NIH Genomic Data Sharing Policy:
An Overview & Update for IRBs

Laura Lyman Rodriguez, Ph.D.
Division of Policy, Communications, and Education
National Human Genome Research Institute

Columbia 8th Annual IRB Conference
May 12, 2014
“We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals.”

- NIH 2003 Data Sharing Policy
Genomics: A Culture of Sharing

Sharing Data from Large-scale Biological Research Projects:
A System of Tripartite Responsibility

Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale, USA.

OPINION

Prepublication data sharing
Rapid release of prepublication data has served the field of genomics well. Attendees at a workshop in Toronto recommend extending the practice to other biological data sets.
Values Influencing Data Sharing

Genomics
- Public Domain
- Rapid Access
- Communication

Community Resource Projects

Human Subjects Research
- Protection
- Individual Autonomy
- Distributed Burden

Institution by Institution
Project by Project
Finding Common Ground

Genomics through Human Subjects Research

- Participant Respect
- Transparency
- Advance Research
- Data Aggregation
- Improve Health
The NIH GWAS Data Sharing Policy

- Establish community resource to promote maximum public benefit
- Consistent participant protection policies and practices
- Rapid and broad pre-publication data sharing
- 12-month publication embargo for PI to publish primary analyses
- Expectation that primary data will not be subject to IP protections

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)

AGENCY: National Institutes of Health, HHS.

ACTION: Notice.

Background

The NIH is interested in advancing genome-wide association studies (GWAS) to identify common genetic factors that influence health and disease. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood...
Guiding Principle for Genomic Data Sharing

The greatest public benefit will be realized if data from genomic studies 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.

• Respect for Participants
• Freedom to Operate
• Data Sharing
Genomic Data Management Overview

Data Collection

Submission & Management of Data

Distribution & Secondary Use of Data

Research Participants

- Informed Consent

Submitting Investigators

- NIH Genomic Data Repository

Recipient Investigators

- Data Access Request for Coded Data

Data Use Limitations

- Identifying information removed, replaced with random unique code
GWAS Data Access is Two-Tiered

Genotype & Phenotype Data

dbGaP—NIH Genomic Data Repository

Public Access

Study Protocol Descriptive Information

Controlled Access

Coded Genotypes Phenotypes Pre-computes

All potential users

Data Access Committee

• Review data use limitations

Requested Research Use

• Co-signed by institution
• Agree to terms of use
• PI agree to Code of Conduct
Data Access Use (2007-2013)

Total Studies Available in dbGaP = 432

Data Access Requests (DARs) = 19,982
Project Requests = 4,384

Most Common Research Uses
• Statistical methods research
• Software development
• Developing medical therapies
• Basic scientific investigations

Secondary user publications = ~251/yr
(2010-2013)

Data Compiled by NIH Office of Science Policy
Trends in Data Requests and Usage

Data Compiled by NIH Office of Science Policy

<table>
<thead>
<tr>
<th>Year</th>
<th>DARs</th>
<th>Approved</th>
<th>Downloads</th>
<th>Number of Approved DARs (63%)</th>
<th>Number of Downloaded Datasets (~54%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trends in Data Usage

- 2,231 investigators and 5,400 collaborators from 41 countries approved (7/2007 – 11/2013)
- > 500 organizations across the research community

Data Compiled by NIH Office of Science Policy
Data Management Incidents (DMIs)

Project Requests
4,384

27 Total DMIs

• Unapproved Research or Data Access = 10
• Data Security Incidents = 3
• Publication Embargo Violations = 6
• Data Submission Incidents = 7
• Improper Acknowledgement of Data Use = 1

Data Compiled by NIH Office of Science Policy
Need for a Broader Policy

- Technology advances and dramatic drops in cost create opportunity for genomics (beyond GWAS) to be part of many studies
  - The data generated are far richer than what a single investigator or a collaborative team can fully explore
  - Cross-study analyses are possible, which increases the capacity to address complex questions

- Extend participant protections beyond GWAS to other types of genomic data

- Desire for an overarching policy framework that promotes consistent and robust data sharing can best serve the public
Moving Forward with Genomic Science

- Proposed new policy announced September 20, 2013
- Public webinar held November 6th (~200 attended)
- Comment period closed November 20, 2013

Parameters that should be routinely assessed in toxicology studies for INDs, NDAs, and BLAs that are designed to determine the potential for a drug to disrupt the endocrine system. This draft guidance also discusses factors that should be considered in determining the need for additional studies to characterize potential endocrine disruptor properties of a drug.

