National Institutes of Health
Turning Discovery Into Health

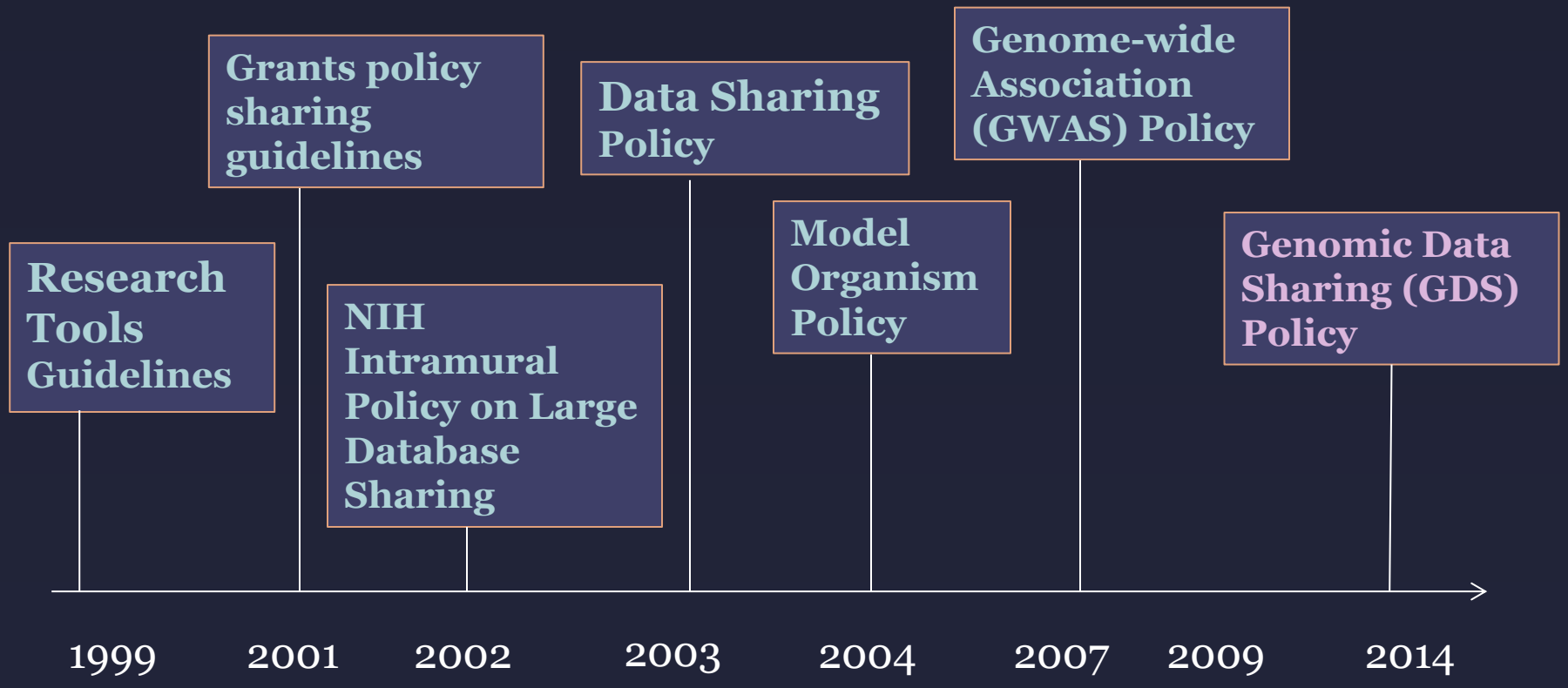
NIH's Genomic Data Sharing Policy

Benefits of Data Sharing

- Enables data generated from one study to be used to explore a wide range of additional research questions
- Increases statistical power and scientific value by enabling data from multiple studies to be combined
- Facilitates reproducibility and validation of research results
- Facilitates innovation of methods and tools for research

Sharing research data supports the NIH mission

NIH: A Culture of Sharing



Guiding Principle of the NIH GWAS Policy

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.

