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