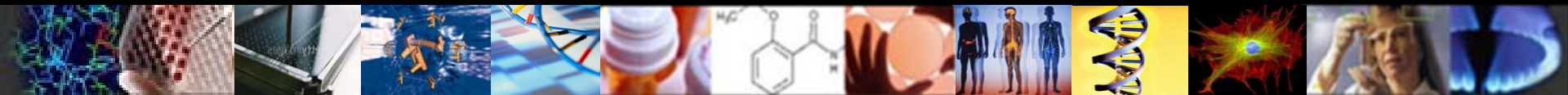


# NIH Wide Policies on Sharing of Data



**Susan B. Shurin, M.D.**  
**Deputy Director,**  
**National Heart, Lung, and Blood Institute**  
**June 7, 2006**



# Guiding Principle:

---

**The greatest public benefit will be realized if data from GWAS are made available, under terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of investigators.**

# Existing Data Sharing Policies

---

- **The existing NIH-wide policy targets wide sharing of data in projects with direct costs exceeding \$500,000 in any given year. Investigators must outline plans for data sharing.**
- **A large number of proposals are now being received for genome-wide studies to link single nucleotide polymorphisms to specific phenotypes.**

# Objectives of Re-examination of Data Sharing Policies

---

- **Extend existing policy to “expedite the translation of research results into knowledge, products, and procedures to improve human health” by making data from GWAS widely available;**
- **Ensure maximum public benefit and avoid duplication of studies, which make relatively large demands upon limited resources;**
- **Encourage rapid and consistent sharing of data, regardless of which Institute or Center provides the funds;**
- **Attempt standardization of both genotypic and phenotypic data submitted to permit comparisons across studies and diseases;**
- **Open a public dialogue with the many constituencies affected by decisions to share such data widely, to ensure transparency, protect privacy, and obtain maximum public benefit.**

# Hot button items

---

- **Existing policies:**
  - NIH-wide apply to >\$500K
  - IC specific – variable in specifics
  - GAIN –FNIH, contracted, NHGRI does the genotyping
- **Can the ultimate identifier (genotype) be de-identified?**
- **How do we balance the sometimes competing interests of individuals, groups and society?**
- **NIH is not a regulatory organization. What can/do/should we mandate?**
- **How can we optimally facilitate communication and change so to gain wide acceptance of what may be in the public interest?**

# What the Policies Will Address

---

- Expectations for submission of GWAS data: timeliness, repositories, format.
- Access to GWAS data: who, for what purpose, how do we protect consent and privacy.
- Human subjects implications of submission and access
- Publication of results based upon GWAS data: should investigators have a head start?
- Intellectual property rights derived from GWAS data: when is an association obvious?

# Study Participants: Consent

---

- **Consents are often quite specific , and do not include sharing of data**
- **Sharing data may have implications beyond the participants: relatives, ethnic groups, members of communities, others affected by disease**
- **Initial consent for existing studies may not permit data submission, even when performance of genetic studies is included**

# Study Participants

---

- **Most want to ensure maximal impact from their participation, maximal benefit to themselves and family members**
- **Privacy concerns:**
  - Can I be identified?
  - Will I/my family/community be compromised?
  - Insurance/discrimination
  - Forensic use
- **Communities/ethnic groups: who speaks?**



# Study Participants: Human subjects

---

- **Does submitting/using this data constitute human subjects research?**
- **No, per OHRP, if data are:**
  - **Fully de-identified (stripped of all 17 personal identifiers used to define human subjects research)**
  - **None of investigators having access to the data will have any way of identifying the participants**

# Protected Health Information-I

---

- **Names**
- **All geographic subdivisions smaller than a State**
- **All elements of dates (except year) for dates directly related to an individual**
- **Telephone numbers**
- **Fax numbers**
- **Electronic mail addresses**
- **Social security numbers**
- **Medical record numbers**
- **Health Plan beneficiary numbers**

# Protected Health Information-II

---

- **Account numbers**
- **Certificate/license numbers**
- **Vehicle identifiers and serial numbers, including license plate numbers**
- **Device identifiers and serial numbers**
- **Web Universal Resource Locators (URLs)**
- **Internet Protocol (IP) address numbers**
- **Biometric identifiers, including finger and voice prints**
- **Full face photographic images and any comparable images**
- **Any other unique identifying number, characteristic, or code**

# Where Will the Data Go?

---

- For optimal user accessibility and integration, a single database should be created and maintained.
- Goal: NIH-supported databases maintained by the NCBI or a similar data repository.
- Interim or longer:
  - existing data are in multiple sites
  - Reluctance to submit/share for multiple reasons

*(tentative, under discussion)*

# Implications for Investigators

---

- **Submitting:** Encourages all investigators who receive NIH support to conduct high-density, genome-wide genotyping to submit curated and de-identified phenotype, exposure and pedigree data. Initial or follow-up consents will need to be compatible with such submission.
- **Access:** Encourages extensive data mining and sharing of potential benefits of data. What limits will be imposed, who decides, and how are these limits managed to protect integrity?