NIH expectation that data would be shared in the NIH database of Genotypes and Phenotypes (dbGaP)

Data Submission and Access Under the GWAS Policy

Data Submission

Genotype & Phenotype Data
(data use limitations specified)

↓

dbGaP—NIH Genomic Data Repository

↙ ↘

- Study Protocol
- Descriptive Information

Unrestricted Access

- Coded Genotypes
- Phenotypes
- Aggregate Data

Controlled-Access

Data Access

All potential users

↓

Data Access Request

- Co-signed by institution
- Agree to terms of use
- PI agrees to Code of Conduct

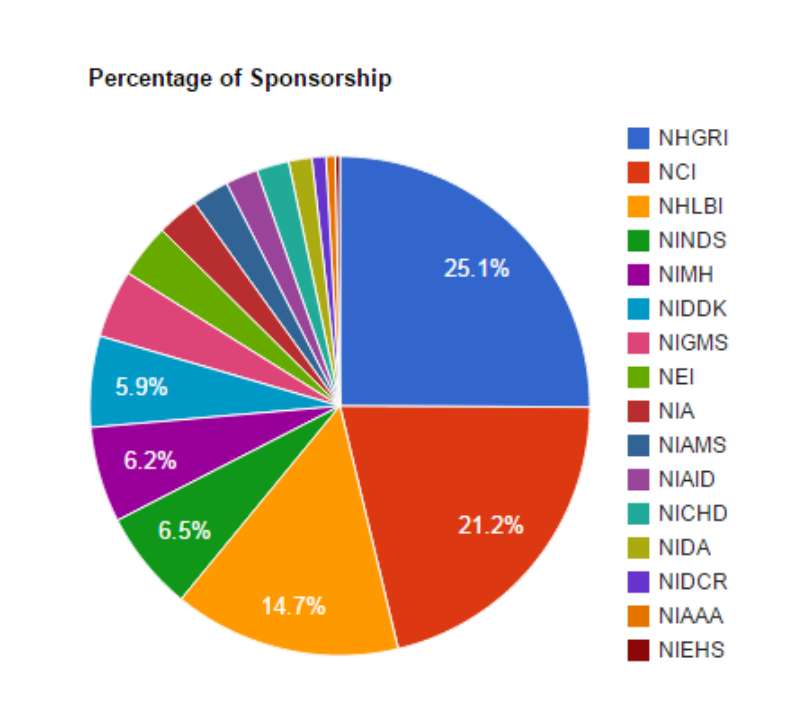
↓

Data Access Committee

- Review data use limitations

Data Submission, Access, and Use Statistics

NIH ICs Sponsoring dbGaP Studies (currently 550)



Data Access and Use Snapshot (since 2007)

30374= Number of Data Access Requests submitted
20272= Number of Data Access Requests approved

920+ = Number of publications resulting from secondary use of dbGaP data

3597 = Number of PIs requesting data
41 = Number of PI countries

1,000,000+ = Number of participants represented In dbGaP studies, collectively

In an effort to increase transparency, NIH is providing information and statistics on data submitted to dbGaP and the subsequent use of that data on the GDS website ("Facts and Figures" tab):
Link to the Facts and Figures tab on the NIH website.

Reasons for a New Policy

- Generation of larger volume of genomic data funded by NIH
- Calls for expanded data sharing from the private and public sectors
- US Federal Government Initiatives
  - February 2013 **White House initiative** to increase access to publications and the results of federally funded scientific research
  - **Proposed Common Rule revisions** (July 2011) supports broad consent to maximize utility of bio-specimens and data
  - **NIH Big Data to Knowledge Initiative (BD2K)** supports the broad use of digital assets and resources to enhance the utility of biomedical big data and accelerate discovery

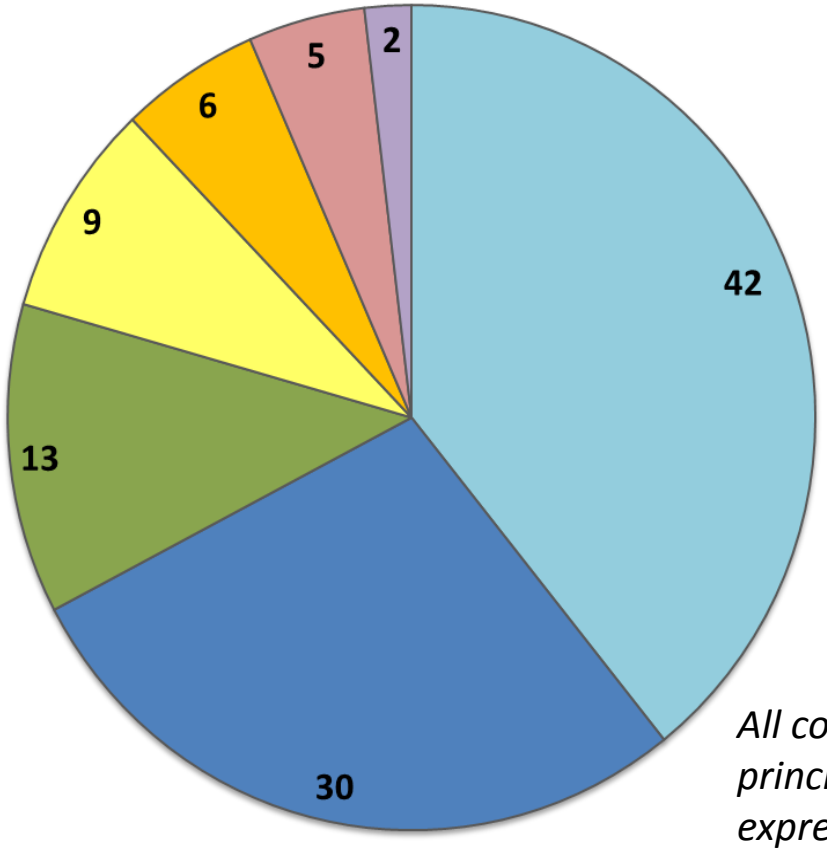
Purpose and Scope of the GDS Policy

- **Purpose**
  - Set forth expectations and responsibilities that ensure the broad and responsible sharing of genomic research data in a timely manner
- **Scope**
  - Applies to all NIH-funded research generating **large-scale human** or **non-human genomic data** and the use of these data for subsequent research
    - Examples of large-scale genomic data include, but are not limited to, GWAS, Single Nucleotide Polymorphism (SNP) arrays, and genome sequence, transcriptomic, epigenomic, and gene expression data
  - Applies to all funding mechanisms (grants, contracts, intramural support)
  - No minimum threshold for cost

Public Consultation Process for the Draft GDS Policy

- Draft GDS Policy was released for a 60-day public comment period: September 2013
- Public webinar and Q and A session: November 2013

Public Commenters on the Draft GDS Policy



N = 107

- Academic Institution
- Research Organization
- Professional Society
- Other
- Industry or Commercial Organization
- Tribal Organization
- Patient Advocacy Organization

All commenters expressed support for the principle of data sharing and nearly all expressed general support for the Policy

Overview of Comments on Draft GDS Policy

- Major comments about specific elements of the Policy:
  - The scope of the Policy is unclear
  - Timelines for submission and release of human data are too short
  - Data submission standards should be provided
  - Additional funds and infrastructure resources will be required for implementation
  - NIH should continue to enable sharing of data for more limited uses
  - NIH should acknowledge risks of re-identification
  - Penalties for misuse of controlled-access data should be outlined in the Policy
  - Tribal sovereignty should be acknowledged in the Policy
- Draft Policy was revised to reflect the public comments and discussions with NIH GDS governance and leadership

The Final NIH GDS Policy

- Final NIH GDS Policy was released August 27, 2014
- The GDS policy became effective January 25, 2015, and applies to:
  - Competing grant applications that are submitted to NIH for the January 25, 2015, due date or subsequent receipt dates;
  - