the centre of decision–making. The DC model

involves a user interface (such as a website or an application for a mobile phone or tablet) that acts as a personalised communication portal, a source of information and a system for making and changing consent choices that will link directly to the information system where genomic and other health data are stored. DC interfaces can also be tailored to meet the needs of particular research infrastructures and projects, in terms of which consent options are practical and workable and determining the most appropriate information to provide to particular patient populations or members of the public. The primary advantages of employing a DC system are that it:• Recognises users' autonomy and allows their interaction to be tailored to meet their individual needs• Promotes scientific literacy as participants become more informed about the research carried out on their samples and information, which encourages public trust by making research more transparent and accountable. • Creates an online, responsive and highly engaged cohort of participants for researchers to contact about further studies or to collect additional types of information (e.g. self-reported health status, adverse drug events) to add new dimensions to current studies.• Allows research governance to respond to changes in law and regulation, new scientific techniques and capabilities, and changing social perspectives, by engaging with participants to discuss changes rather than making assumptions about what they 'would probably be comfortable with'.• Makes consent meaningful and allows nuanced consent choices that avoid the 'all or nothing' involvement of broad consent.• Meets the highest international ethical and legal standards for consent in a world where data protection laws are changing.In 2007, the NIH conducted a similar public consultation to gather comments relating to the policy for sharing of data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS). The Notice outlining the result of this exercise stated that the 'NIH recognizes that the ethical considerations relevant to GWAS data sharing are complex and dynamic'. Consent was a specific area of concern for respondents, with the Notice stating that efforts to address the complex nature of these issues would include 'discussion of the optimal methods for communicating with participants about relevant issues through the informed consent process for prospective studies'. It also conceded that '[t]he NIH anticipates that a number of GWAS proposing to include pre-existing data or samples may require additional consent of the research participants', providing a clear example of the difficulty of setting up a broad consent that adequately caters for future research developments. While this previous exercise specifically focused on GWAS studies, many of the concerns raised are directly applicable to the data sharing issues discussed in the Genomic Data Sharing Policy. Learning from this past experience, it would be beneficial to both the researchers and the participants to install a system of dynamic consent that enables these points to be addressed from the outset, providing a vehicle to cascade information to participants to discuss the complex and dynamic issues relating to data sharing, a remit to optimise the method of communicating with participants, and a tool by which researchers can re-contact participants quickly and easily, if new consent issues arise. The development of new policy is an opportunity to reflect on best practice and incorporate new ideas. We would strongly recommend the NIH to take advantage of this opportunity to consider the merits of nuanced dynamic consent over broad consent, on-going digital engagement over static paper-based systems and participant-centred research partnerships over passive human subjects research.The Dynamic Consent platform is being developed and implemented by a partnership between HW Communications Ltd, the London School of Economics and Political Science and the Centre for Health, Law and Emerging Technologies (HeLEX), University of Oxford in the UK. Key members are: Dr Jane Kaye, Director of the Centre for Health, Law and Emerging Technologies (HeLEX), Nuffield Department of Population Health, University of

OxfordDr Edgar A. Whitley, Department of Management, London School of Economics and Political ScienceDr David Lund, Head of Research and Development, HW Communications Ltd.Dr Michael Morrison, Centre for Health, Law and Emerging Technologies (HeLEX), Nuffield Department of Population Health, University of OxfordDr Harriet Teare, Centre for Health, Law and Emerging Technologies (HeLEX), Nuffield Department of Population Health, University of OxfordFor more information see: Dynamic Consent: http://www.publichealth.ox.ac.uk/helex/about/research-projects-1/dynamic-consent-project EnCoRe (the project that dynamic consent grew out of): http://www.publichealth.ox.ac.uk/helex/about/research-projects-1/encore-projectIndicative publications:Whitley EA (2013) Towards effective, consent based control of personal data. In Digital Enlightenment Forum Yearbook (O'Hara K, Hildebrandt M and Waidner M, Eds), pp 165-176, IOS Press, Amsterdam.Kaye J, Curren L, Anderson N et al (2012) From patients to partners: participant-centric initiatives in biomedical research. Nature Reviews Genetics 2012; 13: 371–376.Kanelloupoulou N, Kaye J, Whitley E, Creese S, Lund D, Hughes K (2011) Dynamic consent – a solution to a perennial problem? BMJ Rapid Response. Available online at URL< http://www.bmj.com/rapid-response/2011/11/08/re-broad-consent-informed-consent> [Accessed 23 October 2013]

Commenter: Joseph Kenary

Date of comment: 11/19/2013 13:12

Comment:

Section II. Scope and Applicability: I have a severely retarded 14-year-old granddaughter who is missing chromosomes 6 and 10. She is about 3 feet tall and walks with difficulty. At birth, the only visible manifestations of her condition were crossed eyes and a cleft palate. Both were surgically corrected at about age one. At birth, after several days of trying to have her suckle, it was abandoned because the milk or formula would exit her nose. A tube to her stomach was inserted until a surgical incision was made in her stomach for tube feeding with formula. My daughter and I have not discussed this in detail, but I wonder if this is why she has never spoken any words nor ingested any food -ever.

She has two very normal sisters, one older, one younger.

Section IV.A. Data Sharing Plans: This is an inquiry, and I am not sure where this might lead. I recall my daughter saying, in the months after my granddaughter's birth, that she was told there was little data available.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Sundeep Khosla

Date of comment: 11/15/2013 17:29

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: One point that raised at the last eMERGE consortium meeting was that there are no real consequences (i.e. negative consequences) articulated in the proposed policy for investigators who knowingly compromise confidentiality of individuals participating in genomic research. This is a point of concern that has been raised by a number of individuals over the last several years in the context of broad consent and wide data sharing.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Douglas P. Kiel

Date of comment: 11/12/2013 12:10

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: Study sections reviewers and SRA's should be educated about budgetary considerations required to prepare data for sharing such that grant applications are fairly evaluated for these important budgetary considerations

Section IV.B. Non-human and Model Organism Genomic Data:

Given the complexity of human genomic data and the time required to pursue functional validation of findings, a six month embargo on allowing publication by other users is too short. A one year embargo should be considered.

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Under the data submission and data release section, more careful thought should be given to the requirements for levels 3 and 4 release. Analysis of omic data often involves multiple approaches and generates many types of output that are unique to a given project. Requiring all analyses to be submitted, for example, "capturing gene expression patterns," will require additional budgetary support to employ analysts who will track different approaches, data output, etc. for submission under this rule. The table in the document regarding data submission lacks sufficient detail.

The problems with submission of level 3 data are magnified even more with level 4 data submission. The number of final analyses for a manuscript will again require excessive time and financial burden on investigators to track and upload the many statistical models, analysis approaches, etc. required for a manuscript. Further clarification of the extent of such a requirement is needed a the very least. Also the NIH must allow for budgetary increases to handle all of this data submission. The submission requirments of this level data should be reconsidered carefully.

Commenter: Thomas Kosten MD

Date of comment: 11/4/2013 13:28

Comment:

Section II. Scope and Applicability: no comment

Section IV.A. Data Sharing Plans: It should be made clear that investigators seeking NIH funding need not contact the appropriate officials if they are experienced with the submission of data for sharing.

Section IV.B. Non-human and Model Organism Genomic Data: no comment

Section IV.C. Human Genomic Data: The deadline for release of data of six months or at the time of acceptance of the first publication should be extended to one year after data submission. As it is the investigator who initiated and conducted the research, adequate time should be given for that investigator to analyze and publish their investigations.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: no comment

Section VI. Intellectual Property: no comment

Any other aspect of the draft GDS Policy: no comment

Commenter: Maria Teresa Landi

Date of comment: 11/20/2013 22:00

Comment:

Section II. Scope and Applicability: I am concerned for the amount of data that will be uploaded into dbGaP. This would require huge storage capacity, personnel and costs. Since there are no standardization guidelines for the data features, different types of data, and possibly even different versions of the same data (if further QC or analyses will modify the original version submitted) may be uploaded, with exponential complications. Only investigators at big Institutions with large computational capabilities may be able to sort through all the large amount of data. And most of all, this will require that investigators that analyze any genetic data (from just one gene in 100 people to large whole genome sequencing analyses in large populations) have the financial and personnel support to upload all the data into dbGaP. Or dbGaP personnel will be devoted to help coordinating this work? Moreover, a full-time DAC maybe required to go through and approve all the applications for each piece of data. It may be more appropriate to require to share only cleaned final versions of meaningfully large datasets with more stringent criteria for QC, format and reporting characteristics.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Neil McKenna

Date of comment: 11/13/2013 13:38

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Are expression microarray datasets included in the GDS policy? They are not specified anywhere in the policy. Expression microarrays remain very widely used for transcriptomic research and indeed given the vast amounts of data and relative lack of standards in the field for RNA-Seq data, microarrays are often used preferentially. I think restricting this policy to sequence data only is a mistake and will continue to allow a large amount of transcriptomic data to be lost to posterity. Please please please include expression microarray data in this policy.Thank you

Commenter: Neil McKenna

Date of comment: 11/15/2013 13:16

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: I would urge NIH to engage publishers as early as possible in the process of implementing this policy so that they can modify their Instructions To Authors accordingly. Doing so will accelerate the rate at which datasets reach the community and will also decrease the administrative burden on NIH in tracking down published datasets, which will be an extremely onerous task. Publishers are best-positioned to assist NIH in ensuring this policy achieves its intended aim of universal accessibility of NIH-funded datasets. Being prominent financial beneficiaries of NIH-sponsored research, this is something that publishers should be required to do as a service to the scientific community. Authors will do what they have to do to get a paper published and if they have to deposit a dataset, they will do so. If they don't, they won't, and the burden will fall on NIH administrative staff to follow up on undeposited datasets.

Comments on Draft NIH Genomic Data Sharing Policy

Name of Individuals Submitting Comments:
Michael Lenardo, Koneti Rao, Helen Su, Morgan Butrick

Commenting on behalf of:
Molecular Development of the Immune System Section, Laboratory of Immunology, NIAID, NIH
Bethesda, MD

Word Count: 2984

Our group of scientists and clinicians within the National Institute of Allergy and Infectious Disease at the NIH values this community-wide effort for open and efficient use of genomic data. We have extensive experience over the past 20 years identifying, understanding the mechanism, and finding new therapies for genetic diseases of the immune system. We are concerned that the proposed timeline for data release is too restrictive (i.e., up to 6 months after data submission or at the time of first publication, whichever occurs first). This concern stems from two sources: ability to complete functional studies related to genomic findings and obligation to disclose relevant findings to human subjects before sharing publicly.

It is increasingly required to demonstrate the functional effects of newly identified genetic mutations for publication. Functional studies to establish pathogenicity can include validation in both in vitro assays as well as animal models. These assays can take well over 12 months to complete. Requiring the release of genomic data before the functional studies are completed diminishes the possibility that the primary investigators, who have already invested considerable time and effort beyond the sequencing, will be able to publish their findings first in a peer reviewed journal. Further, recent research has shown that the incidence of mutations of deleterious cellular consequences is higher than expected. Initially, it was anticipated that deleterious mutations would be rare and have obvious effects on phenotype. However, Yue and colleagues (Amer. J. Hum. Genet. 91:2012) demonstrated that healthy volunteers carry 40–110 variants classified by the Human Gene Mutation Database as disease-causing mutations, 3–24 homozygous variants, and many polymorphisms putatively associated with disease. Therefore, every genome poses a puzzle that must be interpreted with judicious experimental validation beyond simple bioinformatic analysis. Such variation requires rigorous interpretation that takes an unspecified amount of time, as it may be medically actionable for our participants.

Relatedly, we are concerned that the timeline for data release will outpace the timelines for the functional studies and variant interpretation such that research participants' genomic data will be publicly available before consequential findings are replicated under CLIA conditions and have been discussed with the study participants. We feel that this threatens our ethical and moral obligation to respect and inform of our participants of potentially medical relevant information before that information is made public.

We suggest that the policy be revised to reflect these realities. The release of data following publication seems to be a reasonable requirement which has been used in the past, but the addition of an arbitrary release of data at 6 months in the absence of any published analysis of the data does not seem to be in the best interests of the participants, the investigators, the scientific community, or the public at large. While an extension of the timeline would lessen these concerns, we also suggest the committee considers alternatives. For example, the investigators who generated the data could use their discretion to allow data release if they are unable to publish and future or continued funding could be tied to compliance with data sharing, independent of any specific timeline. We appreciate the consideration of these concerns and proposed solutions.

Commenter: Lindsay Morton

Date of comment: 11/13/2013 16:13

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

In light of the variability in files that are generated from various platforms, what kind of centralized standardization and quality control of data submissions will be put in place?

The quality control process for use of these data is extensive, and some key aspects of data quality may not be revealed until more in-depth analyses are complete. How will it be handled if a key quality control issue is identified by an investigator after data submission? Related to the issue of quality control, the timing for the submission is unclear. On slide #19 from the webinar, the GDS policy indicates data should be submitted 3 months after generation and will be released 6 months after submission with no publication embargo. Presumably this only applies to level 2 & 3 data, because slide #25 indicates that level 4 data (relating genomic data to phenotype) should be submitted after analyses are complete, with data to be released at the same time as publication.

What resources are expected to be put in place to handle these requirements? The expansiveness of the policy suggests that substantial resources will be needed to support each aspect of the policy, from data submission to database management and subsequent release to qualified investigators, but the details of these resources are not clearly outlined. For example, webinar slide #18 says “Plans should include resources necessary to support sharing.” Specifics regarding such resources would be very helpful.)

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: In light of the dramatic expansion of the data that fall under the new data sharing requirements, can you clarify how the current Data Access Committees will be expanded/restructured? The current de-centralized structure may mean that individual DACs are ill-equipped to handle both the volume and diversity of data that will fall under their purview.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

National Congress of American Indians

Comments on
Draft NIH Genomic Data Sharing Policy

November 20, 2013

The National Congress of American Indians (NCAI) is the oldest and largest national organization representing the interests of American Indian and Alaska Native tribal governments in the United States. NCAI is a membership organization that serves the interests of the 566 federally-recognized tribes, state-recognized tribes, and American Indian and Alaska Native tribal citizens. As stated in the Preamble to the NCAI Constitution, NCAI serves:

> "to secure to ourselves and our descendants the rights and benefits [of] the traditional laws of our people to which we are entitled as sovereign nations; to enlighten the public toward the better understanding of the Indian people; to preserve rights under Indian treaties or agreements with the United States; to promote the common welfare of the American Indians and Alaska Natives."

As part of our work to affirm tribal sovereignty and secure our ability to continue to live as Native peoples, NCAI recognizes that research can add value to Native communities when it is driven by tribal leaders and developed in an ethically and meaningfully way. As such, NCAI established the National Congress of American Indians Policy Research Center (NCAI PRC) in 2003 to serve as a tribally-driven center, focusing solely on issues facing tribal communities. We assert that tribes have sovereignty over research that happens on their land and with their citizens and that research ethics must acknowledge the need to *both protect and benefit* Native people through research development.

NCAI advocates that all research conducted with American Indian and Alaska Native tribes and peoples should be developed in full consultation and in equal partnership with tribal leaders over the course of the entire research process, including: research design, data collection, data analysis, and reporting and dissemination. Tribal leaders have the best sense of what kinds of research and data would be most helpful to their citizens. Furthermore, given the diversity and uniqueness of American Indian and Alaska Native communities, the potential risks, benefits, and considerations related to participating in a research study will vary by tribe and by research study. For this reason, American Indian and Alaska Native individuals and tribes must have the opportunity to consent to participate in research in an informed and ethical way.

The NCAI PRC provides the resources and tools necessary to inform public policy debates with meaningful information and assist in shifting the discourse in Native policy from a problem-focused approach to truly proactive, future-thinking strategy development. The NCAI PRC's tribal research regulation work serves to support tribal leaders in ensuring research that is conducted on their lands and with their citizens is ethical, affirms tribal sovereignty, and contributes to community well-being. A major part of the work of the NCAI PRC has been to engage with tribal leaders and federal partners around data sharing and genetics research. For example, in September 2013, we launched the American Indian & Alaska Native Genetics

Resource Center (http://genetics.ncai.org) to provide tribal leaders and researchers with information on genetics research development in Native communities.

This initiative recognizes the long and challenging history of research in American Indian and Alaska Native communities. American Indian and Alaska Native people are one of the most heavily-studied groups in the United States. Unfortunately, the long history of research in Indian Country has included some instances of harm to American Indian and Alaska Native tribes and peoples. Many Native peoples are wary of research and do not trust researchers. This is largely due to the fact that the term "research" generally reminds Native peoples of the myriad projects historically conducted that did not benefit Native communities, and even, in some cases, resulted in harm to these communities.
It is in the spirit of affirming tribal sovereignty, traditional laws, and the role of appropriate research that NCAI submits comments on the Draft NIH Genomic Data Sharing Policy. There are five overarching points we want to highlight in these comments, including:

- **Tribal nations have sovereignty over research conducted on tribal lands and with tribal citizens**;
- **Researchers must secure active tribal approval for the collection, use, and sharing of tribal data**;
- **There are successful models of tribally-driven data sharing that serve to both protect and benefit Native people;**
- **Research ethics need to acknowledge the importance of community consent alongside individual consent; and**
- **Research ethics need to include protections for biological samples collected from both living and deceased human beings.**

NCAI is interested in ongoing engagement with institutions like NIH about how ethics and data sharing protocols need to evolve to acknowledge tribal sovereignty. In addition, our NCAI Policy Research Center has developed a range of educational materials to inform researchers and academic institutions about the particularities of tribal research regulation that are highlighted below:

> *Research that Benefits Native People: A Guide for Tribal Leaders* (2009). With financial support from the Administration for Native Americans, the NCAI Policy Research Center partnered with the First American Land-grant College and Organization Network (FALCON) and the National Indian Child Welfare Association (NICWA) to create a curriculum and in-person training to equip tribal leaders, Native students, and other Native community members to understand and manage research and program evaluation. Participants are presented with typical research scenarios faced by tribal leadership and staff. The curriculum was developed in response to requests from tribal leaders who wanted resources to make better decisions about the proposed research in their communities and was launched in September 2009 following pilot use in several tribal communities. The five modules of this research curriculum have been field tested and are being used with tribal communities at their request and as funding is available. It emphasizes the validity of Indigenous knowledge while highlighting the benefits of western research methods when used in an ethical and community-informed manner.

'Walk Softly and Listen Carefully': Building Research Relationships with Tribal Communities Report (2012). In partnership with Montana State University's Center for Native Health Partnerships, the NCAI Policy Research Center developed a resource guide to provide insights for researchers committed to developing research that benefits Native peoples.

Data Control Options for American Indian/Alaska Native Communities (2012). An information sheet that highlights data sharing concerns and methods for use with tribal nations.

Research Regulation in American Indian/Alaska Native Communities: Policy and Practice Considerations. This paper describes different ways to institutionalize research regulation in communities and reviews the legal basis for tribal regulation of research. It then describes different kinds of research review board structures communities might use and the pros and cons of each board structure. Possible review board options include Institutional Review Boards (IRBs), community advisory boards (CABs), and other review board structures. This paper also discusses jurisdictional issues, such as what kinds of research should be reviewed by community boards and how these boards might relate to federal and university research regulatory bodies. Finally, a brief discussion on methods for enforcing community research review decisions is included in this paper.

Research Regulation in American Indian/Alaska Native Communities: A Guide to Reviewing Research Studies. This paper provides a detailed discussion of each stage of research review from study proposals to publications. This paper is meant to serve as an interactive guide for communities to consult when they are reviewing research studies and includes a detailed checklist that can be used in the review process. The paper begins with a description of components that should be included in research proposals, such as informed consent procedures, data collection/storage methods, and budget/funding sources. Next, the paper describes issues communities may wish to consider when reviewing research proposals including control of data through written contracts and tribal law. Finally, the paper discusses community review of ongoing research studies and research publications, which can be a complex and challenging process.

Federal Data Collection in American Indian/Alaska Native Communities. This paper presents recommendations to federal agencies for data collection in American Indian/Alaska Native (AI/AN) communities. The National Congress of American Indians Policy Research Center has developed this paper in response to numerous requests from federal agencies soliciting advice about how to improve data collection processes in AI/AN communities. Overall, we recommend that federal agencies openly consult with tribal governmental officials, and seek their insights and support. This paper is meant to serve as a guide for federal agencies engaged in data collection, as well as the analysis, interpretation, and implementation of data in the development of policies and programs.

In addition to the overarching points noted above, NCAI provides the following comments on the specific elements of the Draft NIH Genomic Data Sharing Policy:

I. Purpose

As part of the purpose and expectation set forth in the Draft NIH Genomic Data Sharing Policy, we recommend an amendment to the sentence that currently reads, "Sharing research data supports the NIH mission" to "Sharing research data in an appropriate way supports the NIH mission" in order to acknowledge the spirit of the mission that calls for the application of knowledge in a way that enhances health, lengthens life, and reduces illness and disability. Current research ethics protocols have emerged in large part because there has been (and continues to be) inappropriate sharing of research data.

II. Scope and Applicability

It is not clear from the language provided under the Scope and Applicability section whether and how this policy applies to NIH-funded research that involves large-scale genomic data that has been collected from humans who have since passed away (or who are now deceased). It will be important to include information on the ethical protocols and policies involving biological samples and other data from deceased human beings.

In addition, the Scope and Applicability section should speak to how this Policy takes into account the National Institutes of Health Guidance on the Implementation of the HHS Tribal Consultation Policy, specifically as the HHS Tribal Consultation Policy affirms the following:

> "Indian Tribes exercise inherent sovereign powers over their citizens and territory. The U.S. shall continue to work with Indian Tribes on a government-to-government basis to address issues concerning Tribal self-government, Tribal trust resources, Tribal treaties and other rights. Tribal self-government has been demonstrated to improve and perpetuate the government-to-government relationship and strengthen Tribal control over Federal funding that it receives, and its internal program management. Indian Tribes [sic] participation in the development of public health and human services policy ensures locally relevant and culturally appropriate approaches to public issues" (pp. 2-3).

Control of data collected from tribal citizens and on tribal territory is a critical component of tribal sovereignty and impacts tribal participation in the development of public health and human services policy.

III. Effective Date

No comments.

IV. Responsibilities of Investigators Submitting Genomic Data

A. Data Sharing Plans

There should be a protocol established for Institute or Center Program or Project Officials to follow if the data sharing plans involve tribal data to ensure that tribal sovereignty is being maintained. The National Institutes of Health Guidance on the Implementation of

the HHS Tribal Consultation Policy refers to the development of an NIH Tribal Consultation Advisory Committee (TCAC) who could assist with the development and/or oversight of this protocol. In addition, HHS has an American Indian/Alaska Native Health Research Advisory Council (HRAC) and the Secretary's Tribal Advisory Committee that could also assist.

In addition, as part of the Briefing Book published for the 2013 White House Tribal Nations Conference that took place in November 2013, NCAI recommends President Obama and his Administration take the following action:

> "Convene a tribal consultation at the level of the HHS Data Council and develop an agency-wide policy on data management in Indian Country. Echoing the call from the HHS American Indian and Alaska Native Health Research Advisory Council (HRAC), NCAI requests that HHS convene a tribal consultation at the level of the HHS Data Council and develop an agency-wide policy on data management in Indian Country" (p. 20).

B. Nonhuman and Model Organism

1. Data Submission Expectations and Timeline

It is not clear from the language provided under the Nonhuman and Model Organism section whether nonhuman data includes data that has been collected from humans who have since passed away (or who are now deceased). It will be important to include information on the ethical protocols and policies involving biological samples and other data from deceased human beings.

2. Data Repositories

The Alaska Area Specimen Bank (see http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3629262/) has established tribally-driven and culturally appropriate protocols for the management and sharing of biological data collected from Alaska Native peoples. It could be listed as a resource for researchers and Institute/Center Program or Project Officials who need guidance on appropriate methods for managing and sharing genomic tribal data in a way that honors tribal sovereignty and the need to both protect and benefit tribal people through research.

C. Human Genomic Data

1. Data Submission Expectations and Timeline

Guidance to govern human genomic data submission timelines and data release expectations needs to acknowledge tribal sovereignty over data collected on tribal lands and with tribal citizens. Most recently we have raised concerns about "passive approval" language included in NIH Funding Opportunity Announcements (see http://grants1.nih.gov/grants/guide/pa-files/PAR-11-346.html) where publication

timelines and research expectations are seemingly put at odds with tribal sovereignty over data. In order to uphold tribal sovereignty and ensure that research protects and benefits Native people, NIH policies and published language must honor tribal oversight of research that takes place on tribal lands and with tribal citizens. We are eager to continue to engage with NIH about how to affirm tribal sovereignty in research and produce research that has meaningful impact – we believe these are not conflicting aims.

De-identification to protect individuals from whom data is collected may need to take place at both an individual and a tribal level, meaning that there is a need for data sharing protocols that protect a person's identity and the identity of the tribe that person belongs to in the case of American Indian and Alaska Native data. This is especially true given the persistent violations and stigma facing tribal members when these protocols are not in place and data is shared without both individual and tribal consent and de-identification. De-identification of tribal data may be complex due to the small size of communities and unique characteristics of American Indian and Alaska Native tribes and peoples in the larger population. Protocols for the review of tribal data sharing requests must take this complexity into account.

2. Data Repositories

The registering of studies with human genomic data should take into account tribal sovereignty over data as detailed above with regards to de-identification of data, individual and tribal consent, and management of biological data collected from individuals who have since died. Timelines for registering this data should acknowledge the time it takes to secure tribal approvals. In addition, the Alaska Area Specimen Bank (see http://www.ncbi.nlm.nih.gov/pmc/articles/ PMC3629262/) has established tribally-driven and culturally appropriate protocols for the management and sharing of biological data collected from Alaska Native peoples. It could be listed as a resource for researchers and Institute/Center Program or Project Officials who need guidance on appropriate methods for managing and sharing genomic tribal data in a way that honors tribal sovereignty and the need to both protect and benefit tribal people through research. Tribes and researchers working with tribal data should not be compelled to share data if tribes have not approved data sharing.

3. Tiered System for the Distribution of Human Data

While current ethics protocols require informed consent from individuals, there is a need to expand these protocols in the case of data collected on tribal lands and with tribal citizens to also require the informed consent of tribes for data usage. This is important for both primary data collection and use and secondary data collection and use as there have been documented instances of harm to individual tribal members and tribal nations from inappropriate and unethical secondary use of data (e.g., diabetes research data collected by researchers at Arizona State University that was later used in secondary research on schizophrenia). While case was settled out of court, it sent waves throughout Indian Country and the research world, with many

tribes and American Indian and Alaska Native organizations, including NCAI, passing resolutions expressing support for the tribe's lawsuit against the Arizona Board of Regents. This case also caused many American Indian and Alaska Native communities to seek new ways to protect themselves from being deceived about the purposes of research projects and to control how their communities are portrayed in publications or presentations by researchers. Again, protocols in use by the Alaska Area Specimen Bank may be instructive.

4. Informed Consent

NCAI recommends that DNA and biospecimens should be considered identifiable in and of themselves because genome sequencing technology is making it more possible to link DNA with an individual. NCAI is concerned about secondary use of data, so rigorous data protections should be applied to genetic information and specimens containing DNA. NCAI advocates specific informed consent be required for all studies in which an individual's DNA or data are used, and that general informed consent not be allowed.

NCAI recommends that future research use of data require informed consent for secondary analysis. Regardless of whether the secondary data could be identifiable or not, some American Indian and Alaska Native peoples believe that human tissue, blood, and other biological specimens are sacred as they contain a person's essence and spirit. For this reason, sharing specimens between investigators or moving them from facility-to-facility is worrisome and spiritually concerning for tribal nations and peoples. Other potential harm may occur when tribal nations' names are linked to biological specimens, genetic material, or other kinds of data. Even when a sample or data point does not identify the individual participant, the tribal nation may be named. If specimens and data are then used for secondary analysis in ways not authorized by the tribe, there is the potential for group harm and stigmatization of the tribe in resulting publications and reports.

NCAI recommends that all secondary uses of collected specimens and data should require an additional consent process. Additionally, clearly defined choices or checkboxes should be incorporated into the informed consent form for participants to specify which types of studies and how they would or would not like to participate. Individuals should have option to identify their own categories of research they would permit or disallow. The ability of participants to self-identify their own categories of research they would permit or disallow should be clearly explained and defined in the informed consent process. However, NCAI cautions against using consent processes to garner blanket consent before future and secondary aspects of research design and data use have been determined. While many members of the general population may have a better sense today than in past about research and their rights, researchers and research review bodies should not transfer responsibilities around consent processes to potential participants. Researchers and research review bodies have significant responsibilities to ensure consent processes are informed and that human subjects are protected throughout the entire research process.

While current ethics protocols require informed consent from individuals, there is a need to expand these protocols in the case of data collected on tribal lands and with tribal citizens to also require the informed consent of tribes for data usage. This is especially crucial in the context of open-access to data. There should be an additional layer of consent required for data from tribal citizens to ensure appropriate de-identification and to prevent harm in the case of providing open-access to that data. Again, protocols in use by the Alaska Area Specimen Bank may be instructive. Where university Institutional Review Boards can oversee research through a particular set of ethics, tribal research review bodies may also need to be consulted to ensure for cultural and community protections that existing research ethics ignore.

NCAI recommends that limited data sets should not be shared outside the original research team without permission from individual research participants and tribal nations involved in the study. The sharing of data outside the original research team falls under NCAI's broader concern about secondary use of specimens. There are models for making data accessible to outside research teams without compromising tribal confidentiality, such as a data enclave – or a secure space for researchers to perform analyses that require a protected or controlled environment. The National Institutes of Health has offered data enclaves as an option for the original research team to retain control over data, but to provide the aggregate results of secondary analyses to outside requesting research teams in an ethical way.

NCAI recommends that the regulations be clarified regarding the current practice of allowing research on biospecimens that have been collected outside the research study to require consent, regardless of whether a research participant's identity is never disclosed to the investigator. NCAI is concerned with the secondary use of these specimens without informed consent due to potential for harm of the individual participants and tribal communities as groups. Biospecimens that are collected outside of the research study such as "left-over" tissue and blood may be considered sacred by tribal nations and peoples and so sharing them between investigators or moving them from facility-to-facility may circumvent the human subject protection provided as part of informed consent processes.

5. Institutional Certification

NCAI recommends that IRBs work to ensure that researchers abide by data sharing, use, review, and dissemination agreements stated in research review applications; and that IRBs pay particular attention to the complexities around de-identification of data due to the small size of tribal communities and unique characteristics of tribal nations and peoples in the larger population that may require initial and continued research review.

Further, where university Institutional Review Boards can oversee research through a particular set of ethics, tribal research review bodies may also need to be consulted to ensure for cultural and community protections that existing research ethics ignore. Risks to tribes are a priority and must be considered and prevented as they are never justified. The same survey instrument or types of questions might be considered

minimal risk in one population, but greater than minimal risk with another group. For example, questions about topics that have been historically sensitive in American Indian and Alaska Native communities, such as alcohol use or genetic risk, may be considered higher risk than if the same questions were asked of other groups. Individual studies should be assessed by local IRBs or review boards to determine what level of risk is posed to potential study participants. Notably, tribal nations have a variety of research review structures. Some tribal nations have their own formal IRBs, while others have developed alternative forms of research review committees or processes. The local research review process a tribe has developed, regardless of its form, can help to ensure risks specific to the population will be minimized. Tribal IRBs and other review boards may have more insight about potential participants' ways of life, cultures, languages and community traditions that could inform decisions about human subject protection and research risk. They may also know and understand more about the issues and disparities the community faces and have ideas of how to be proactive and best address these issues. University and federal review boards should also be encouraged to include American Indian and Alaska Native peoples and researchers to serve on research review bodies, especially when research with American Indian and Alaska Native tribes and peoples have been put forth. This is particularly important in the case of research review in an urban Indian context, where there may not be a formal tribal governance mechanism to provide research review.

While current ethics protocols require informed consent from individuals, there is a need to expand these protocols in the case of data collected on tribal lands and with tribal citizens to also require the informed consent of tribes for data usage. This is important for both primary data collection and use and secondary data collection and use as there have been documented instances of harm to individual tribal members and tribal nations from inappropriate and unethical secondary use of data (e.g., diabetes research data collected by researchers at Arizona State University that was later used in secondary research on schizophrenia). This is also especially crucial in the context of open-access to data. There should be an additional layer of consent required for data from tribal citizens to ensure appropriate de-identification and to prevent harm in the case of providing open-access to that data. Again, protocols in use by the Alaska Area Specimen Bank may be instructive.

De-identification to protect individuals from whom data is collected may need to take place at both an individual and a tribal level, meaning that there is a need for data sharing protocols that protect a person's identity and the identity of the tribe that person belongs to in the case of American Indian and Alaska Native data. This is especially true given the persistent violations and stigma facing tribal members when these protocols are not in place and data is shared without both individual and tribal consent and de-identification.

6. Data Withdrawal

Removal of data from NIH-designated repositories should also be possible when a tribe withdraws its consent to data that is identifiable at a tribal level.

7. Exceptions to Data Submission Expectations

   Allowable exceptions to submitting data to NIH-designated data repositories should include instances where data is identifiable at a tribal level and the tribe has not provided consent for the sharing of that data.

V. Responsibilities of Investigators Accessing and Using Genomic Data

A. Requests for Controlled-Access Data

   NIH Data Access Committees should have particular protocols in place related to requests to access controlled tribal data that speak to de-identification at a tribal level, tribal consent, and tribal protections. The NIH User Code of Conduct should also include language about protocols related to accessing, sharing, and using tribal data.

B. Acknowledgement Responsibilities

   The NIH should also expect investigators who access genomic datasets from NIH-designated data repositories to acknowledge all provisions related to data sharing set out in the particular study for which data was originally collected.

VI. Intellectual Property

NCAI encourages the NIH to ensure its policy language about the patenting of genomic or genotype data and technology is consistent with its mission that calls for the application of knowledge in a way that enhances health, lengthens life, and reduces illness and disability. The interests of health and life must come before market and property interests.

Commenter: National Society of Genetic Counselors, Molly Giammarco

Date of comment: 11/20/2013 18:02

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: The National Society of Genetic Counselors (NSGC) appreciates the opportunity to comment on the National Institutes of Health's (NIH) draft Genomic Data Sharing (GDS) Policy, published September 20, 2013 in the Federal Register. NSGC is the voice, authority, and advocate for over 2,800 genetic counselors – the largest group of clinical genetics care providers in the United States. NSGC recognizes genomic data sharing's potential to enhance research collaborations and improve our understanding of the contribution of variations in the human genome to health and disease states. Improved research collaboration will undoubtedly lead to enhanced clinical outcomes. In facilitating data sharing, NSGC strongly recommends that NIH's final GDS Policy protect the rights, welfare, wellbeing, and privacy of subjects by implementing an informed-consent process that thoroughly addresses incidental findings, duty to re-contact, data accessibility, and prospective and retrospective data collection. Recommendation One: Develop standard consent language that clearly describes methods in place to protect privacy and provide this model consent language to investigators. NSGC recommends that NIH-funded genomic research informed consent language: a) explicitly state that participants' genomic data will be submitted to an NIH-designated data repository; b) explain the varying levels of access that the general public and investigators may have to deposited data; c) implement steps to protect participant privacy and identity; d) address the possibility of participant identification; and e) explain how a participant's deposited data could still be used after the participant withdraws from the original research protocol. The informed-consent process should clearly explain that the NIH GDS Policy requires participants to submit their de-identified and coded data to an NIH-designated data repository (i.e., dbGaP) to maximize the potential opportunities for discovery and enhance public benefit.The informed-consent process should explain dbGaP's two-tiered structure: the data available to the general public without restrictions and the data only available in a controlled fashion to investigators with authority to access specific data sets. NSGC supports the draft Policy's expectation that participants explicitly consent to sharing their data through open-access mechanisms.Participants should understand the specific steps that investigators at the submitting institution will take to protect their privacy before submitting their data to an NIH-designated data repository. This

includes explaining that data will not include any personally identifying information and will be assigned a randomly generated and unique code known only to the submitting institution. The final Policy should also explain the Certificate of Confidentiality and any additional measures that protect participant privacy. The NIH should develop policies that encourage institutional IRB's to allow the sharing of genomic data with limited or no consent if it was collected before the enactment of these new proposed rules. Many of these studies involved individuals with a personal or family history of a genetic disease. These individuals participated in research to improve our knowledge about these medical conditions. Prohibiting data sharing in these cases reduces the impact of their research participation.The final Policy should also address the risk of personal identification from coded data, either through a security lapse or by virtue of the highly specific data collected (i.e., individual genetic variants). Although participants have the right to withdraw from original research at any time, the final Policy should explain that data already submitted into an NIH-designated depository could still be used. The final Policy should also explain that even if investigators at the submitting institution destroy locally held data upon participant withdrawal, they cannot retrieve information that has already been deposited and/or disseminated.Recommendation Two: Describe scenarios in which investigators should use anonymized data rather than de-identified data and implement clear and strict penalties for misusing information within an NIH-genomic database. The possibility of identifying individuals by triangulating sequencing data sets in publicly available databases with other accessible personal information, as Gymrek et al., 2013 have demonstrated, will only increase as more genomic data is shared. NSGC strongly recommends that the NIH implement and consistently enforce penalties for investigators who misuse information in NIH-sponsored genomic databases. Individuals who attempt to identify patients and/or research participants with information in these databases should incur severe penalties (e.g. losing NIH-funding for a pre-determined and significant period of time) to discourage such activity.NSGC understands and supports the fact that the NIH will de-identify most data provided, but recommends that the NIH include an option for anonymized data in specific studies that affect high-risk study populations (i.e.: mental health studies, sexually transmitted disease studies, and criminal behavior studies). Investigators' desire to avoid re-contacting consented participants regarding medically actionable incidental findings (MAIFs), however, should not justify anonymous enrollment. Recommendation Three: Address the duty to re-contact individuals regarding medically actionable incidental findings that are obtained using de-identified data acquired through NIH open-access or closed-access databases.As genetic knowledge continues to grow and technologies improve, the data that NIH open-access or closed-access databases acquire will likely identify more MAIFs. Because the draft Policy proposes to de-identify, rather than anonymize data, re-contacting participants regarding medically actionable results will be possible. MAIFs can lead to timely preventive care or treatment (Simon et al., 2012) and the informed consent process should explain the possibility of MAIFs and the potential for re-contact. The final Policy should accordingly address investigators' obligations to search for MAIF, report identified MAIFs to subjects, and include specific information about MAIF in the informed consent form (Wolf et al., 2008; Gliwa et al., 2013). The NIH should also require investigators using information from NIH databases to report MAIFs to the Primary Investigator, who should abide by a patient's consent choices when determining if he/she should re-contact a participant. The consent should include a protocol for returning MAIFs that details a plan for disclosure and addresses situations in which the primary investigator finds an MAIF that is not within his/her specialty. Federal regulations stipulate that the informed consent process include

both risk and benefit information (Wolf et al., 2008), but adequately encompassing the many aspects of whole-genome research, such as addressing releasing incidental findings (IF), is a challenge (Caulfield et al., 2008). The American College of Medical Genetics and Genomics recently published clinical guidelines for disclosing pathogenic or suspected pathogenic MAIFs within 57-known genes that are implicated in 24 conditions (Green et al., ACMG guidelines, 2013). Current research guidelines, however, recommend that disclosing results in a research setting should reflect the participant's preference – which should be determined in the consenting process (Fabsitz et al. 2010). Recommendation Four: Adhere to a timely data release schedule.NSGC supports timely releasing data to promote knowledge and improve patient care. The final Policy should prevent investigators from publishing other investigator's data for an additional six months after data release to encourage the early submission of variant data.Additional Considerations:The final Policy should implement mechanisms to allow data collected prior to the NIH enacting these rules be de-identified and included in the databases. De-identification may require posting single variants rather than linked variants from panels, whole exome, or whole genome testing.NHGRI should encourage grant proposals and provide funding for education and informed consent about data sharing.The final Policy should require that research projects work with NHGRI's Clinical Genome Resource (ClinGen) project (http://www.genome.gov/27555151) to ensure assure the deposition and curation of variants in ClinVar. The final Policy should explicitly state that these rules do not apply to clinical testing. The NIH may need to develop a second set of guidelines to encourage depositing clinical data into ClinVar and other NIH databases.The final Policy should clearly indicate what, if any, penalties apply for failing to release data in a timely manner. Research projects should clearly name the individual responsible for submitting the data. That individual alone should be subject to any proposed penalty.NSGC appreciates this opportunity to provide comments on the NIH's Genomic Data Sharing Policy. We look forward to collaborating with the NIH to continue to properly facilitate genetic information's role in healthcare advancements.ReferencesCaulfield T, McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA, Juengst ET, Kaye J, Kedes L, Knoppers BM, Lemmens T, Meslin EM, Murphy J, Nussbaum RL, Otlowski M, Pullman D, Ray PN, Sugarman J, Timmons M. 2008. Research ethics recommendations for whole-genome research: Consensus statement. PLoS Biol 6: e73.Fabsitz RR, McGuire A, Sharp RR, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. Circ Cardiovasc Genet 2010; 3: 574-80.Gliwa C, Berkman B (2013). Do Researchers Have an Obligation to Actively Look for Genetic Incidental Findings? The American Journal of Bioethics, 12 (2) 32-42.Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin, CL, McGuire A, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker, LG. ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing (March 2013). The American College of Genetics and Genomics. http://www.acmg.net/docs/ACMG_Releases_Highly-Anticipated_Recommendations_on_Incidental_Findings_in_Clinical_Exome_and_Genome_Sequencing.pdf Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y (2013). Identifying personal genomes by surname inference. Science 339(6117): 321-4.Simon C, Shinkunas LA, Brandt D, Williams JK. Individual genetic and genomic research results and the tradition of informed consent: exploring US review board guidance. J Med Ethics 2012; 38; 417-422.Wolf SM, Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton EW, Fletcher JG, Georgieff MK, Hammerschmidt D, Hudson K, Illes J, Kapur V, Keane MA, Koenig BA, Leroy BS, McFarland

EG, Paradise J, Parker LS, Terry SF, Van Ness B, Wilfond BS (2008). Managing incidental findings in human subjects research: analysis and recommendations. J Law Med Ethics. 36(2): 219-48, 211.

November 20, 2013

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

The NHGRI/NHLBI Centers for Mendelian Genomics is comprised of three sequencing centers—one at Yale University, one at the University of Washington and a joint Center at Baylor College and Johns Hopkins University. We respectfully submit the following comments regarding the Draft NIH Genomic Data Sharing Policy, on behalf of the CMG IRB/Consent Review working group.

Drs. Debra Mathews, Holly Tabor and Deborah A. Nickerson
dmathews@jhmi.edu
holly.tabor@seattlechildrens.org
debnick@uw.edu

GENERAL COMMENTS

1. **Policy Clarification, Supporting Documents and Definitions**
   We believe that the policy should explicitly state whether or not it replaces the existing GWAS data sharing policy. For implementation of this policy to be effective, it will be necessary to update supporting documents, such as the GWAS points to consider document with draft consent language. These documents should include information about how IRBs should evaluate both prospective and retrospective informed consent language for data sharing and data use limitations under this policy. Additionally, it would be helpful if the policy or supporting documents contained definitions for some of the terms in the policy, including "large-scale human and non-human genomic data" and "NIH designated data repository."

2. **Policy Application**
   It is unclear from the document to which studies/analyses this policy would apply. While the policy states that investigators should correspond with NIH program officers about details, more clarification in the policy itself would be beneficial. Specifically, the examples provided in Appendix A suggest that data sharing would be required for a study that sequenced more than one gene in more than 100 people. Would partial targeted sequencing of two genes in 101 people require sharing to comply with this policy? If so, at which level of data release? Similarly, the examples suggest that the exome or whole genome sequence in one person would require data sharing to comply with this policy. For both of these examples, the reasons for and benefits of data sharing are not clear, and it is not clear if the significant resources involved to facilitate approval and submission would be worthwhile. We respectfully suggest that the relative merits and challenges/possible harms of these very different kinds of projects and data sharing be carefully articulated and evaluated in the consideration of this policy.

3. **Privacy Protections for Rare Diseases/Phenotypes**
   Based on our collective experiences in the CMG project, we know that individuals and families with rare diseases and phenotypes are at substantially greater risk of identification through data sharing, even in controlled access databases. For example, if samples are indexed in dbGaP by phenotype, individuals with extremely rare conditions may be easily identified and

linked to identifiable information, including articles with photographs or other potential identifiers. Our experience suggests that the relative benefits and harms of data sharing for data from these samples/cohorts should be carefully evaluated, and that special protections and enforcement of protections may be recommended for use of these data.

4. **Enforcement**
   In order for this policy to be effective, it will be necessary to have adequate enforcement, and, specifically, enforcement of consequences for those who access data and violate the protections/rules for data use. Such an expansion of data sharing will increase the likelihood of misuse and related harms. In our opinion, such enforcement will be resource-intensive, but without it the potential harms of data misuse may be unacceptably high and may erode public trust in science and research.

5. **Promoting Competition and Discovery**
   The policy should take into consideration the potential impact on researchers, specifically those studying rare phenotypes, of mandatory sharing within a relatively narrow time window. For example, if a researcher has samples from two or three families with a rare phenotype, s/he may not be able to identify the causal genes/variants. They may ask another researcher who has additional families to collaborate, but the second investigator may refuse. The first investigator could be obligated to share their genomic data from these families within a six month window, and at that time the second investigator could apply for and obtain access to that data, use it to identify the causal gene and publish the results, without collaborating with the first investigator and without including them as coauthors on any publications. In this way, the implementation of the proposed data sharing policy could stifle collaboration and data sharing, and create a competitive advantage for those who do not comply with the policy or those who do not receive NIH funding and, therefore, are not obligated to share their data. ***The policy and its implementation should consider strengthening the obligation for researchers, specifically those getting access to data in the repositories, to collaborate with data submitters who are studying rare phenotypes or with small sample sizes.***

INFORMED CONSENT COMMENTS

6. **Informed Consent**
   Informed consent is a cornerstone of human subjects protections, and the primary mechanism for demonstrating and fostering respect for persons in research. We worry that the ambiguity in the draft policy around questions of both evaluating previously signed consents and developing prospective consents will frustrate efforts to appropriately respect those whose tissues and data are used in research. The CMGs have spent two years grappling with how best to honor the informed consent of those who have contributed samples and data to research, not only with respect to the inclusion of previously collected samples in CMG research, but also regarding the deposition of sequence data from those samples in dbGaP. Each of the three Centers has evaluated hundreds of previously signed consents, and we have seen dramatic variability in both the provision of information about research and in the terms to which subjects have been asked to agree. Given this experience, we do not think that it is appropriate, in many, if not most cases, to permit the submission to dbGaP of all data generated prior to the effective date of the GDS policy, regardless of consent. We believe that it would be more appropriate to develop clear, specific and justifiable criteria for inclusion or exclusion of these

data based on the terms of the original consent documents and that consents should be evaluated against these standards prior to making any decision regarding the deposition of data into dbGaP. This leads us to our next set of concerns, regarding the evaluation of informed consent documents.

7. **Consent Evaluation**

Key questions regarding evaluation of the appropriateness of consents for deposition of data into dbGaP include: What are the standards for appropriateness? Who will conduct the assessment? How are risks assessed? What are the mechanisms for enforcement or audit?

The standards for appropriateness should be described in the GDS policy or supporting documents to avoid inconsistent and unjustified variation in the application of the policy. Possible standards include the presence in the consent document of specific language regarding data sharing, or large-scale data sharing, or merely the absence of language restricting such sharing. Even with standards available, it is unclear that IRBs are equipped or resourced to conduct the evaluation of the very many individual consents that will require such review. As we have learned at the CMGs, even with clear standards for evaluation in place, many consents require discussion and debate, and sometimes the gathering of additional information followed by re-evaluation prior to determination about the appropriateness of data deposition.

The GDS policy, or supporting documents, should also describe standards for evaluating risks to individuals, their families and, where applicable, groups or populations associated with submitted data. Again, who will perform this consideration of risks? How will it be done? What are the anticipated outcomes of such consideration? In addition, what counts as a group to which a specific risk evaluation ought to apply? For example, do individuals/families with a rare disease count? Indeed, how will the new policy facilitate and protect participant privacy for individuals with rare diseases/phenotypes? It may be difficult to develop clear guidance within the current GDS draft policy, given that risks are critically dependent on governance of data use and access, standards for consent evaluation and provisions for enforcement.

As noted above, the draft policy lacks discussion of enforcement or audit. The NIH might consider the inclusion of unsigned/de-identified copies of consent documents with the uploaded data as a potential mechanism for audit and enforcement.

8. **Exceptions**

The last major category of concern is what exceptions will be permitted and how determinations will be made about the appropriateness of any particular exception. First, given the known variation in individual views and preferences regarding research participation and data sharing, we believe that going forward, individuals should be offered the option to opt out of large-scale data sharing. ***However, subjects who choose to opt out of broad data sharing should still be allowed to enroll in research.*** The majority of people will permit such sharing, but the minority who feel strongly about not sharing should be permitted to restrict it.

Second, what counts as adequate justification for exceptions to dbGaP deposition? The draft policy states that this information will be made available on GDS website, but given how important this information is to the evaluation of the policy as a whole, a draft set of mechanisms and standards for exceptions should also be made available for comment.

Finally, as noted above, we also have concerns about enforcement. In particular, how will compliance, including with individual data use limitations, be monitored? What are the mechanisms for identifying, reporting and investigating alleged violations?

DATA SUBMISSION COMMENTS

9. Timeline Definitions

The draft policy is not clear regarding the data submission timelines. Specifically, at what point does the six-month window begin? Is it six months after data generation is completed? Is it six months after the creation of a draft vcf file, or six months after the data are recalled and a final data analysis set created? These details are important and must be clarified. We recommend that the start date be either when the final data set is completed or the date of first publication, depending on the context.

10. Data Submission Timelines

Based on our experiences in the CMG, we think that data release six months after the creation of a *final data set* is a reasonable window for most projects, and will allow investigators a sufficient amount of time to complete their analysis and necessary functional studies.

11. Data Sharing Levels.

It is not clear to us what data are required to be submitted for Level 4, how it differs from the data in levels 2 and 3 or how this submission would be accomplished. More clarity and detail are required for this to be understood and implemented. Additionally, across all levels, more detail is needed about the kinds of phenotype information that will be required for submission, and how this will be balanced with protecting the privacy of participants, especially participants with rare phenotypes or from small or potentially, identifiable populations.

12. Resources for Submission and Review

We do not believe that most investigators and institutions have the resources or knowledge to prepare/upload both genomic and phenotypic data into dbGaP for all the levels of data submission described in Appendix A. In our experiences, both in this project and in other projects, this work is very complex and time-consuming, and supplemental funds would need to be provided with each grant in order to facilitate the scope of data sharing described in the draft policy. Similarly, we are concerned that the NIH/Institutes/DACs may not be prepared to process and certify all of the data described for the various levels of submission in Appendix A. Again, our experience is that this is tremendously labor intensive and time consuming, even under the existing policy, and the proposed policy would require substantially greater resources for implementation.

Commenter: David A. Nielsen, Ph.D.

Date of comment: 11/4/2013 13:27

Comment:

Section II. Scope and Applicability: The Scope and Applicability requirements are adequate and inclusive.

Section IV.A. Data Sharing Plans: It should be made clear that investigators seeking NIH funding need not contact the appropriate officials if they are experienced with the submission of data for sharing.

Section IV.B. Non-human and Model Organism Genomic Data: No comment.

Section IV.C. Human Genomic Data: The deadline for release of data of six months or at the time of acceptance of the first publication should be extended to one year after data submission. As it is the investigator who initiated and conducted the research, adequate time should be given for that investigator to analyze and publish their investigations.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: No comment.

Section VI. Intellectual Property: I support this.

Any other aspect of the draft GDS Policy: No comment.

Commenter: John Nurnberger MD PhD

Date of comment: 11/14/2013 11:23

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: The plans as described make sense to me. If I understand correctly human sequence or GWAS data would be shared at the time of initial publication or within 9 months of data release to the investigators, whichever comes first. This is the policy that NIMH has been operating with for several years, and I am not aware of any problems with it.

The more sharing of these data, the better, as the complex disorders require large sample sizes to make headway.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Office of Research Integrity for the University of Cape Town, South Africa, Robert H. McLaughlin

Date of comment: 11/19/2013 23:28

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: The Office of Research Integrity for the University of Cape Town, South Africa is pleased to benefit from the opportunity afforded by the U.S. National Institutes of Health to review and comment on the Draft NIH Genomic Data Sharing Policy Request for Public Comments.The University of Cape Town aspires to become a premier academic meeting point between South Africa, the rest of Africa and the world. Taking advantage of expanding global networks and our distinct vantage point in Africa, we are committed, through innovative research and scholarship, to grapple with the key issues of our natural and social worlds. In light of our mission, the connection in the draft NIH policy between data sharing, responsible stewardship of scientific resources, and the promotion of fundamental knowledge about nature and living systems for the benefit of humankind resonates strongly with the programmatic direction of biomedical sciences at UCT as a research-led institution. We offer the following comments with the hope that they may contribute to the final form of the NIH policy on sharing genomic data.First, we note that where the draft NIH policy refers to the paramount importance of informed, individual consent, the policy appropriately sets a high standard of education for meaningful consent. "NIH expects the informed consent process and documents to state that a participant's genomic and phenotypic data may be shard broadly and for future research purposes." (§IV.C.4). To meet this expectation both in statements and in mutual understandings between investigators and research participants, funding for scientific investigation, data collection, and specimen management will need to include significant support for participant and community education. The concepts of genomic data and phenotypic data and the distinctions between them are sophisticated ideas. They necessitate education and engagement in many contexts—especially those involving vulnerable populations—if participants are to participate in research on the basis of meaningful, informed consent. Education of both investigators and research participants is especially important where participants' conceptions of biological specimens are relevant to the availability and use of such specimens in scientific research. In the work of many UCT investigators and their peers in biomedical and social sciences, freely shed hair or nail clippings may be understood differently

among research participants than blood, which must be drawn, and specimens such as saliva and urine are widely recognized to embody different ritual, symbolic, and cultural meanings.Second, and also with concern for establishing clear expectations among the research participants who may contribute to genomic research, we note that although the policy contemplates future research with flexibility and sensitivity to a range of scientific interests, it does not address whether participant interests change or terminate when the living individuals to whom data relate are no longer living or could be determined statistically to be deceased. U.S. regulations for the protection of human subjects define a human subject as: "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) identifiable private information. (45 CFR 46.102(f)).Given that the genomic data contemplated by the draft NIH policy may be useful over a prolonged period of time, the NIH policy might productively consider the limitation of this definition to living individuals in regard to whether a right of withdrawal for unused, re-identifiable genomic data may be an enduring right, and whether it might be inheritable or transferable? In the context of research conducted with individuals from vulnerable populations, in particular, the policy might similarly explore whether (instead or by complementary consideration) long-term stewardship of scientific resources can be informed by community-based and/or cultural factors. Such factors can be integrated into the terms and conditions for the use of data as those factors inscribed in consent forms are presently integrated in the data use provisions of the draft NIH policy. Doing so would expand and elevate the consideration of collective interests identified as an IRB consideration in §IV.C.5 (risks to groups or populations), and incorporate such interests as a structural factor alongside informed consent in §IV.C.3 (Tiered System for the Distribution of Human Data).Third, to the extent that the draft NIH policy makes a categorical distinction between human and non-human genomic data, defining microbiome data as non-human, we believe the policy might accommodate some additional flexibility where conceptions of the self and body among research participants who do not support or share the distinction as drawn in policy. Willing and motivated research participants might disagree, for example, about whether their gut bacteria are "theirs" and of-the-body or, by contrast, distinct, internally hosted organisms within the body. Where microbiome data supports comparative studies across populations and a family, community, and/or people may share in, but also be distinguished by microbiome data, different sorts of human qualities (belonging, normalcy, difference, etc.) may attach for which the heightened standards associated with human genomic data might be appropriate and serve to reduce the anxiety and fears associated with biopiracy.Lastly, we appreciate the effort to identify a broad range of data repositories in §IV.B.2 and through which genomic data can be made available to the global community of scientists engaged in genomic research. We recognize that a resource is made "global" not only by its composition originating from widely distributed sources, but also from a structure of access and dissemination that is equally broad and also equitable in terms of supporting the increasingly global and increasingly collaborative community of scientists interested in genomics research.In offering the above comments, I wish to acknowledge the generous contributions of the following UCT colleagues: Profs. Marc Blockman and Nicola Mulder, and Drs. Jantina de Vries, Lesley Henley,and Marilet Sienaert.At UCT, we are anxious to participate in the realization of the NIH goal to pursue genomics research for maximum knowledge and public health benefits, we commend the efforts of the Genomic Data Sharing Policy Team, and we look forward to a final policy that supports our common scientific interests and endeavors

Commenter: Olufunmilayo Olopade, MD, FACP; Sarah Nielsen, MS, CGC

Date of comment: 11/19/2013 13:40

Comment:

Section II. Scope and Applicability: It is our hope that the Genomic Data Sharing Policy will enable the collaborative efforts necessary for the sharing and utilization of the rich data generated by genomic technology, while providing the appropriate safeguards unique to this data. We strongly support a policy that promotes ease of data sharing amongst collaborators as this is the only way to harness the potential of big data. We envision sustainable solutions to data curation and data access as the key to progress and without giving researchers increased opportunity for secondary use of data that are now locked up, progress will be slow. At the institutional level, increased layers of security have hindered data access. We advocate for streamlined and transparent processes for accessing data within institutions and sharing of data between institutions, and perhaps a revision of the rules penalizing hospitals for HIPAA violations. We also recognize the importance of informed consent and specifying how results could be shared with the larger research community (open vs. controlled access). In our experience, it is also critical to inquire whether patients desire relevant research results be returned to them and if they are willing to be re-contacted for future follow-up. Medical and family histories are dynamic and can further inform research results, which is why we also support open access to data for longer periods of time without renewing requests for access. Finally, we feel that the guidelines set forth should be applicable internationally in order to facilitate data sharing through various international consortia.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Response to Draft NIH Genomic Data Sharing Policy Request for Public
Nov. 20, 2013

Thank you for the opportunity to provide comments on the proposed changes to the genomic data sharing policy. We support the goal of sharing this data for the purpose of improving research and ultimately supporting important biomedical advances to improve health.

Before providing comments on specific sections, we would like to note several over-arching concerns and requests.

We realize that the current interface between genomic information, identifiability and human subject issues is complicated, often controversial and evolving. The status of genetic information, particularly the extent to which it should be considered inherently identifiable, is a moving target. It is not clear from the Advance Notice of Proposed Rulemaking on "Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators" at 76 FR 44512 (ANPRM) how the Department of Health and Human Services will move on the issue of tissue, and hence genetic identifiability. NIH's draft policy does not acknowledge these open and evolving issues directly and does not even mention the ANPRM. NIH maintains that because it does not have identifiable data (limited to coded data), NIH is not engaged in human subjects research – yet there seem to be inconsistencies to this presumption. NIH limited open access to dbGaP coded data in response to concerns of identifibility, and obtained a Certificate of confidentiality for dbGaP. Why? We do not minimize the difficulty of these issues. However, we do request that the NIH set forth the regulatory basis for each of the proposed requirements and discuss how the NIH intends the data sharing policy to respond to proposals in the ANPRM and the evolving global discussion vis-à-vis genetic data.

Secondly, we suggest that NIH take a leadership role in educating the public as well as the medical and research community about genomic information, the promise it holds for understanding disease, the need to share this data and why the federal government is mandating this sharing. It is difficult explaining these concepts not only on an institution-by-institution level, but on an informed consent by informed consent level, and the potential for inconsistency and misunderstanding with such an approach is significant.

Specific comments:
Note that we present our point-by-point comments in the order that issues were presented in NIH's draft – not in order of significance.

<u>Section IV.A Data Sharing Plans</u>

> Investigators are encouraged to contact "Institute or Center (IC) Program or Project Officials" for details regarding data sharing and this policy.

Comment: We suggest a trans-NIH process to develop and implement consistent, standardized procedures. This will minimize or hopefully prevent the current confusion that results from IC-specific interpretations that are often in conflict with other IC interpretations. For example, certification language accepted by one Institute has not been accepted by another.

Section IV.B.2 Data Repositories

The draft policy states that "Data should be made available through any widely used data repository, whether NIH-funded or not." Several examples are then listed. The draft policy is silent on how such repositories may be identified and designated.

Comment: We suggest clarifying any standardized vetting process for these other repositories. Also clarify if deposition into a different repository is in addition to or in lieu of deposition into dbGaP/GEO/BioLINK?
Also, if data is submitted into more than one repository, will there be any cross referencing (or even common study number) to maximize integration of the genomic data?

Section IV.C.1 Data submission expectations and timeline

The draft states "NIH will release data submitted to NIH-designated data repositories without restrictions on publication or other dissemination no later than six months after the initial data submission…"

Comment: We suggest clarifying that 'other dissemination' is in reference to secondary research findings and NOT other dissemination of the data itself.

Section IV.C.4 Informed consent (IC)

1. There is some internal inconsistency regarding what institutional body is tasked with consideration of the IC form. In the first paragraph of this section, the IRB is designated: in the third paragraph the IRB, Privacy Board or equivalent group is referenced.

   Comment: Please clarify whether the expectation is for an IRB to review – or if a variety of groups can complete this review.

2. The draft policy says that for studies initiated after the date of this policy, the IC should include whether access will be open or controlled. The expectation is that there will be explicit consent for open access.

Comment: Mandating this detail for an informed consent form (ICF) raises a number of logistical issues as listed here.

- The ICF cannot simply state this option – to make this understandable there would have to be an explanation and description of open versus controlled access and how each are implemented – complete with descriptions of the Data Access Process etc. This will increase the length of the ICF for questionable benefit.
- Because NIH states that many repositories may receive data – do all have the same rules for (and implementation of) open versus controlled access? How would this be included in the ICF?
- Can delineation of open versus controlled access change over time – either as a result of reinterpretation of the data itself or operations of the repository? How will this be handled?
- We suggest that NIH propose some standard language to explain the difference between open and controlled access – some minimum floor of language to include.

General comment regarding next three specific issues (numbers 3, 4 and 5) that relate to the use of data from cell-lines and clinical specimens:
To date, most institutions are working under the premise that a requirement for submission to dbGaP is an informed consent that is not inconsistent with such submission. Despite a FAQ (http://gds.nih.gov/13faqs.html#i2) suggesting the possibility for applying the waiver of informed consent, most institutions have not exercised that option. Institutions are closely reviewing existing consent forms and based on their analysis deciding whether or not data can be submitted to dbGaP – and if submitted whether or not uses must be limited. Data from clinical specimens, which by definition do not have robust research consent, are not allowed to be submitted. This draft policy now proposes mechanisms by which data from clinical specimens and cell-lines can be submitted to the repository. The presumed goal of increasing data capture and sharing raises concern regarding protections of the patients from whom this data came.

3. For studies *initiated after the effective date of this Policy* that propose to use cell lines or clinical specimens, even if de-identified – the expectation is informed consent for future use and broad sharing.

   First a request for clarification: It seems clear that a study initiated after the effective data that proposes to use cell-lines/clinical specimens collected after the effective date would be covered. Question - does this requirement also affect a study initiated after the effective date that

proposes to use cell lines/clinical specimens collected *prior* to the effective date?

Second, a concern that relates to the broader concern noted above about clarifying NIH's view whether genetic information is inherently identifiable and the regulatory basis for the requirements in this draft Policy. NIH can set a standard for dbGaP that is more protective of individuals than the requirements of the human subjects regulations. And in this draft Policy, NIH has done that by not applying existing rules that delineate use of de-identified excess clinical material as not-human-subjects research and hence do not require informed consent. It would be important that NIH state that the more protective requirement in this Policy only applies to this specific data sharing Policy and that it does NOT suggest broader change to current implementation of the existing regulations in other contexts. Otherwise, this new mandate could be interpreted as creating a broadly applicable new standard independent of existing regulations and without consideration of the process for changing these regulations.

The consequences could be significant for all uses of excess de-identified clinical tissue and the scope of not-human-subjects research. This issue deserves focused discussion and justification – and it should not be simply one of many proposals enacted as part of a data sharing policy.

It seems possible that the draft Policy is anticipating the eventual adoption of the ANPRM proposal for 'brief consent' for future, broad use of clinical specimens and data. If this is correct – NIH must understand that the ANPRM 'brief consent' would not be an IRB reviewed and approved document. As NIH proposes more detail in informed consent regarding open versus controlled access, this is not something that would likely be included in the 'brief consent' as seemingly envisioned by the ANPRM.

We cannot emphasize enough the need for NIH to be explicit about its determinations about identifiability and the regulatory implications thereof, and for NIH to work with others within the Department of Health and Human Services to ensure consistent proposals in all related domains or articulate the reasons for different standards in different contexts.

Finally, clinical specimens and cell-lines raise very different issues. We suggest addressing any requirements for clinical specimens or cell-lines separately.

4. The draft Policy allows use of cell lines or clinical specimens, obtained after this policy is in place, and for which there was no consent, to be used for research if "there are compelling scientific reasons that necessitate the use" of those lines. Investigators are asked to "provide a justification for the use of any such materials in the funding request."

   Comment: This presents a difficult situation – as noted above in #3 - NIH suggests mandating consent for use of specimens that are considered not-human-subjects research under current regulations and then proposes a work-around for this new mandate in the event of compelling scientific reasons.

   There are many issues that must be addressed – some of which relate to seeming internal inconsistencies of the NIH data sharing proposals.
   - Submission to dbGaP requires institutional certification that the submission of the data is appropriate vis-a-vis the informed consent form. In fact this draft Policy increases the demands on what should be included in the informed consent form. How then can an institution agree to submission of data when there is no informed consent at all? This inconsistency puts into question the basic rules for submission of data to dpGaP.
   - If one were to agree to submit without consent – how and by whom are 'compelling scientific reasons' determined? Can NIH provide examples of such reasons that it would find acceptable?
   - The draft Policy instructs investigators to provide justification in their funding request. What is the role of the IRB in this determination? What if there is a difference of opinion between the funding agency and the IRB?

5. The draft Policy states that NIH will accept data derived from cell lines or clinical specimens created or collected before the effective date of this policy – even if they lack consent.

   Comment: As noted above, this seems inconsistent with the understood requirement not only of general consent, but the inclusion of specific details in that consent.
   If this is allowed – will this be open or controlled access? How will NIH determine this?

Section IV.5 Institutional Certification

Details of required elements in the Institutional Certification are outlined. Currently we have amended the proposed certification language – and this has been accepted by most all ICs without further discussion. Will this draft Policy remove the possibility of local changes? Specific requirements of the draft Policy include the following:

The certification is to include that "Data submission is consistent with applicable laws, regulations and institutional policies."

Comment: We have to date limited this to applicable federal and specific state laws/regulations in an attempt to be certain we are not responsible for reviewing and considering all state regulations in research that is multi-state.

The certification is to include" "risk to individuals and their families associated with data submitted to NIH-designated data repositories were considered…"

Comment: Local Data Access Committees (DACS) – and not the local institutions -- control access to the data submitted to dbGaP. Therefore we suggest limiting the local institutions' responsibility to the risk of submission to the database itself – and that institutions be permitted to rely on DACs for downstream access.

Section IV.5 Institutional Certification:

Institutions are asked to certify if data submitted with limited uses can be aggregated and used for general research use.

Comment:

We suggest deleting this provision for the following reasons:

- Promised limitations should be respected. Aggregating the data should not override the stated limitations. If a submission is limited to research on diabetes, aggregating this data and permitting general research use ignores this original limitation.
- "General research" is an extraordinarily broad category and includes research of different levels of risk and acceptability. There is a difference between general research validating statistical methods versus the proposed use in "the study of population origins."

Thank you for the opportunity to provide comments on this important draft Policy. We remain available to provide clarifications and/or additional comments upon request.

P. Pearl O'Rourke, MD
Director, Human Research Affairs
Partners HealthCare
Boston, MA

PERSONALGENOMES.ORG™

Personal Genome Project

November 20, 2013

Personal Genome Project: Public Comments on NIH draft Genomic Data Sharing Policy

The Personal Genome Project (PGP) is a global network of research studies with thousands of participants dedicated to the creation of public resources composed of genome and phenotype data. The first PGP research study was founded at Harvard Medical School in 2005, and international sites now exist in three additional countries.[1] The PGP has been at the forefront of participatory research in genome sequencing and has extensive experience with the ethical, privacy, and consent issues involved. We welcome this opportunity to publicly comment on the NIH draft Genomic Data Sharing (GDS) Policy and make recommendations for improvements. Our recommendations can be summarized as two areas for improvements in section IV.C. of the draft policy:

1. to adequately inform researchers and participants of the inherent identifiability of genetic data, and
2. to require researchers to share with participants their personal research data in order to establish reciprocity and to increase data sharing

The inherent identifiability of genetic data

The draft GDS Policy makes no mention of the inherent identifiability of genetic data. All genetic and phenotype data shared is mandated to be "de-identified". Footnote eight of the draft states: "'De-identified' refers to removing information that could be used to associate a dataset or record with a human individual. Under this Policy, data 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."

This definition of "de-identified" is inconsistent: genetic data is inherently identifiable. Using nothing more than genetic data and other publicly available data, researchers were able to identify nearly 50 individuals whose samples were "de-identified" (i.e. all public data met the same standards mandated by this draft).[2] It is now a documented fact that this type of genetic data, even if scrubbed of personal information as described in this draft, "could be used to associate a dataset or record with a human individual". Genetic data itself violates the draft's definition of "de-identified".

In the past, de-identification of samples or data sets by stripping personal data (name, social security number, date of birth, etc.) was sufficient to avoid re-identification of a particular subject. Genetic data was not seen as an equivalently identifiable piece of information. This is demonstrated to no longer be the case, and the identifiability of genetic data is likely to increase and may eventually become trivial. Ancestry

[1] Three PGP sites exist currently outside the United States: (1) PGP-Canada, based out of the McLaughlin Centre, University Toronto & Sick Kids Hospital (2) PGP-UK, based out of the University College London and (3) another site in the EU with ethics approval, set to launch in early 2014. The Global PGP network is coordinated by PersonalGenomes.org, a 501(c)(3) nonprofit based in Boston, Massachusetts. To learn more please visit: http://www.personalgenomes.org/mission

[2] Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. "Identifying personal genomes by surname inference." Science. 2013 Jan 18;339(6117):321-4.

PERSONALGENOMES.ORG™

Personal Genome Project

databases currently link genetic elements to surname and in the future are likely to link genetic elements to individual ancestors. Controlled-access databases create a legal barrier to re-identification, but data security breaches are possible and have been an increasingly high profile issue in recent years. If the NIH is to mandate that all participants in NIH-funded studies producing large-scale genetic data agree to broad sharing of their genetic and phenotypic data, it is mandating an exposure of many participants to a known re-identification risk.

If the NIH wishes to uphold the public trust in biomedical research, it must respect the right of research participants to be informed of relevant risks. If all potential participants in these studies are asked to agree that their "genomic and phenotypic data may be shared broadly for future research use", they must also be adequately informed regarding the identifiability of that data.

We recommend this draft be amended to:

1. Add language that acknowledges the inherent identifiability of human genetic data.
2. Add to section IV.C.4 instructions for researchers to inform participants regarding the potential identifiability of the genomic data they are sharing (despite planned de-identification procedures) and, in the case of controlled-access data sets, the potential for data security breaches.

Sharing research data with participants

The draft GDS Policy mandates all NIH-funded research studies that wish to produce "large-scale"[3] human genetic data require that all participants from whom samples are collected consent that their "genomic and phenotypic data may be shared broadly for future research uses". This is elsewhere defined as NIH-designated controlled-access or open-access databases (the latter only if participants "have provided explicit consent for sharing their data through open-access mechanisms").

What is not addressed in this draft is a statement about genomic data sharing with the participants themselves. We strongly recommend the NIH consider including such a requirement for two reasons.

The first reason is to establish reciprocity in the data sharing mandate. This draft mandates all participants in NIH-funded studies generating large-scale genetic data allow broad access to their genomic and phenotypic data to unknown individuals -- without ever having access to that data themselves. Participants' genetic data is sensitive, meaningful, and identifiable. Participants deserve the reciprocal mandate that their personal data being shared with others also be shared with them.

The second reason is that this is a significant opportunity to further the NIH's data sharing goals. Participant-managed data sharing is a promising mechanism for open-access data sharing. Even if participants would not have agreed to open-access at the outset of a study, their attitudes may change. Additionally, participants may wish to share their data with future studies in a selective manner. Participant access to data

[3] Defined as more than 100 participants for genotyping or multi-gene sequence data, or whole genome sequence from a single participant.

PERSONALGENOMES.ORG™

Personal Genome Project

enables an additional participant-managed model for data sharing, and we can imagine a future where numerous studies benefit from participant-donated data.

We recommend the following:

1. For participants consented after the effective date of this policy, add a requirement for researchers to give these participants access to their personal data that is shared with other researchers.
2. Because some researchers may be unable to comply with this requirement, also allow researchers to instead provide specific reasons for why this data sharing cannot be performed. Some mechanism should also be provided for participants to access these reasons in a study-specific manner (such as in a public database).

Minor suggestions

1) In section IV.C.4: "If there are compelling scientific reasons that necessitate the use of cell lines or clinical specimens that were created or collected after the effective date of this Policy and that lack consent for research use and data sharing, investigators should provide a justification for the use of any such materials in the funding request." We suggest clarification of whether the lack of informed consent automatically exempts the researcher from data sharing, or if data sharing is expected to occur despite the exemption.

2) We suggest clarification confirming that "sample identification" using genomic data or other genotypic assays which are not intended to identify individual human participants is acceptable (e.g. detection of duplicate samples across different studies for statistical validity or for quality assurance).

3) "Binary alignment matrix (BAM)" should probably be "Binary Alignment/Map (BAM)". Assuming this is a reference to SAM and BAM files, there is no clear definition what the BAM acronym abbreviates ("B" could potentially mean "BZGF" or "Binary"), but a SAM file is defined here as a "Sequence Alignment/Map": http://samtools.sourceforge.net/SAMv1.pdf

Many thanks to the Harvard PGP staff that contributed to these recommendations
Madeleine Ball, Jason Bobe, Michael Chou, George Church, Tom Clegg,
Preston Estep, Jeantine Lunshof, and Alexander Wait Zaranek

Commenter: Thomas D. Petes

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Dear Colleagues: I have several concerns about the policies proposed. My main concern is that it is very difficult for NIH to propose a document that will address many of the types of experiments that I and other investigators perform. For example, suppose I identify a yeast strain with an interesting mutant phenotype. I then sequence the strain to determine which gene is mutated. Do I need to provide the sequence information from the whole 13 Mb genome even though the relevant information may be a single base difference. Because of these uncertainties, many investigators will be unsure whether they are in compliance with the GDS policy. They may waste time submitting various types of data that are unlikely to be desired by anyone. Second, submission of the data should not be based on when the data are obtained, but when the data are submitted for publication. Most researchers do not want to submit data until the research project is completed and ready for publication. Third, the time line for comments is rather short considering the importance of this issue, and the possibility of applying these regulations will impede rather than aid the research effort. Thank you for considering my comments.Sincerely,Tom Petes

Commenter: Paul Pharoah

Date of comment: 10/3/2013 4:10

Comment:

Section II. Scope and Applicability: Why are genomic data treated differently than other data? Irrational genetic exceptionalism is alive and well.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

RECEIVED
NOV 20 2013
NIH/OBA

PRIM&R
PUBLIC RESPONSIBILITY IN MEDICINE AND RESEARCH

November 20, 2013

Submitted by fax to: (301) 496-9839

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

RE: **FR Document 2013-22941**, Draft NIH Genomic Data Sharing Policy Request for Public Comments (78 FR 57860).

Dear Members of the Genomic Data Sharing Policy Team,

Public Responsibility in Medicine and Research (PRIM&R), a nonprofit educational and professional development organization, appreciates the opportunity to submit comments on the National Institutes of Health (NIH) draft Genomic Data Sharing Policy, as requested in the September 20, 2013 *Federal Register* notice.

For 39 years, PRIM&R has been dedicated to advancing the highest ethical standards in the conduct of research. We accomplish this goal by serving the full array of individuals and organizations involved in biomedical, behavioral, and social science research, particularly the members and staff of human research protection programs and institutional review boards (IRBs). Through conferences and other educational activities, PRIM&R provides balanced, thorough, and accurate information on a range of ethical and regulatory issues affecting research.

PRIM&R agrees that making genomic data available for broad research has the potential to provide important public benefits. As a research ethics education organization, rather than an organization that conducts, oversees, funds, or otherwise has a direct stake in research using such data, our comments focus on dimensions of the Policy that could help to ensure this work is conducted in the most ethically defensible way.

Before we move into our specific recommendations and comments, we wish to make a preliminary point. We strongly urge NIH to clarify that this proposed Policy covers only data and data sharing, and not the collection or sharing of biospecimens. The draft Policy refers in several places—notably in section C.4 on informed consent—to "clinical specimens" and "cell lines." Though these are presumably references to data derived from clinical specimens and cell lines, and not to the specimens/cell lines themselves, the mention of biological source materials may lead to confusion. It is essential that NIH make absolutely clear that this Policy applies solely to data, to avoid confusing guidelines dealing with data and guidelines dealing with specimens. In accord with this clarification, our comments refer only to data sharing, and not to specimen collection or sharing.

Below, we make one broad recommendation, and then add two narrow requests for further guidance and clarification.

I. Informed Consent as an Educational Opportunity

The draft Policy presents a number of requirements regarding informed consent as a condition of submitting data to the appropriate databases. Two of PRIM&R's guiding principles are that protecting the rights and welfare of human research subjects is of primary importance, and that a robust informed consent process is a principal mechanism by which such protection is operationalized. However, the activity discussed in this draft Policy does not qualify as human subjects research under the Federal Regulations codified at 45 CFR 46, because it solely involves the submission of de-identified genomic data to a database. Therefore, as long as these data are truly de-identified and the proposed uses do not constitute human subjects research, PRIM&R urges NIH, as it considers the infrastructure needed to promote responsible genomic data sharing, to move away from a model that focuses narrowly on what consent documents for discrete studies say about the collection, de-identification, and secondary use of any research data generated. Nevertheless, the ethical principle of respect for persons (as described in the Belmont Report and elsewhere) supports the contention that those who provide genomic information for research purposes be told that their genetic material or data is going to be so used; we do not suggest otherwise. Rather, we urge NIH to utilize these data sharing possibilities as opportunities to educate the general public, as potential sources of specimens and data, about genomic research and research in general, and to adopt a model of disclosure for data that has this broad educational goal at its core.

More specifically, we suggest that NIH create, or underwrite the creation of, a toolkit of understandable, accessible, layperson-friendly information about how and why biological specimens and genomic data are being collected and shared.[1] This toolkit should be provided to all investigators who might be required to submit information to a database, as well as to clinicians who must be aware that data and specimen obtained in routine clinical care may be used in research. The toolkit might involve a script for investigators with visual aids for potential donors, a brochure to review, and/or a short video to discuss. It should be generic, rather than specific to any particular study or data repository, so that it could be used by any NIH-supported researcher in conversation with any individual whose de-identified data might be collected and shared.

The goals of this conversation, as facilitated by such a toolkit, should be (1) to inform individuals in the research setting that their genomic and phenotypic information from discrete research studies may be used for additional research in the future; (2) to inform patients in the clinical setting that information and specimens obtained for their clinical care may be used for research; (3) to explain to all potential donors why such information sharing is an important part of the research enterprise and public health; and (4) to encourage all individuals to accept the collection and retention of their

[1] While the collection of clinical biospecimens is beyond the scope of this policy, we acknowledge that it is difficult to dissociate the collection of specimens and the collection of the genomic information derived from them. A layperson-friendly communications toolkit about data sharing will likely need to provide information about the source materials from which those data are derived.

genetic information for research. It should not be to solicit their consent for research, as such, though a mechanism should be available for individuals to opt out.

We suggest that NIH consult with health communications specialists to determine how the relevant information can be best communicated. Below are some points we think it would be important to clearly communicate in any educational materials created, and some suggestions for the scope of the language that might be used:

- The National Institutes of Health (NIH) stores genetic information from people who have been in research.
- If researchers can use this information it will help us better understand the role that genes play in human health.
- It is routine to send genetic information to the NIH.
- None of the information sent to the NIH will include your name **or other identifying information.**
- Researchers must receive permission from the NIH to use the information for research.
- If you do not want us to send your genetic information, just let us know.
- Do you have questions?

Again, this is just a rough sketch of the kind of information we think ought to be communicated. The development of an effective communications toolkit will require assembling individuals who are skilled in health communication and adult education in order to create materials with the appropriate tone, level of clarity, and scope to be understandable to the average person. Such experts will be essential because few clinicians and investigators have the training to facilitate this type of conversation.

We strongly urge NIH to seize this opportunity to educate the public about the emerging research landscape and the role each person can play in the advancement of scientific knowledge, by investing in the creation of a general communications toolkit.

We now offer two additional points focused on aspects of the draft Policy that require further clarification.

II. Guidance Regarding "Compelling Scientific Reasons" to use Tissue Without Consent

The following language appears in section IV.C.4 of the draft Policy:

> 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. *If there are compelling scientific reasons that necessitate the use of cell lines or clinical specimens that were created or collected after the effective date of this Policy and that lack consent for research use and data sharing, investigators should provide a justification for the use of any such materials in the funding request.* (78 *FR*57862, emphasis added)

It is unclear what would count as a "compelling scientific reason" to use the data derived from cell lines or specimens collected without consent for research purposes, or who is expected to make the determination that there are such reasons. If the Policy intends for an institution to determine that the research is scientifically compelling enough to conduct without consent, prior to the submission of a funding application to NIH, then the Policy should make clear who is charged and authorized to make such determinations. NIH should also develop a plan to assure this exception to the rule is not abused. Further, more detailed criteria for the application of this exception need to be created to both guide and limit its use. This guidance is absolutely necessary even if the NIH believes that this is a role for the IRB.

III. Clarification Around "General Research Use"

Our final comment is a recommendation for clarification and consistency within the proposed Policy. In Section IV.C.5, regarding institutional certification, the draft Policy states that,

> Institutions should indicate in the certification whether aggregate genomic data from datasets with data use limitations may be appropriate for general research use.... If so, the aggregate genomic data will be made available through the controlled-access compilations of aggregate genomic data to facilitate secondary research." (78 *FR* 57862-63)

This statement seems inconsistent with other points in the Policy that indicate that research uses of genomic data must be consistent with what is indicated in the informed consent document signed by the individual whose genomic information it is. If a person has given consent to use his or her genomic information for specified purposes only, such as for research on a particular disease, rather than for future research broadly, then on what basis would an IRB decide that the "general research use" of such data is "appropriate"? What are the boundaries around such "general research uses," and how—from the perspective of an individual who has allowed use of his or her information only for a specified set of purposes—would allowing that person's information to be used for "general research" be relevantly different from using it in research the person has explicitly ruled out?

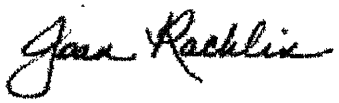
Perhaps another way to articulate what is perplexing here is to ask how the category of "general research use" is distinct from the category of research uses to which information that is designated as "open-access" would be put. If it is not different—and we admit that we do not understand, as written, how it would be—then it seems that no data accompanied by data use limitations could appropriately be used for such "general research uses" since the person, by not electing "open-access," presumptively meant to remove his or her data from use in such "general research." At the very least, NIH needs to clarify these categories, to provide better reasons why general research uses would not routinely be ruled out by the limitations placed on research uses during the consent process, and to specify by whom determinations that such uses are "appropriate" may be made.

IV. Conclusion

We hope that our recommendations regarding a communications toolkit, and our requests for further clarification and rethinking, provide useful direction to NIH as it further develops its

genomic data sharing policy. We welcome the opportunity to collaborate with NIH on the development of a the toolkit, should that be of interest, and, more broadly, on promoting the goal of responsible genomic data sharing that is so important to all stakeholders in the research enterprise.

Respectfully Submitted,



Joan Rachlin, JD, MPH
Executive Director, PRIM&R

Cc: Board of Directors, Public Policy Committee

1808 Dorchester Drive
Oklahoma City, OK 73120-4706
November 15, 2013

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

Dear Sirs:

I am a member of the Kiowa Tribe of Oklahoma, one of the important constituent Tribes involved for more than twenty years with the Strong Heart Study, which is supported by the National Heart Lung and Blood Institute. I hereby respond to the call for comments regarding the proposed NIH policy: "NIH Genomic Sharing Policy". Although I have a number of affiliations, I am responding only as an interested member of an affected American Indian Tribe with a certain degree of experience in the field.

The NIH, as a federal agency, continues to demonstrate that it is either ignorant of, or disregards, its responsibilities to American Indians and Alaska Natives. It does this in opposition to stated policies of the Department and of the President of the United States. This deficiency is compounded by a related failure: requirements for the ethical conduct of community based participatory research. The continued disregard by the NIH for Departmental and Presidential policy in regard to American Indians and federal consultation is rightly condemned.

Concerns about the proposed policy extend to all policies of the NIH regarding disposition of specimens and data obtained from certain American Indian Tribes and their respective citizens. Certain precepts must guide the imposition of such sweeping and paradigm-shifting policies when they apply to sovereign American Indian Tribes. It is important to note that the NIH, as a federal Agency, has dual responsibilities to American Indian Tribes: treating with the latter on *a government to government basis* and at the same time *fulfilling federal trust obligations* to the same tribes. The trust responsibility is as important in most respects as the sovereign nature of the Tribes. To date, there is no evidence that the NIH exhibits the slightest cognizance of these federal, and, one might add, ethical, responsibilities. An important consequence of these responsibilities is that questions relating to NIH policy simply have to be conducted between the Agency and the respective Tribes. Leaving such negotiations to investigators is *an abrogation of federal responsibility*.

Many Indian people have special concepts regarding proper a priori respect, even *reverence*, for certain things and situations, even those that the dominant society may think of as no more than quaint tribal superstitions. Tribes have long been concerned about the handling and disposition of their biological specimens, and the data derived therefrom. These concerns are made even more acute by the advent of genetic studies.

The value of American Indian data resides in the very fact that the data are from a specifically identifiable population. Thus, in many if not most cases, it is not reasonable to expect that the identity of the Indian group can be made anonymous. This point should be addressed in the proposed policy.

Another problem with entering of American Indian data into various databases is that it is highly unlikely that American Indians studies will have sufficiently identical consents that their data can be readily commingled with studies having a different set of consents. The NIH policy is unsatisfactorily silent on this important matter. There remains the hint that some repositories are, or often may be, relatively open ended, obviously an unacceptable situation. This matter is not made clear in the proposed policy.

The NIH position is also unacceptable in the urge, however subtle, to secure consents for data dissemination. This is at the heart of the NIH policy and its unacceptability is obvious. This attitude has been severely criticized many times and one would have thought that it would be avoided by modern scientists. The implied coercion pushes the bounds of ethical behavior.

The policy emphasizes that investigators should provide assurances of good intentions, but fails to list any penalties for failing to fulfill the assurances provided.

In conclusion, the inadequacy of the proposed policy is brilliantly illustrated by the following policy language:

> 7. Exceptions to Data Submission Expectations. The NIH acknowledges that in some cases, circumstances beyond the control of the investigators may preclude submission of data to NIH-designated data repositories (e.g. country or state laws that prohibit data submission to a U.S. federal database).

One could not ask for a more striking example of the failure of NIH to take into account the sovereign nature of American Indian Tribes. It is obvious that this section must contain the same provision for the sovereign Indian Nations, all of whom occupy a status equivalent, at least, to the various states of the union.

Including the sovereign Indian Nations in the language of No. 7, along with a paragraph describing other special considerations would put the NIH in a much more favorable position regarding its special relationship with, and its obligations to, the various American Indian Tribes. It could even help bring the NIH into compliance with Departmental and Presidential policies. Further, it just might well facilitate the achievement of the stated goal of wide sharing of data among investigators.

These observations are offered, not only as criticisms, but as a solution that would serve the NIH well. May I offer some further fairly concise guidance:

Kickingbird, K and Rhoades, ER. 2000. *The Relation of Indian Nations to the Federal Government.* In Rhoades, ER (Ed.). American Indian Health-Innovations in Health Care, Promotion and Policy. Johns Hopkins University Press. Baltimore. pp 61-73.

Reid, R and Rhoades ER. 2000. *Cultural and Traditional Considerations in Providing Care to Indians.* In Rhoades, ER (Ed.). American Indian Health-Innovations in Health Care, Promotion and Policy. Johns Hopkins University Press. Baltimore. pp 418-425.

Rhoades ER, Rhoades DA, and Freeman WL. 2000. *Research Ethics and the American Indian.* in Rhoades ER (ed.), American Indian Health-Innovations in Health Care, Promotion and Policy. Johns Hopkins University Press, Baltimore. pp 426-433.

Thank you,

Everett R. Rhoades MD
R/Adm. (ret.) USPHS
Prof. Emeritus of Medicine

Commenter: Ellen Rothenberg

Date of comment: 11/5/2013 18:47

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data: It is quite reasonable to ask to have this sequence data made available no later than the time of initial publication, as in the main text of Section IV.B.1. However, the timelines indicated in Appendix A are much shorter, and are not at all realistic for anyone using large-scale sequence analysis as an integral part of a classic experimentally based, hypothesis-testing study in a model organism. This is perhaps different from GWAS studies or other work with human samples in which the sequence analysis is the end product. But mechanistic, hypothesis-testing studies do not generate sequence tracks as an endpoint of the study, but rather, as a form of data that is used at each step of the study interactively to design the next set of experiments, as the team progressively tests the logical tree of hypotheses which emerge from the initial data. The series of analyses that has to be included in order to meet the standards of a high-quality publication may often include further deep sequencing of samples subjected to defined perturbations or analyses of cell subpopulations fractionated by different methods, costly and time-consuming studies which can only be designed as a result of data from the previous rounds of sequencing. Such work can easily span 2-3 years from the initial collections of sequence data before a study is complete. It is completely inappropriate to force investigators to leak fragments of preliminary data to the public all along the way for studies which remain incomplete and in progress. First, this penalizes any investigator who seeks to include mechanistic experimental science in a study using deep sequencing – results will be pirated by everyone else in the field long before the authors even arrive at their main conclusions, much less get their work in press. This could drive a rush to publish poorly thought-out and logically shoddy or incremental work rather than strong, innovative, well-controlled and well-analyzed research. Second, in many cases where experimental variables are important aspects of the study, it will not be clear until the study is mature whether a given sample was stressed, impure, perturbed inadvertently, or otherwise a poor exemplar. Results from pilot sequence data rounds may be important to determine which samples should be re-analyzed by generating multiple new biological replicates to obtain statistical credibility, and which ones are unlikely to yield insight into the problem at hand. Quality control is most definitive when the results from the study are complete. A lot of poor data will end up in the databases if this deadline is enforced. In analyses of endpoint clinical samples of human patient tissue, there may be a better argument for rapid release. My point concerns sequence data from experimentally created model organism samples that are generated as intermediate steps in classic mechanistic studies, where hypotheses are refined through a logical progression of experimental tests.

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Melissa Rotunno

Date of comment: 11/20/2013 18:28

Comment:

Section II. Scope and Applicability: “This Policy applies to all NIH-funded research that involves large-scale human and nonhuman genomic data produced by array-based or high-throughput genomic technologies, such as GWAS, SNP, whole-genome, transcriptomic, epigenomic, and gene expression data”. Considering the wide range of data type here mentioned, it is not clear where the standards for data submission will be defined. If the datasets will not be submitted using comparable standards (e.g., MIAME guidelines for gene expression arrays), analyses across datasets will not be feasible.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: In reference to paragraph: "The NIH will release data submitted to NIH-designated data repositories without restrictions on publication or other dissemination no later than six months after the initial data submission to an NIH-designated data repository, or at the time of acceptance of the first publication, whichever occurs first." Regarding sequencing data, the timeline is then clarified in the following paragraph of the Appendix: "In general, it is anticipated that this work [QC] could reasonably be completed within three months, and data submission would follow shortly thereafter. Data files may be held in an exchange area accessible only to the submitting investigators and collaborators for a period not to exceed six months from the time of submission. Following this period of exclusivity, the data will be available for research access without restrictions on publication." Therefore 9 or 10 months after production, sequencing data would have to be made available for public access requests from secondary investigators. This timeline is extremely unrealistic. The primary investigators team has basically no chance to get their work published (most likely not even submitted) during this time frame. Suggestion: make the requirement for sequencing data the same as for array data, i.e. at the time of acceptance of the first publication. It is not clear what "register" mean in the following sentence (assuming it differs from the actual data submission), and what type of clock starts after such registration: "Applicable studies with human genomic data should be registered in the database of Genotypes and Phenotypes (dbGaP) no later than the time that data cleaning and quality control measures begin." In general the presently proposed timeline, particularly for sequencing data, incredibly penalizes the primary investigators team: they would basically barely have the time to clean the data, use their time and money to format the data in a comprehensible way and to submit the data, (submitting 100s of BAM files will require non negligible resources), etc… just on time to make the data available for the secondary investigators to publish it closely and possibly even earlier than the primary team depending on editorial luck. It is very important to share data within the scientific community, but it is not clear why for sequencing data it has to happen in such a short timeline, and without running a pilot in order to address issues of data format, feasibility, and required resources.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Edward A. Ruiz-Narvaez, ScD

Date of comment: 11/20/2013 12:22

Comment:

Section II. Scope and Applicability: It is unclear whether the examples provided in Appendix A are exhaustive of the types of research that would be covered under the proposed GDS policy. For example, are studies with less than 100,000 variants exempted from the GDS policy regardless of the number of participants?

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: EXCEPTIONS It is unclear the type of exceptions that would be allowed for submission of human genomic data. EXPECTATIONS FOR DATA SUBMISSION AND DATA RELEASE The expected timeline for data submission and data release is excessively short. The proposed policy allows up to 3 months for data submission after data generation. Concerns about this timeline are: • If a genotyping project is carried out in different batches at different times, what is the data generation time for the project? Is it expected that the investigator submit data to NIH in batches? Or the 3 months clock will start to run after the last batch of genotyping allowing for a single data submission to NIH?• 3 months is an extremely short period for time for QC analysis and cleaning of data after data generation. Such rush to have data submitted after data generation may lead to mistakes in the QC of genomic data.Of great concern is the fact that the proposed policy allows only for 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first. It is materially impossible to have analysis of data (including discussion with all collaborators about the appropriate analysis), preparation of tables, writing of the manuscript (including circulation of different draft versions to all co-authors), submission, and acceptance in just 6 months. Scientific quality of published research is most likely to suffer due to this unnecessary rush to publish within an extremely short period of time. I would suggest a more reasonable period of time (12 – 18 months) for data release after data submission.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Steven L. Salzberg, Ph.D.

Date of comment: 9/27/2013 15:44

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data: Sharing human genomic is vital to the process of accelerating scientific discovery. NIH has led the way in encouraging greater data sharing; however, most human data collected today is not shared, and much of it is never even released to public repositories. Some of the leading journals do not enforce data sharing; for example, JAMA and the New England Journal of Medicine routinely publish genetic studies based on exomes and other human sequence data without requiring the authors to deposit the data anywhere. I have tried in vain to obtain such data, which often isn't in dbGaP, SRA, or anywhere else.

NIH and all other federal agencies should require any data collected using federal funds to be shared as widely as possible, and as quickly as possible. I'm well aware of human subjects protection issues, but these are being used (unfortunately) as an excuse to prevent data sharing. The countless subjects who agree to share their data would, I am certain, be happy to share it more widely, rather than with just the investigators doing a single study. NIH should insist that investigators use consent forms that allow human data to be shared by the entire scientific community. If a subject wants to share his/her data in a more limited fashion, an alternative consent form could be developed - but I think this would rarely be used.

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property: Sequence data collected from subjects should not be subject to intellectual property claims. The genetic information about an individual is not the "property" of someone else, despite a variety of legal cases allowing such claims in the past. The court system is finally beginning to acknowledge this, with the overturning of the BRCA1 and BRCA2 patents. NIH should discourage investigators from filing such claims on any genetic data.

Any other aspect of the draft GDS Policy:

Commenter: Steven Salzberg

Date of comment: 11/14/2013 13:23

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data: The policy on non-human data is absurdly weak - it is weaker than current policy! This is alarming. It says only that the data should be released by the time the genome is published.The problem - and it is quite a severe one in many cases - is that the major NHGRI-funded centers (Baylor, the Broad Institute, and Washington University) are sequencing species at a far, far greater rate than they are publishing papers. Many species sit in the archives for years, and no one (including me and my collaborators, in several instances) can do anything with the data, even though it's technically "available", because we can't publish our findings. For example, I've been working with a collaborator (Rob Norgren, Univ of Nebraska) on a new, improved genome assembly of rhesus macaque for 3 years now. We could have submitted our findings at least a year ago, but Baylor controls some of the data, and they won't permit publication. (Prof. Norgren generated new data, but we combined his data with the original data to improve the assembly.)

Section IV.C. Human Genomic Data: Human data collected for any large-scale "resource" project such as ENCODE or GTex should be shared broadly, with no restrictions on publication. Releasing data while telling others they can't publish on it is not much better than keeping it secret. As a high-profile example, the recent ENCODE data release was not really available until the papers appeared last fall (late 2012). Much of the data was available for a year or more, but because no one could publish any analysis based on that data, it was no different from data that was unavailable.Data release policies need to enforce rapid, unrestricted sharing. Any restrictions are too much, because some investigators will always do whatever they can to keep their data secret and to prevent others from using it.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Andrew Sardella, Debra Lochner Doyle

Date of comment: 11/20/2013 13:45

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: The policy draft outlines the responsibilities of Primary Investigators to submit resources needed to support proposed data sharing plans in their funding application. Increased competition for research funding and the attempt to obtain the most "bang for your buck" may provide an advantage to applicants who do not require additional resources for data sharing. Research groups with established data sharing programs in place will have the ability to use funding resources for other aspects of the research.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: The policy draft describes the two groups data will be stored under. Open access data are available for public access while the controlled-access data will require secondary researchers to submit a request for access to be reviewed by an NIH Data Access Committee. External and internal demands experienced by the NIH as a result of the tiered approach have the potential to bias the type of research that receives awards by the NIH.

The policy draft encourages informed consent to state that data may be shared broadly. This type of language may be an obstacle for researchers attempting to obtain institutional IRB approval if IRB committees are not familiar with the NIH policy. Education of IRB committees will be essential to reduce barriers.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: The policy draft indicates that only NIH-funded research results are required for submission. However, much research is carried out through collaborative efforts between multiple research groups. The policy fails to outline the association between the NIH-funds dedicated to producing research data and the data required for submission. For example, if NIH-funding is used to support 0.5 FTE of research personnel from a collaborative group, what data are the submitting PI required to release from the collaborative group? Are requirements different if only 0.1 FTE are supported by NIH-funds?

As a note, but not pertinent to the draft policy, as a result of the prevalence of collaborative research, efforts should be made toward collaboration between NIH-depositories and other national and international data depositories.

Commenter: Gerard Schellenberg

Date of comment: 11/5/2013 16:35

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: The proposed language is in some ways more restrictive and in some ways less restrictive than current practice under the Fort Lauderdale model which is often used for these types of data. The Ft. Lauderdale model specifies rapid release of pre-publication date with the stipulation that the investigators who generate the data are allowed to publish first. Others can work with the data but not publish before the group generating the data publishes. This has evolved to a model where rather than first publication as the complete release date, a time period is specified (often one year). The rationale is that to attract excellent scientists to these projects, they should be given the opportunity to publish on their work. At the same time, others can analyze and work with the data and subsequently publish after an appropriate period of time. Since much of the genomics data being generated is big, being able to get rapid access to results for what is often a lengthy analysis period is attractive. This model works, protects the investigators who generate the data, while allowing access by others.

As written, the proposed language in section IV.C.1."

"The NIH will release data submitted to NIH-designated data repositories without restrictions on publication or other dissemination no later than six months after the initial data submission to an NIH-designated data repository, 27 or at the time of acceptance of the first publication, whichever occurs first."

There is a 6-month delay before outside investigators can get access to the data which is really more restrictive than the Ft. Lauderdale model. It does give the investigators who generate the data a period of time to develop a publication. In some ways the two different models are similar in the time frame that the data generating team gets to develop a publication. However, the proposed approach holds data back for six months. As a policy, it really gives a bad impression in terms of our responsibility to the public who fund this work.

The Ft. Lauderdale principles work well, probably should be a formal policy. I am not sure why there is the need to develop something different to replace what currently works well.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Chuck Selden

Date of comment: 11/6/2013 21:23

Comment:

Section II. Scope and Applicability: The scope is reasonable for the importance of the data as a resource.

Section IV.A. Data Sharing Plans: Expectations for data sharing plans appear to be informed by GWAS experience and previous policies and appear to be workable. I did not see any description of enforcement procedures NIH will take if plans are not followed in fact. Once the grant is closed out, how will NIH ensure sharing? Will some grant money be held in escrow until the plan is fulfilled?

Section IV.B. Non-human and Model Organism Genomic Data: no comment.

Section IV.C. Human Genomic Data: no comment

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: As in my comment 2, NIH lists expectations of what should not be done after shared data are obtained. But what will NIH do if data are sent to open web sites or otherwise shared beyond allowed expectations? Or sold? Or if personal identities are found and linked to the shared data?

Section VI. Intellectual Property: no comment

Any other aspect of the draft GDS Policy: no comment

Commenter: Kyle Serikawa

Date of comment: 11/20/2013 20:04

Comment:

Section II. Scope and Applicability: In the appendix A, examples of types of research, many of the definitions are vague and may be subject to interpretation, or seem arbitrary. What is a "gene" or "gene-sized" region? How about an exact KB, which is derived from known SNP frequencies and known needed number of SNPs for de-identification purposes?Similarly, why 10,000 "genes or regions"? Why is 10,000 a magic number? What about organisms that have fewer than 10,000 genes?

Section IV.A. Data Sharing Plans: Should there be a provision for changing ethical standards? If a typical RO1 is 5 years, between application and completion of a project the standards for data sharing may change. Perhaps there should be a suggestion for periodic revisiting the data sharing plan?

Section IV.B. Non-human and Model Organism Genomic Data: Is there any concern about release of information about bioterrorism agents, pathogens, etc?

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Lawrence Siegel

Date of comment: 11/20/2013 22:13

Comment:

Section II. Scope and Applicability: The scope of the data collection and sharing should be quite narrow, initially, and systematically broaden as information becomes integrated.

Section IV.A. Data Sharing Plans: I think it is necessary, provided the plan's parameters are followed in earnest.

Section IV.B. Non-human and Model Organism Genomic Data: Definitely the starting point.

Section IV.C. Human Genomic Data: Something that needs to be collected. It is in the application and interpretation that greatest care needs to be taken.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: There should be a system of identifying those accessing the data and a system of accountability in place to help avoid abuses.

Section VI. Intellectual Property: DNA cannot be considered property.

Any other aspect of the draft GDS Policy: Necessary and beneficial. PLEASE PROCEED WITH CAUTION...

Commenter: Slone Epidemiology Center at Boston University, Lynn Rosenberg

Date of comment: 11/20/2013 14:55

Comment:

RE: NIH draft genomic data sharing policy

It is widely understood that data sharing is necessary to obtain sufficiently large sample sizes for informative results in genomic research. Investigators have been willing and able to share the data they have collected, as demonstrated by the publication of results from numerous collaborative efforts. While promoting data sharing, NIH also recognizes that the investigators who have devoted their time, energy, and intellect to obtaining genomic data should have the chance to consider and publish the results of those efforts before sharing the data with others. I believe that the proposed new rules on the allowable time period before sharing are unrealistic. The proposed time period for investigators to be able to analyze their data, discuss findings with collaborators, try different models, and write and submit a manuscript—6 months-- before sharing their data would require an all-out effort in a time period too short for proper thought, analysis, and interactions among the collaborating investigators. Imposition of the new rules would likely result in shoddy research resulting from poorly cleaned datafiles and errors in the analyses. A more appropriate time period would be 12-18 months.

Lynn Rosenberg, ScD
Slone Epidemiology Center at Boston University
1010 Commonwealth Avenue
Boston, MA 02215
Email: lrosenbe@bu.edu
PH: 617 734-6006
FAX: 617 738-5119

Commenter: Douglas Stewart

Date of comment: 11/19/2013 12:16

Comment:

Hi,

I read and reviewed the recently proposed NIH Genomic Data Sharing Policy. As part of the request for comment, I share my thoughts below.

I have been affiliated with the NIH intramural program for 9 years. Prior to that I worked in an extramural lab at a large research university. I have spent my entire career working on genetics and genomics. There are three areas of concern to me in the new policy:

1) *Cost.* The policy as proposed is ambitious in the degree to which genomic data is shared. The costs in time and infrastructure to store, annotate and transfer the massive files generated by exome and whole-genome sequencing are substantial. I am concerned that the requirements to fulfil the GDS policy amount to an "unfunded mandate" on investigators and institutions by the NIH. Fulfilling that mandate will come at the cost of continuing to do science, especially in the budget era in which we live. I am not an expert in this matter, but my impression is that NCBI is not adequately funded or structured to handle to deluge of data that will descend on them once the policy goes into effect.
2) *Acceptability to research participants*. As the GDS policy notes, language to sharing next-generation data needs to be included in consents. Whole-exome and whole-genome data are different in their scope and implications for privacy, something that in my experience many lay people recognize even prior to a formal consent process. Even with the use of data access committees I am concerned that the proposed GDS policy might act as a disincentive to enroll in studies.
3) *Burden on investigator*. Aside from the infrastructure and institutional costs noted in #1, there are considerable additional burdens placed on the individual PI by this policy. From experience, I know that submitting data to a publically-available website is time consuming and labor-intensive, even for SNP/GWAS data. The problem is magnified many fold given the scope of whole-exome and whole-genome data. I am concerned that the time cost to submit these data will be substantial and may require hiring extra staff to do so. Who will pay for these costs?

I have benefitted from the long-standing NIH culture to openly share data. I agree that a mechanism needs to be in place to share next-gen data. However, I believe that the broad implications of the policy, as currently written, are unrealistic without substantial extra funding to implement it successfully.

Thank you

Douglas Stewart

-

Douglas R. Stewart, MD
Investigator, Clinical Genetics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute
9609 Medical Center Drive, Room 6E450 MSC 9772
Bethesda, MD 20892-9772 (US Mail)
Rockville, MD 20850 (FedEx, deliveries, visitors)

Office: (240) 276-7238
Fax: (240) 276-7836

Lab: (301) 496-0927 (Alex Pemov, NIH Main Campus, Bethesda)

Email: drstewart@mail.nih.gov

Commenter: Robert Stewart

Date of comment: 9/25/2013 17:38

Comment:

Section II. Scope and Applicability: Genetic research shouldn't be allowed to be private when a new SNP or haplogroup is found. Keeping private the name of the original person whose DNA found the new SNP/haplogroup is fine, but they shouldn't have the option to keep the result private.

Section IV.A. Data Sharing Plans: If government money was used to fund it, private companies and educational institutions should not be allowed to charge for it.

Section IV.B. Non-human and Model Organism Genomic Data: See above.

Section IV.C. Human Genomic Data: See above.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: See above.

Section VI. Intellectual Property: See above. But additional, copyright needs to be revised to make it more in line with what our Founding Fathers wanted and less what Sonny Bono and Disney wanted.

Any other aspect of the draft GDS Policy: See above.

Commenter: Katherine Sutherland, MD

Date of comment: 11/15/2013 22:20

Comment:

Section II. Scope and Applicability: I would like to see the scope expanded even further to include data produced by privately funded research organizations. I also think any data generated by clinical testing of patients should be submitted to national databases. Maybe you should give an "award" that researchers could site in their publications or hang on their walls for voluntarily submitting data.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: I strongly favor collection of extensive genotype/phenotype databases. As a clinician, it is clear to me that we will not be able to use genomic data for clinical purposes until we are able to interpret the meaning and interactions of a huge number of variants. The massive amount of data needed to reach reasonable conclusions requires sharing of data in a centralized database. This is a necessary step to advance clinical research. In my opinion, potential privacy concerns are outweighed by the promise of improved health/healthcare for a significant number of the population. I think a Certificate of Confidentiality is a great idea.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Ileen Sylvester

Date of comment: 11/20/2013 20:41

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans: We request that specific language be added pertaining to tribal laws or regulations that prohibit the submission of data to NIH-designated data repositories. Furthermore, we request that language be added that NIH extramural investigators obtain in letters of support in the proposal from community partners the understanding that deidentified data sharing will occur in all NIH-funded research and reference the final NIH Genomic Data Sharing Policy.

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Pertaining to Article 4, Informed Consent, as well as Article 5, Institutional Certification, we request that specific language be added to the listing re: IRB, Privacy Board, or equivalent group to include in the listing tribal authorities as American Indian and Alaska Native tribes have the right to establish regulations and laws. Pertaining to Article 7, Exceptions to Data Submission Expectations, we request that specific language be added pertaining to tribal laws or regulations that prohibit the submission of data to NIH-designated data repositories.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Pertaining to Article B, Acknowledgment Responsibilities, we request that all investigators who access the genomic datasets from NIH-designated data repositories be expected to acknowledge in all resulting oral or written presentations, disclosures, or publications the populations, communities, or groups of individuals who consented to provide their genomic data for research. In instances when no consent was obtained, such as, for example, when a waiver of written or oral consent was issued by an IRB, investigators should be expected to acknowledge that no consent was required or obtained for collection of the data used in the research, as well as acknowledge the people from whom data were collected without consent.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: It is the expectation from Southcentral Foundation that any researcher using data derived from a study conducted originally with tribal members would receive tribal approval of secondary data. Furthermore, Southcentral Foundation requires that researchers provide any publication of research findings or presentation of research findings using data that initially required tribal approval to receive tribal review and approval from the same tribal organizations.

Commenter: Philip Taylor

Date of comment: 11/17/2013 8:17

Comment:

Section II. Scope and Applicability: The underlying rationale for this effort is laudable. However, my primary and overriding concern for the impact of this extremely broad policy change relates to the effect on investigators who will be required to submit genomic information under the proposed GDS. The request may appear simple and uniform, but as one who has already been impacted by the GWAS submission policy and the considerable effort required (personal and institutional) for that effort, this policy will expand prior related requirements enormously, both in terms of the scientists impacted and the volume and complexity of the data which will be required to be submitted.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: I see no criteria for standardization of human data. If data submission is required within an unreasonably short time period, without standardization, without adequate data cleaning (as usually occurs in the course of manuscript preparation), and as an unfunded mandate, this will become a junk in junk out effort. The net result of this could actually end up promoting the use of inappropriate or poorly annotated data that in fact does not advance the scientific mission as intended. Bad data or badly used data can be worse than no data at all.

With funding reductions now a regular occurrence and the standard of future expectation, this means that a significant portion of the shrinking pie will be required to go to the equivalent of more administrative overhead rather than science.

The added complexity of the new consents required for broad sharing of genomic data will affect patient acceptance and response rates, and not in a positive way. We have sufficient trouble with 'informed consent' as is, and this will make an already over-the-top complex process more so. The current complexity and the level of detail in the consent documents mean, in my opinion, that few if any patients today are truly 'informed'.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: Have effects of the already over-burdened DACs been considered? This sounds like the equivalent of a data oversight nightmare to me.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: I remain unconvinced that the volume and complexity of the access process will be manageable. dbGaP recently informed me of an incident in which our GWAS data was inappropriately shared with about a dozen investigators

who had not requested the data. It is apparent that we can not properly handle the data that we currently received, let alone the flood that will be involved under these proposed new guidelines.

Commenter: Sharon Terry

Date of comment: 11/20/2013 19:48

Comment:
Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: Response to the NIH draft Genomic Data Sharing (GDS) policyComment 7: Any other aspect of the draft GDS policyDear NIH Genomic Data Sharing Policy Team,The Genomic Data Sharing Policy provides an excellent opportunity for the NIH to establish methods to engage research participants in a truly participant-centered approach to genomic medicine. The participant consent process provides an opportunity to engage individuals and their families, to explain the intentions of research, and to initiate dialogue about the participant's role in the research process.First, we believe more emphasis should be put on Fair Information Practice Principles, so that the burden of engagement is not placed upon informed consent alone, and particularly not upon a form, rather than a process.Second, we believe that the proposal to adopt a "broad consent" approach undermines the NIH's focus on ensuring that participants are appropriately informed about the research to which they are contributing. NIH wishes to engage participants and the public in a much broader understanding of biomedical research, and those who 'raise their hand' to participate in biobanks, registries and clinical trials are prime stakeholders in this engagement.Instead, we propose that NIH adopts at least a dynamic consent approach, and perhaps a granular (allowing sharing of specific subsets of information) and dynamic consent process. Using dynamic consent will empower participants to understand the potential of the proposed research, improve their level of engagement, and provide input in the process. The recent Institute of Medicine recommendations for the Centers of Translational Science Awards (CTSA) highlighted community engagement as an essential element of the research enterprise. NIH welcomed those recommendations, and it would be inconsistent for the agency to use broad consent instead of dynamic and participatory consent. The deficiencies of broad consent are considerable and well articulated. Participants cannot make genuinely informed decisions when sharing and decisions about secondary use of their data is beyond their reach and control. Broad consent is effectively consenting to have all the important decisions made by other people—its primary effect in practice is to marginalize and trivialize the trust and involvement of donors in research. Dynamic consent will provide an opportunity for researchers to gain participant input as the research field develops and progresses, and will enable participants to receive timely information about the research that is being undertaken.The primary argument for broad consent—that it

relieves researchers from having to engage in expensive, time-consuming recontact and re-consent of participants—is limited by the fact that broad consent does not protect research from changes in law and regulation, from innovative new technologies that permit novel and un-anticipated uses of data, or from changing demands from publics or policy makers. Further, it diminishes the power of the connection between individuals and their data and samples. Only by integrating the whole of the individual, their family and their community into the research enterprise will researchers have the data they need to understand stratified medicine, and the contribution of the environment and microbiome.In an era of participant-centered innovation and increased public engagement in science, research that treats participants as 'subjects' rather than participants, and static paper-based consent models, are becoming increasingly out-dated and unfit for the purpose of patient consent.Dynamic consent is an alternative to broad consent that addresses the changing nature of biomedical research. Dynamic consent maintains and upholds participant respect by actively producing research as an ongoing partnership between participants and researchers. To achieve this, dynamic consent uses information technology to place patients and research participants at the center of decision–making. These technologies are ubiquitous in other sectors, but new to biomedical research. It makes what seemed onerous and impossible in the past, possible and simple. There are advantages in employing a dynamic consent system. 1. This participant-centered paradigm of consent recognizes user autonomy and tailors the experience to meet individual needs. 2. Engaging participants promotes scientific literacy, transparency, and trust in research as participants become more informed about the research carried out on their samples and information. 3. An engaged and dynamic consent process creates an online, responsive, and highly engaged cohort of participants for researchers to contact regarding further studies or further collection of information. 4. A dynamic participatory process allows research governance to respond to changes in law and regulation, new scientific techniques and capabilities, and changing social perspectives by engaging with participants to discuss the changes rather than making assumptions about what patients 'would probably be comfortable with.' 5. It makes the consent process meaningful and allows for nuanced consent choices that avoid the 'all or nothing' flaw of broad consent. 6. In this age of abundant information, an engaged dynamic consent process meets the highest international ethical and legal standards for consent in a world where data protection laws are changing. In 2007, the NIH conducted a public consultation to gather comments relating to the policy for sharing of data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS). The Notice outlining the result of this exercise stated that the 'NIH recognizes that the ethical considerations relevant to GWAS data sharing are complex and dynamic.' Consent was a specific area of concern for respondents, with the Notice stating that efforts to address the complex nature of these issues would include 'discussion of the optimal methods for communicating with participants about relevant issues through the informed consent process for prospective studies.' It also conceded that '[t]he NIH anticipates that a number of GWAS proposing to include pre-existing data or samples may require additional consent of the research participants,' providing a clear example of the difficulty involved in setting up a system of broad consent that adequately caters to future research developments. While this previous exercise specifically focused on GWAS studies, many of the concerns raised are directly applicable to the data sharing issues discussed in the Genomic Data Sharing Policy. The white paper produced by the global alliance states: "Within research, there are a number of participant-centric initiatives (PCIs) that use social media tools, offering new ways to engage with research participants. These can enable on-going communication, allowing individuals to give consent to research, specify

personal privacy levels and to become partners in the research process in ways that have not been possible before. By enabling control over personal information and the potential to give on-going consent in real time, these initiatives meet international legal standards for the protection of privacy. Active engagement with the public and relevant governmental and regulatory officials will be needed to encourage the use of PCI and promote beneficial research while providing adequate privacy protections. In the long term, there needs to be greater transparency in data handling, commensurate punishment for mishandling of data, and governance procedures that include public input…". Renowned experts produced this white paper after much deliberation. NIH is a champion for the centrality of the participant. Broad consent is not participant-centric. We strongly recommend that NIH also champion dynamic consent.Sincerely,Genetic Alliance Council and StaffGenetic Alliance BioTrust Ethics TeamKelly Edwards Seattle, WAJane Kaye Oxford, United KingdomGreg Biggers Palo Alto, CAKieran O'Doherty Ontario, CanadaNick Anderson Davis, CALeila Jamal Baltimore, MDDavid Winickoff Berkeley, CAOrganizations:Basal Cell Carcinoma Nevus Syndrome Life Support Network Burton, OHCardio-Facio-Cutaneous International Vestal, NYGenomera Palo Alto, CAHHT Foundation International Monkton, MDKS&A Pine, COLynch Syndrome International Vacavilli, CAMLD Foundation West Linn, ORPachyonychia Congenita Project Salt Lake City, UTPhelan-McDermid Syndrome Foundation Venice, FLPXE International Washington, DCRASopathies Network USA Altadena, CA

Commenter: Texas Biomedical Research Institute, Shelley Cole

Date of comment: 11/20/2013 17:05

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: A. Most of the concerns expressed in the policy, regarding the release of genomic data, focus on individual privacy and confidentiality. However, a majority of my research involves underserved minority populations, which despite the fact that they tend to carry more of a disease burden, also tend to be understudied and thus lacking in benefits from research advances. My concerns regarding the broad scope of this draft policy are therefore at the community level, which is not addressed in the draft policy.

1. While participants' data are de-identified from personal identifiers, the ethnic and cultural identifiers remain, usually as necessary covariates in analyses (especially analyses which propose to combine datasets for analysis). There are obviously many instances in human history where ethnic affiliation has been used to discriminate, and those who have (or would) discriminate have never needed genomics data to do so. But more subtle forms of discrimination are clearly possible through broad data sharing, including denying affordable treatments and preventions to a defined community because of their genetic differences. Communities that broadly share their genetic data may open themselves up to such discrimination.

2. While the policy speaks to exceptions to data sharing for some studies, these are poorly defined. The exceptions are granted at the discretion of each Institute, which means they may be subject to the politics and priorities of individual Institutes. This could potentially result in inconsistent application of the policy, leaving some disease investigations unfunded because of a denial for an exception that was granted elsewhere. If a risk to a community exists from broad data sharing of genomics data for one disease investigation, then it will most likely exist for another. There should be a consistency across the NIH regarding data sharing exceptions, and it should be defined in this policy.

3. There is a risk associated with broad data sharing of genomics data. If there were not, then there would be no need for de-identification of data, nor controlled access data sharing. In many cases, the participants from minority communities understand that there might be risks and are willing to participate in research for the greater good. But they are not willing to shoulder more than their share of the risk, including risk to their communities. Such communities might have a fear of federal policies and assurances, real or imagined, but rooted in a history of abuse and/or reneging on established promises (e.g. American Indians, African Americans). Because of this history and mistrust of government, participants may feel that they have no choice but to say no to broad data sharing of their genomic data. This would result in further disenfranchisement of

groups that would benefit the most from state-of-the-art genomic research, especially regarding personalized medicine which relies upon distinct genetic markers that may only be discovered by population-specific investigations.

4. The investigator is put in the unenviable position of convincing underserved minorities that they should share their data freely. The NIH's role in promoting human health seems to stop when it comes to discussing the advantages of sharing of genomic data with minority communities. The policy should provide for some direct communication regarding data sharing approaches between NIH and population groups where research results could have the most impact on alleviating the financial burden of disease.

B. My research also focuses on family cohorts, where privacy and confidentiality extend to family members who may not be research participants and may not have signed consents. Family data deserves special consideration, especially family data collected in a well-defined (i.e. geographically) minority population. Family datasets are valuable for genetic studies as they allow the testing of true genetic inheritance in a pedigree. This is becoming more important as genomics investigations attempt to sort out inherited genetic changes from those due to environmental and epigenetic effects. However, family data includes information on potential non-paternities and other differences from self-reported relationships. Even though personal identifiers are removed, pedigree structures and generational information remain, and may allow the identification of specific families from small communities. Family-based cohorts should be exempt from mandated, broad data sharing.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: The policy does not define the consequences a researcher will encounter if s/he violates or does not meet her/his responsibilities when accessing and using genomic data. (During one of the open, public seminars introducing the draft Data Sharing policy, it was stated that steps may include contacting the investigator and notifying the affiliated institutional official, requesting a plan to prevent re-occurrence, and having the investigator stop work with the data and lose access to dbGaP. No regulations were noted.) Contrast that with the threatening consequences, included in the written policy (section I), that are imposed upon a researcher who does not comply with the Data Sharing policy: "Failure to comply with the terms and conditions of the funding agreement could lead to enforcement actions, including withholding of funding, consistent with 45 CFR 74.62 and/or other authorities, as appropriate". More importantly, there is no mention of what avenues research subjects have if their consents and/or privacy are violated. It appears that NIH claims no responsibility in such cases. Therefore, the draft policy is written to more towards forcing compliance with genomic data sharing than in punishing offenders who violate data use agreements, and in protecting the rights of human subjects.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: There is a lack of justification for the need for a new, more expansive NIH Genomic Data Sharing Policy.
1. The original GWAS Data Sharing policy implemented in 2007 was justified due to cost. The federal government was investing millions of dollars in GWAS data generation, and felt the need to promote the use of GWAS data in order to "recoup" the investment. Now that there are

"advances in DNA sequencing and other high-throughput technologies as well as a steep drop in DNA sequencing costs" it would appear that the need to recoup the initial investment is not as great. This would appear to argue for a cancellation, not an expansion, of mandatory data sharing.
2. NIH keeps records on data requests from dbGaP, and reports great interest and use of public data. However, there are very poor records, if any, on research results generated from the use of shared data, especially outlining how they have accelerated research and advanced improvements in human health. NIH relies on investigator self report, and that is inefficient (e.g. Ramos et al., PMC3617375). In fact, I would challenge that most data sharing that is productive occurs as a natural consequence to current scientific approaches; meaning that consortia are formed to harmonize and share their data to generate "large and information-rich datasets" long before the data are available to researchers (not affiliated with the consortium) on public databases such as dbGaP. I am unaware of any large, information-rich dataset that has been formed solely by a third party investigator (not already affiliated with a consortium or projects) that combined datasets through the use of public databases.
3. A significant amount of resources have gone into developing public databases, and preparing data for deposition in databases. Given the lack of justification, it is possible that a more expansive data sharing policy will result in a massive waste of federal research dollars that could instead be spent on funding more research projects that actually do accelerate research discoveries.

The following ideas are poorly defined in the draft policy:
1. punishment/consequences for an investigator who violates or does not meet their responsibilities when accessing and using genomic data (see section V comments)
2. what is meant by aggregate data and does it meet the policy requirements?
3. what will happen to research funding if some research subjects opt out of data sharing?
4. what constitutes exceptions other than country or state laws?
5. Why would legal circumstances (country or state laws) still require NIH to consider whether it would grant an exception? Shouldn't an exception under these circumstances be automatic, since NIH does not have the authority to violate these laws, and it would be discriminatory to a potential research population governed by such laws to deny them benefits from genomics research?

Commenter: Chloe Thio

Date of comment: 11/19/2013 16:24

Comment:

Section II. Scope and Applicability: Data sharing is critical and important for the advancement of all science. The scope of this data sharing plan may be harder to implement and enforce since it includes sequencing of one or more genes in more than 100 participants. Will this type of data sharing be as fruitful or manageable as compared to release of sequence array or exome or whole-genome data?

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data: The proposed 6 month plan for release of data is not long enough especially for smaller academic, based groups, like myself, who have one analyst working on several projects. If people are forced to analyze their data this quickly, then I think the data will be less accurate since less time will be spent on proper QC of the data. it is especially important that investigators have the time necessary to properly analyze and interpret their data before contributing to the literature. I suggest extending this date to 12 months. After 12 months data can be released for data sharing without an additional embargo period. I am concerned about the requirement that all future studies require the broad consent of participating individuals to have their data publicly available and that if they don't agree, they cannot participate. First, I think that few people will understand the implications of this policy, so it won't be informed consent. Even scientist in the field can't predict the future and how such a consent will be interpreted in the future when more genomes are sequenced and we will be able to identify people easier with little genomic data. Second, it may bias study populations and exclude those who are marginalized since they may be less likely to agree to such a broad consent. I think more discussion with citizens and focus groups are needed before such a broad consent is required.

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: David L. Thomas

Date of comment: 11/20/2013 10:47

Comment:

Section II. Scope and Applicability:

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: The intent of this proposal is praiseworthy. There are some concerns that can be addressed. One concern is the 6-month time frame. One of the principle responsibilities of the team is to distinguish between false and true findings. This process requires time to analyze, to test confirmatory panels and, in some instances, to explore functional studies. A six-month time frame is too short. A longer interval will improve the quality of the data released and have a much better net impact on human health. Safeguards of human rights are paramount, and studies on specimens collected after the enactment of this policy should require that genetic consent be specified. However, new research on de-identified tissues that were previously collected using practices that were approved as ethical at the time should not be restricted even if genetic consent was not specified. The rationale for this position is established and forms the basis for the current policy that allows each IRB to consider whether the benefits of a proposed study justify the risks of unintentional disclosure of information. The 'controlled access' mechanism provides ample protection against unintentional disclosure. It is impracticable that this could ever occur without wanton misrepresentation. Access to those controlled data can be more carefully scrutinized and unjustified use more stringently sanctioned. However, retrospective research on de-identified tissues collected in an ethical manner but without specified genetic research should not be restricted. The NIH should provide real, substantive assistance to simplify and assist with this process. Expanding the scope of the reporting requirement to include 'more than one gene or gene-sized region in more than 100 persons' the policy will unintentionally discriminate against small laboratories without dedicated personnel who are dedicated to discharging this requirement. The guideline is disproportionately burdensome to the labs least able to comply. NIH can develop nimble systems to assist in a meaningful way to make this easy, not just for someone who does it monthly but for the person doing it the first time. We have all experienced web based systems that really work and those that are simply possible once one already knows how to use them.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

OFFICE OF THE VICE PRESIDENT - RESEARCH AND GRADUATE STUDIES

OFFICE OF THE PRESIDENT
1111 Franklin Street, 11th Floor
Oakland, California 94607-5200

November 20, 2013

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

[Submitted via e-mail to: GDS@mail.nih.gov]

Re: University of California Comments on Draft NIH Genomic Data Sharing Policy

Dear Genomic Data Sharing Policy Team:

I am writing on behalf of the University of California (UC) system (that comprises research universities at Berkeley, Davis, Irvine, Los Angeles, Merced, Riverside, San Diego, San Francisco, Santa Barbara, and Santa Cruz), and the University of California-managed Department of Energy-funded Lawrence Berkeley National Laboratory, to offer comments on the Draft NIH Genomic Data Sharing (GDS) Policy released on September 20, 2013. The researchers of UC engage extensively in the genomic research and data sharing that are subject of the proposed draft policy.

The University of California shares the goals of promoting maximum public benefit by facilitating broad sharing of genomic data and ensuring responsible oversight of the sharing of such data that were expressed by NIH in issuing this draft policy. We are committed to accelerating research through the power of combining large and information-rich datasets because of the opportunities for discoveries and improved understanding of common diseases that this presents. Researchers across the UC system are experienced in collaborative research and understand the value of cross-studies analyses to address complex questions.

However, there are elements of the Draft GDS Policy that would generate new unfunded administrative burdens on genomic researchers and academic research institutions. Certain aspects of the policy are unclear and would likely make implementation confusing. As further detailed below, UC asks that NIH consider clarifying the Draft GDS Policy to better serve its intended purpose of promoting broad sharing of genomic data to serve legitimate research purposes.

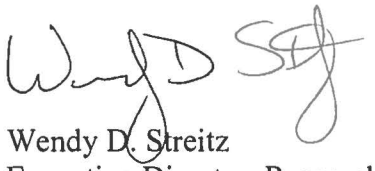
The scope and nature of the review under the Draft GDS Policy creates an additional burden for IRBs. The Draft GDS Policy's requires IRBs to assess research participants' consents for use of their de-identified data from existing cell lines and clinical specimens in order to designate

data as open or controlled access, and to assure that data submission is not inconsistent with those consents are challenging undertakings. Although we currently do our best to undertake these tasks, IRBs will struggle to assess subjects' expectations of privacy regarding their de-identified data in a scientific /technological environment where the ability to re-identify genomic data is so rapidly evolving. It is incumbent on the repositories to share in the responsibility to protect against the misuse of the shared genomic data that they house.

The Draft GDS Policy mandates Institutional Certifications that at most academic research institutions, including UC, constitutes work that will fall to already heavily-burdened IRBs. This burden is exacerbated by the requirement that certification be made prior to the award. With an increasing number of applications being submitted per funded award due to current decreased federal research funding, the large majority of the Institutional Certifications will be undertaken for studies that will not be funded. UC supports the Federation of American Societies for Experimental Biology (FASEB) recommendation that, rather than requiring a data sharing plan and related budget at an early stage, the NIH should "consider adopting a "just-in-time" process that allows investigators to submit basic data sharing plans at the time of proposal submission. Institutional certification and associated documentation should only be required for those proposals recommended for funding. This is commensurate with current policy regarding development and submission of detailed project budgets and has also proven to be successful with human subjects and animal research protocols.

Thank you for this opportunity to comment. We greatly appreciate your efforts to seek stakeholder input regarding the proposed Draft NIH Genomic Data Sharing Policy.

Sincerely,



Wendy D. Streitz
Executive Director, Research Policy Analysis and Coordination
University of California Office of the President

cc: Vice President for Research & Graduate Studies Beckwith
Director Hall

Commenter: Unnamed

Date of comment: 11/20/2013 16:02

Comment:

Section II. Scope and Applicability: The proposed scope and applicability seems reasonable.

Section IV.A. Data Sharing Plans: The proposed data sharing plans seems reasonable.

Section IV.B. Non-human and Model Organism Genomic Data: I have no comment here.

Section IV.C. Human Genomic Data: Sharing of data generated from research is appropriate and important for moving the science forward. However, the proosed timelines of sharing the data, 3 months after generation, is not at all reasonable. First, teams do not necessary have full time staff to work on these large data. Second, institutions may not have the computing resources to analyze the data within 3 months time. Third, investigators do not have full effort on a project to evaluate results within this timeframe. Fourth, there is a major concern of being scooped by institutions who do have Major computing/ FTE staff to evaluate genomic data quickly (e.g., industry). As a result this will hurt reserachers who are performing good research but working with less resources. A proposed timeline would be to submit the data after the acceptance of the first manuscript using the data.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: I agree with the proposed policy.

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

VANDERBILT UNIVERSITY MEDICAL CENTER

Jeffrey R. Balser, M.D., Ph.D.
Vice Chancellor for Health Affairs
Dean, School of Medicine

November 19, 2013

Lawrence A. Tabak, Deputy Director
National Institutes of Health
Genomic Data Sharing Policy Team
Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Dear Dr. Tabak:

We are writing on behalf of Vanderbilt University and as leaders in the development of Vanderbilt's large DNA repository, BioVU, regarding NHGRI's draft data sharing policy. We urge the NIH **not** to adopt the rules proposed on September 20, 2013, which require explicit donor consent for data sharing before genomic data can be deposited in repositories. We believe this policy would seriously impact the free flow of valuable information among scientists. Our rationale for arguing against adoption is provided in this document.

By way of background, Vanderbilt University is one of the nation's leading academic research institutions and is home to one of the nation's top medical schools. Vanderbilt currently is the recipient of over thirty-five billion dollars in NIH grants, and since 2010 has been ranked among the ten largest recipients of NIH support. Its biomedical research enterprise has long been recognized for important advancements to medicine and science. Vanderbilt University is committed to personalized medicine to improve the quality of care, and has made significant institutional investments in the efficient conduct of genetic and pharmacogenetic studies. One large program is called BioVU, a biobank that contains approximately 175,000 samples. In BioVU, samples are collected when they are about to be discarded within pathology and are de-identified. BioVU fuels important research at Vanderbilt. Researchers have conducted approximately 75 formal, IRB-approved research studies in the last three years. Through use of BioVU, researchers are able to significantly increase their productivity. In one subset of studies focused on capturing serious adverse events related to medications, we determined that researchers were able to generate cohorts needed for research in three months' mean study time as opposed to three years' mean study time for traditional approaches. That time savings means a faster path to reach practice. This saving in time also creates a cost savings which means more research can be conducted with the same funding.

One particular study will highlight the health impact of BioVU. Dr. Kelly Birdwell published a paper last year that used BioVU to confirm that a genetic variation in patients affects the dose requirement for a drug called tacrolimus, which suppresses the body's immune system, thereby preventing rejection in patients receiving new hearts, kidneys and other organ transplants. Tacrolimus has a narrow therapeutic window, meaning that if too little of the drug gets into the circulation, acute transplant rejection may occur, but too much of it can cause serious side effects, including a form of diabetes and squamous cell skin cancer. Dr. Birdwell's study in 2012 confirmed that a single genetic variation accounts for much of the variability in dose requirements. In March

1161 21st Avenue South
D-3300 Medical Center North
Nashville, TN 37232-2104
tel 615.936.3030
fax 615.343.7286
www.VanderbiltHealth.com

2013, Dr. Birdwell's finding was adopted by Vanderbilt's pharmacogenomics testing program, PREDICT – an application of discovery to practice within a one year time frame. Since then, more than 2,800 Vanderbilt patients have been found to carry this genetic variation, more than 600 of whom are adults who have received or are awaiting heart or kidney transplants. Doctors who prescribe tacrolimus receive notifications that they may need to adjust the dose if their patients carry the genetic variation.

Vanderbilt patients are provided with documented notification about BioVU during clinical encounters and are presented with the option to not participate in the program within the patient registration process (15% of patients choose to opt out). For participating patients, DNA is extracted from residual clinical blood samples and then linked to a research-optimized, de-identified database of patient records from Vanderbilt's electronic health record. BioVU's notification required enterprise-scale technical and procedural modifications to the clinic registration process in order to provide and document receipt of the notification via electronic forms and informational brochures about BioVU to clinic patients.

Recognizing the importance of both the ethical issues involved and the public's input and perception, Vanderbilt's effort with BioVU began with and continues to benefit from extensive community input and involvement. There have been 6 separate research efforts conducted among the community, patients, and other stakeholders to ascertain attitudes and preferences, and to assess reactions and responses to the BioVU program. Favorability has always been strong. In fact, we have found consistently in surveys that approximately 90% of Vanderbilt patients and Nashville citizens favor data sharing *even without explicit consent.* They understand and support the value of research.(For example, see Kyle B. Brothers, Daniel R. Morrison, Ellen Wright Clayton, "Two Large-Scale Surveys on Community Attitudes Toward an Opt-Out Biobank," American Journal of Medical Genetics 2011;155(12):2982-90).

Vanderbilt is committed to protecting the interests of its patients, and to that end, has created and applied novel robust software and methods, in addition to the active involvement of a well-defined oversight structure, to protect patient privacy. As a result of the operational model and ongoing support, BioVU has continued to operate for seven years without encountering major concerns from Vanderbilt patients. In this way, BioVU has generated tremendous time and cost efficiencies for research, while upholding and maintaining ethical principles; these methods have been published (see Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clinical and Translational Science. 2010;3:42-8.)

In light of these experiences, stakeholder input, regulatory analysis, and empirical data, we believe that requiring explicit patient consent for data sharing is unfounded for the following reasons:

1. Requiring written informed consent for research use of biospecimens would create increased burdens across the country for little to no enhanced protections for participants. This will be harmful to scientific discovery. Research is currently widely conducted without obtaining prior consent using biospecimens "leftover" from clinical procedures (i.e. tissues that would otherwise be discarded). The loss of access to these specimens would hinder research on a host of disease processes, etiology, and normal growth and development using samples that

have been de-identified; in fact, this regulatory framework already forces the de-identification of the samples, thus protecting patients from harms.

2. Eliminating this pathway and forcing a consent procedure could skew data for large population-neutral research resources, which become less efficient and effective for advancing knowledge regarding specific populations. Moreover, skewed populations can threaten scientific validity.

3. The proposed requirement of consent for data sharing in this policy, while addressing one aspect of governance, has the potential to unilaterally foreclose an array of options that may be needed to address these complex issues. In fact, the appropriate governance of biorepositories and the role of consent have been subjects of much debate. More comprehensive analysis of these issues is currently being undertaken by the Office of Human Research Protection, among others (http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html). These rules must be harmonized to be practicable. Efficient models of scientific inquiry are essential. But any move which forces compliance with a single construct (such as that requiring explicit patient consent for data sharing) is ultimately counterproductive not only because it conflicts with OHRP regulations (see item 5 below), but also because it contradicts what we have heard from our patients, and removes local flexibility in interpreting and applying guidance in a local context.

4. To support local adaptations, we would recommend at a minimum that the provisions that relate to the informed consent process and documents (which must state that a participant's genomic and phenotypic data may be shared broadly for future research purposes) be generalized so that they include multiple forms and methods of patient notification, agreement, understanding, awareness, acceptance, and consent. Alternatively, NHGRI could propose principles of operation rather than specific methods of implementation of these principles.

5. The proposed policy could potentially contradict the federal Office of Human Research Protection (OHRP). For example, it is possible to operate under the provisions of research under CFR §46.102(f)(1)(2) regulating that discarded samples can be used for biomedical research if the clinical data are de-identified. Under certain criteria, there are no human participants as defined by OHRP, and thus there are no consent requirements. If the proposed policy were adopted, NIH would have effectively rejected the deposit and sharing of data that has been obtained in compliance with federal regulations for the protection of human subjects.

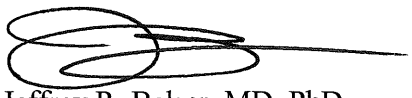
6. The new rules are premature. The NIH has just funded a very large study to ten institutions as a supplement to the eMERGE Network examining factors needed to assess and ensure the acceptability of broad consent. To date, very few data have been collected on what the public actually thinks about broad consent or how to effectuate it or indeed, about whether other approaches would be acceptable. This study will survey 100,000 people (expecting 15-20,000 to provide responses) to ascertain current opinions on these and other topics. We

believe it is premature to issue this policy before the NIH-funded two-year study is complete and patient voices can be heard.

In summary, Vanderbilt investigators collaborate widely and are committed to the free flow of data among scientists while also respecting and complying with the regulatory protections for patients' private health information. Investigators at our institution have deposited de-identified BioVU data in NIH-created databases that have already been reused 192 times for a wide array of research projects, extending the scientific utility of these valuable samples with no additional financial burden. In putting forth a 'one size fits all' approach, NHGRI's proposed policy negates the value of technical methods for securing privacy, overlooks key national experience, and ignores the considerations of local patient and community guidance.

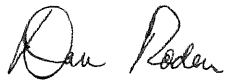
Thank you for the opportunity to comment on this proposed policy.

Respectfully submitted,



Jeffrey R. Balser, MD, PhD
Vice Chancellor for Health Affairs
Dean, School of Medicine

Gordon R. Bernard, M.D.
Associate Vice-Chancellor for Research
Senior Associate Dean for Clinical Sciences



Dan M. Roden, MDCM
Assistant Vice-Chancellor for Personalized Medicine
Director, Oates Institute for Experimental Therapeutics
Professor of Medicine & Pharmacology

Ellen Wright Clayton, MD, JD
Craig-Weaver Professor of Pediatrics and Professor of Law
Co-Founder, Center for Biomedical Ethics and Society

Commenter: Jeffrey Weitzel

Date of comment: 11/21/2013 3:15

Comment:

Section II. Scope and Applicability: In Appendix A, It is not clear why they are excluding smaller studies, i.e., of less than 100 participants or if less than 10,000 genes in one individual. Why not require all if they are requiring any? How was the "10,000" threshold determined?

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data: Some editorials have expressed concern that an individual might be identifiable by their posted genomic sequence data. However unlikely, I don¹t see a way to resolve the potential identification issue. It will be difficult to obtain truly informed consent from research participants. Most genetics/genomics specialists would be challenged to explain much of this to a patient in an understandable way. We already have experience in this area from current clinical testing for cancer susceptibility with NGS-mediated multigene panels.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy:

Commenter: Wellcome Trust Sanger Institute, Giselle Kerry

Date of comment: 11/19/2013 9:35

Comment:

Please find below the comments we would like to make on behalf of the Wellcome Trust Sanger Institute.

"*With regard to the Draft NIH Genomic Data Sharing Policy (Part C Human Genomic Data, Section 4 Informed Consent, Paragraph 2), which presents policy on informed consent and the use of cell lines or clinical specimens, we have consulted with two members of Faculty at the Wellcome Trust Sanger Institute , Hinxton, Cambridge, UK, who have agreed that informed consent for future research use and broad data sharing of cell lines is entirely reasonable and should be sought even for de-identified samples. For studies using data or samples collected before the effective date of the NIH GDS Policy, the consultees agreed that an assessment by an IRB or research ethics committee (REC) should be obtained before samples/data are used in research, and the IRB/REC could make the assessment of whether there was inconsistency between the informed consent and the uses to which the data would be put. In addition, however, for previously obtained samples/data where consent was not explicit at the time of collection, we suggest that an IRB/REC could make an assessment of risk to the research participant and weigh this against the benefit of the research study.*

However, the consultees felt that data obtained from ancient DNA samples (samples from human remains in excess of 100 years old) should be entirely excluded from the Policy and be subjected to a local permission process depending on national policy in the country in which the samples originated, in order to ensure that any ethical issues are dealt with sensitively and in an appropriate context."

Best wishes,
Giselle

Giselle Kerry
Data Access and Regulatory Support Officer
Science Administration
Room D3-04
Wellcome Trust Sanger Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge. CB10 1SA
Tel: +44 (01223) 834244 Ext. 4979
Fax: +44 (01223) 494919

-- The Wellcome Trust Sanger Institute is operated by Genome Rese arch Limited, a charity registered in England with number 1021457 and a company registered in England with number 2742969, whose registered office is 215 Euston Road, London, NW1 2BE.

Commenter: Jeannette Wong

Date of comment: 11/19/2013 16:23

Comment:

Section II. Scope and Applicability: The applicability of the Genomic Data Sharing policy includes a range of sequencing data resources. While some of these technologies are older and more well-established, it may be difficult for research conducted using different sequencing data technologies to adhere to the same guidelines written in this policy. It might be useful to consider establishing levels within the scope and applicability section in regard to how recent the technology became available for research as this can affect the quality of the data and the ability of that data to be submitted according to the deadlines currently recommended in the policy. In addition, although I am not entirely sure that this comment goes here, there is a clause about compliance with the GDS policy as part of the terms and conditions associated with a funding award. I am curious if investigators will be responsible for standardizing sequencing data to meet expectations for submission to the appropriate data repositories, and whether or not funding awards will be adjusted to account for the additional work required to meet these standardization policies. Currently in the GDS policy, there is no mention of how the data would be standardized: file type, file size, minimum variables needed, reference documents, etc. I think clarifying this information would be very helpful for investigators to plan out their budgets and analysis timelines.

Section IV.A. Data Sharing Plans: Again, especially for extramural investigators, it will be important to understand the responsibilities and implications of what quality of data is expected to be submitted to the data repositories. Those expectations will serve as important guidelines for how investigators address the funding needed to meet these demands. In addition, while investigators can contact appropriate officials to discuss expectations needed to conduct a study with genomic data, it would also be helpful to have an online source for future investigators. This online source would serve not just as a home to the final draft of the GDS policy, but also to discussions on policy updates for new sequencing technology and to provide help to investigators on other questions. Especially at its initiation, it will be useful to have additional support in the implementation and clarification of GDS policy.

Section IV.B. Non-human and Model Organism Genomic Data: There is some concern aboutthe required submission of nonhuman and model organism data by the date of initial publication. Is it the policy's expectation for the data to be submitted to a repository by that date and be publicly available at that time? The timeline for data submission seems a bit short given the amount of data cleaning that occurs and, in order to meet the expectations outlined in the current draft of the GDS policy, this could affect the quality of the data submitted to these repositories. The policy needs to consider the time needed to clean and standardize data so that all data submissions to the same respository are of the same quality and can be easily combined with other similar data. Currently, there is no information provided in the policy on the time, budget, and personnel allocated to regulate data quality and standardization measures.

Section IV.C. Human Genomic Data: The same issues I have presented in Section IV.B should also be considered for this section. In addition, the requirement that studies with human genomic data be registered in dbGaP no later than when data cleaning occurs presents an additional concern. Data cleaning is an ongoing process throughout the analysis, and the data will go through several iterations of quality control throughout the analysis. Submission of multiple versions of the data presents an issue in regard to monitoring the quality of the data that other investigators may be utilizing in their research. The GDS policy needs to outline the resources needed for monitoring the quality of data submission. Furthermore, the release of data no later than six months from data submission also presents a problem for sequencing efforts that are ongoing. Completion of all data included in the analysis may take more than six months and presents a conflict with the proposed GDS timeline.

Section V. Responsibilities of Investigators Accessing and Using Genomic Data: How will the NIH Data Access Committees conduct consistent decisions related to requests for controlled-access data? What resources and personnel are required to meet the potentially high level of demand for access to this data?

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: The complexity of research with human genomic data requires consideration for additional resources necessary to provide standardization measures for sequencing data submission and quality. These resources will also be required as these databases need data needs to be edited, updated, or archived. In addition to resources required to maintain the quality of the databases, it is not entirely clear what additional resources will need to be allocated as part of funding awards so that investigators can meet the expectations of the GDS policy. Careful and thoughtful consideration of the time, personnel, and budget required to effectively execute this policy needs to be outlined.

Commenter: World Privacy Forum, Pam Dixon

Date of comment: 11/15/2013 13:03

Comment:

Section II. Scope and Applicability: 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.

Section IV.A. Data Sharing Plans:

Section IV.B. Non-human and Model Organism Genomic Data:

Section IV.C. Human Genomic Data:

Section V. Responsibilities of Investigators Accessing and Using Genomic Data:

Section VI. Intellectual Property:

Any other aspect of the draft GDS Policy: 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 mandatory 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 all 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 PreservationThe 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 WithdrawalThe 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 ChangesIt 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 DixonExecutive Director, World Privacy Forum www.worldprivacyforum.org