This draft guidance is being issued consistent with FDA’s good guidance practices regulation (21 CFR 10.115). The draft guidance, when finalized, will represent the Agency’s current thinking on nonclinical evaluation of endocrine disruption potential of drugs. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative

DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

Draft NIH Genomic Data Sharing Policy Request for Public Comments

SUMMARY: The National Institutes of Health (NIH) is seeking 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.

DATES: To ensure that your comments will be considered, please submit your response to this Request for Comments no later than 60 days after publication of this notice.

Sharing of data generated through genome wide association studies (GWAS), which examine thousands of single nucleotide polymorphisms (SNPs) across the genome to identify genetic variants that contribute to human diseases, conditions, and traits. To facilitate the sharing of genomic and phenotypic data from GWAS, the NIH created the database of Genotypes and Phenotypes (dbGaP) with a two-tiered system for distributing the data: Open access, for data that are available to the public without restrictions, and controlled access for data that are made available only for research purposes that are consistent with the original informed consent under which the data were collected.

Not long after the GWAS policy was issued, advances in DNA sequencing...
Outline of Draft GDS Policy

- Scope and applicability
- Responsibilities for investigators submitting genomic data
  - Data sharing plans
  - Data submission expectations and timeline
  - Data repositories
  - Informed consent
  - Institutional certifications
  - Exceptions to expectations
- Responsibilities of investigators accessing and using data
  - Requests for controlled-access data
  - Acknowledgment for use data
- Intellectual Property
Scope and Applicability

- All NIH-funded research that produce large-scale genomic, metagenomic, epigenomic, or transcriptomic data:
  - Tens of isolates from infectious organisms
  - More than one gene or gene-sized region in more than 100 participants
  - More than 10,000 genes or regions from one participant
  - More than 100,000 variant sites in more than 100 participants
- Applies to all funding mechanisms without regard to costs
- NIH will periodically review and make changes as necessary
Effective Date of the Final GDS Policy

- Effective 60 days after publication of final
- Implemented for research proposals submitted in 2015 for FY2016 funding
Responsibilities for Submitting Data

- PIs seeking funding should contact appropriate NIH Project Officials to discuss expectations and timelines for sharing.

- Plans to comply with the GDS Policy should be included in funding applications.
  - Data sharing plans should include resources necessary to support sharing.

- NIH intramural PIs should contact their IC leadership or OIR for guidance.
## Key Distinctions between GWAS and GDS

<table>
<thead>
<tr>
<th></th>
<th>GWAS Policy</th>
<th>GDS Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope</strong></td>
<td>Applies to human GWAS data</td>
<td>Applies to all genomic data types, human and non-human</td>
</tr>
</tbody>
</table>
| **Consent Standard -- Existing* Collections
d*Before the effective date of the Policy** | If research consent, IRB reviews for consistency. If no research consent exists, data may still be submitted to NIH databases. | Same                                                                      |
| **Consent Standard -- Future* Collections
d*After the effective date of the Policy** | Preamble states “the NIH expects specific discussion within the informed consent process and documentation that participants’ genotype and phenotype data will be shared for research purposes through the NIH GWAS data repository.” | Samples or cell lines should be consented for research use and broad data sharing. Exceptions can be requested. Consent should address whether data should be shared in open or controlled access. |
| **Data Submission**            | Data submitted as soon as quality control procedures are completed          | Timelines vary by data type, but generally as soon quality control procedures are complete |
| **Data Release**               | Immediate data release. 12 month publication embargo                        | 6 month deferral of data release. No publication embargo                   |
Draft GDS Policy – Non-human Data

- Encourages consistent data sharing practices

- Expects sharing to be consistent with current practice and recent Federal policy initiatives
  - e.g., NIH Model Organism Policy, White House initiative

- Current resources and databases will remain the standard mechanism for sharing

- Flexibility for ICs to adjust expectations for different data types or types of projects
  - e.g., microbial data pre-publication, model organism data no later than publication
Guiding Principle for NIH Policy

The greatest public benefit will be realized if data from genomic studies 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.

- Submitting institution and its IRB are expected to review informed consent documents and establish terms for future use.
Data Use Limitations

Informed Consent

Research Participants

Informed Consent

Data Collection

Submitting Investigators

- Standard for data submission: data use “not inconsistent with consent”

- Previously it was implicit that there should be consent for research use (for GWAS)

- GDS Policy proposes an explicit standard of consent for research purposes

- Proposed language allows for exceptions in compelling circumstances
Consent for Research Use under GDS

- Current GWAS Policy implies that there should be consent for research uses, but ultimate decision left to local institution

- Proposed GDS Policy expects (with exceptions) explicit consent for research use for materials collected after policy’s effective date
  - Should discuss open or controlled access

- Proposed policy would “grandfather” extant samples and permit their use under current rules (de-identified, no or not inconsistent with consent)
Informed Consent for Genomics Research

Introduction

Informed consent research projects (research participa-
participant and ongoing explanatory informed consent.

Given the complexity of the project, prospective participants may find it challenging to understand all aspects of the research. Therefore, a collaboration involving individuals with expertise in genetics and genomics is essential.

Explore Information

- Informed Consent
- Special Info
- Additional Info
- Three Cons
- Elements...
Exceptions to Data Deposition

- Policy notes that there will be cases where data deposition may not be appropriate
- Requests for exceptions are to come in within the grant Data Sharing Plan
- To date: ~10 requests granted
  - Limited consent
  - Legal restrictions
  - Localized geographic representation
Expectation for sharing should be broad by default, with specific needs addressed through variable submission expectations.
## Expectations for Data Submission and Data Release

<table>
<thead>
<tr>
<th>Level</th>
<th>General Description of Data Processing</th>
<th>Example Data Types</th>
<th>Data Submission Expectation</th>
<th>Data Release Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Raw data generated directly from the instrument platform</td>
<td>Instrument image data</td>
<td>Not expected</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>Initial sequence reads, the most fundamental form of the data after the basic translation of raw input</td>
<td>DNA sequencing reads, ChIP-Seq reads, RNA-Seq reads, SNP arrays, arrayCGH</td>
<td>Not expected for human data if reads are included in Level 2 aligned sequence file (e.g., BAM) Non-human de novo sequence data</td>
<td>Up to 6 months for non-human data</td>
</tr>
<tr>
<td>2</td>
<td>Data after an initial round of analysis or computation to clean the data and assess basic quality measures</td>
<td>DNA sequence alignments to a reference sequence or de novo assembly, RNA expression profiling</td>
<td>Project specific, generally within 3 months after data generation</td>
<td>Up to 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first</td>
</tr>
<tr>
<td>3</td>
<td>Analysis to identify genetic variants, gene expression patterns, or other features of the dataset</td>
<td>SNP or structural variant calls, expression peaks, epigenomic features</td>
<td>Project specific, generally within 3 months after data generation</td>
<td>Up to 6 months after data submission or at the time of acceptance of the first publication, whichever occurs first</td>
</tr>
<tr>
<td>4</td>
<td>Final analysis that relates the genomic data to phenotype or other biological states</td>
<td>Genotype-phenotype relationships, relationships of RNA expression or epigenomic patterns to biological state</td>
<td>Data submitted as analyses are completed</td>
<td>Data released with publication</td>
</tr>
</tbody>
</table>
Sequence Data Submission Expectations

Level 4: Biological Analysis
- Relating phenotype to genotype
- Data submitted from analyses as completed

Level 3: Basic Analysis
- SNP and Structural Variant Calls
- Project-specific data submission; 6 months privileged access

Level 2: Mapped Data
- Alignment (and Reads)

Level 1: Primary Data Readouts
- Sequencing Reads
- Data submission is not expected

Level 0: Raw data
- Instrument Image Data

Data submission: Timely (enough time for data cleaning & quality control)
Data release: Level 2 and 3, 6 months post-completion of quality control
Publication embargo: None
Responsibilities of Institutions

Provide a signed Institutional Certification assuring:

- The data submission is consistent with laws, regulations, and institutional policies
- The appropriate research uses of the data and any uses excluded are delineated (i.e., Data Use Limitations)
- The identities of research participants will not be disclosed to NIH-designated data repositories
- An IRB has reviewed the investigator’s proposal to submit data
Responsibilities of IRBs

Prior to submission, the IRB reviews the proposal to assure:

- Protocol for data collection consistent with Federal regulations;
- Submission and sharing of data are consistent with the informed consent (or not inconsistent with if grandfathered);
- Risks to individuals, families, groups or populations considered;
- Plan for de-identifying datasets is consistent with the GDS Policy.
Responsibilities for Data Access

- Research Participants
- Submitting Investigators
- NIH Genomic Data Repository
- Recipient Investigators

Data Collection → Submission & Management of Data → Distribution & Secondary Use of Data
Data Use Certification Agreement

There is a common template for all NIH Data Use Certifications (DUCs)

Terms and conditions include that requesters will:
- be responsible for compliance with federal, state, and local policies
- only use the data for the specified research use
- not identify study participants
- not transfer data beyond approved users
- immediately notify the DAC if a security breach occurs
- submit brief annual updates on research and publications
- be identified as an Approved User within the dbGaP
- acknowledge datasets and submitters in presentations
- abide by Code of Conduct
Data Access Committees (DACs)

- Review all requests from the research community for access to dbGaP and annual progress
- Primary question is consistency with data use limitations
- 17 active DACs across 16 Institutes (not 1:1)
  - NHGRI, NCI (2), NHLBI, NIGMS, NEI, NICHD, NINDS, NIAID, NIDDK
  - GAIN, TCGA, NDAR
  - JAAMH, JARDE
  - cDAC
  - NCATS (new)
Top Reasons for DAR Disapproval

- Proposed research use statement not consistent with Data Use Limitations
- Incomplete research use statement
- Research use statement does not reference dataset
- Key personnel or collaborator issues
Annual Report/Renewal Elements

- Summary of research progress
- Proposed plans for further research utilizing currently approved NIH GWAS datasets
- List of all completed or accepted scientific presentations that include (or will include) findings made with the individual-level NIH GWAS data accessed through dbGaP.
- List of manuscripts submitted
- Description of any intellectual property generated as a result of using the NIH GWAS individual-level data
- Summary information on any inappropriate data release incidents or other data security issues
- General comments on process & suggestions for improving dbGaP, study-specific data access, or NIH policy or procedures in general
Supreme Court determined naturally occurring DNA sequences are not patentable in the U.S.

Draft Policy considers basic sequence data and related information (genotypes, haplotypes, $p$-values, allele frequencies) to be “pre-competitive”

Data and any conclusions derived from them should remain freely available
Questions ???

For more information

http://gds.nih.gov
laura.rodriguez@nih.gov or gds@mail.nih.gov
Stewardship & Oversight

NIH Director

Senior Oversight Committee

Technical Standards And Data Submission Steering Committee

Participant Protection & Data Management Steering Committee
# Stewardship & Oversight

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eric Green, M.D., Ph.D.</td>
<td>Co-Chair</td>
</tr>
<tr>
<td>Thomas Insel, M.D.</td>
<td>Director, NIMH</td>
</tr>
<tr>
<td>Amy Patterson, M.D.</td>
<td>Co-Chair</td>
</tr>
<tr>
<td>Story Landis, Ph.D.</td>
<td>Director, NINDS</td>
</tr>
<tr>
<td>Hugh Auchincloss, M.D.</td>
<td>Principle Deputy Director, NIAID</td>
</tr>
<tr>
<td>Sally Rockey, Ph.D.</td>
<td>Deputy Director for Extramural Research, NIH</td>
</tr>
<tr>
<td>Stephen Chanock, M.D.</td>
<td>Chief, Laboratory of Translational Genomics, NCI</td>
</tr>
<tr>
<td>Susan Shurin, M.D.</td>
<td>Deputy Director, NHLBI</td>
</tr>
<tr>
<td>Alan Guttmacher, M.D.</td>
<td>Director, NICHD</td>
</tr>
<tr>
<td>Barbara McGarey, J.D.</td>
<td>NIH Legal Advisor, Office of the General Counsel, HHS</td>
</tr>
<tr>
<td>Michael Gottesman, M.D.</td>
<td>Deputy Director for Intramural Research, NIH</td>
</tr>
</tbody>
</table>