# Data Access: "Approved User"

---

**An investigator who is provided access to a GWAS because the investigator and a responsible official at the institution at which the investigator is employed have both signed the Data Distribution Agreement for that GWAS.**

***(tentative, under discussion)***

# Access: Questions

---

- **Should Approved Users have IRB approval for studies?**
- **How can ICs assist IRBs to address the implications of data submission, adequate consent, and liability?**
- **Could data be used to identify individuals?**
  - Data will be subject to FOIA.
  - Forensic applications are possible.
- **Process for public comment**
  - How do we incorporate comments, with interests/needs/views potentially disparate?

# Publication Policy

---

- **Should the investigators who submitted the data will retain the exclusive right to submit publications based on the submitted data for a period of time? If so, how long, and how will this be regulated?**
- **Approved Users could have access to the data, but are expected to refrain from submitting for publication any results or analyses derived from the use of the data for the period of exclusivity.**

***(tentative, under discussion)***



# Implications for Institutions

---

- **Data submission: data will only be included in the database resource if the responsible IRB has certified that their inclusion**
  - is consistent with the study informed consent
  - does not pose an undue threat to the privacy and confidentiality of the original research participants.
- **How does the institution/IRB learn about or monitor these things?**
- **Should there be a central IRB?**

# Implications for Institutions

---

- **Who keeps the key to the identity? Is this left to individual investigators?**
- **What liability does the institution have?**
- **How does an institution determine what the risks are to participants with data submission or data access, since all that is out of the control of the institution?**
- **Could the institution lose valuable IP?**

# When will the NIH Policy be Issued in Final Form

---

2006 ..... 2007

Trans-NIH  
Discussions

Request for  
Information (RFI)  
Federal Register

Public Broader  
Comment

Final  
Policy

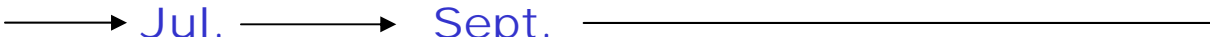
Mar-May

Jul.

Sept.

Dec.

Jan-Feb.



## Notice posted May 15, 2006, in NIH guide

---

- **Notifies the NIH investigator community that the agency intends to begin tracking GWAS applications centrally, through use of a GW code (CSR or IC determined).**
- **Announces that NIH will initiate a public dialogue with constituencies potentially affected by plans to broadly share GWAS data.**
- **The public consultation will seek input on issues relevant to ensuring the protection of research participant privacy, promoting scientific advances across disciplines, and achieving maximum public benefit from investment**
- **The interests involved in access to GWAS data are sometimes competing. Discussion must be very widely based.**

# Next Steps:

## Short-term (next 3 months)

---

- Working group, together with OER and OGC, will develop a Notice regarding a NIH Plan for GWAS.
- The Notice will provide guidance on the overall NIH policy direction, the addition of phenotypes to genotypes, technical standards for genotypes, and DEC and for relevant studies.
- Existing guidelines regarding privacy and human subjects protection apply.

# Next Steps:

## Medium-term (next 6 months)

---

- Working group should refine the draft proposal, (NIH Policy for GWAS), taking into consideration IC, OGC, and OER comments.
- Engage a broad consultative process to discuss the creation of a central NIH database, “In Silico Biobank.”
- Working group further refines the NIH Policy for GWAS based upon public input.
- Implementation

# Next Steps:

## Long Term (6-12+ months)

---

- **Strategic plan for revisions in privacy rule.**
- **Continued strategic communication with the public and scientific community regarding issues surrounding the NIH database, such as privacy protection, IP, access, etc.**

# Public consultation process

---

**An NIH group is refining a draft policy and preparing a request for information to gather public commentary from multiple constituencies:**

- **Investigators: potential suppliers and users of data**
- **Universities, academic medical centers and independent research institutes**
- **Professional societies**
- **Patient advocacy groups**
- **Public interest and privacy groups**
- **Industry groups**
- **Other governmental agencies (FDA, OHRP, others)**
- **Others as identified**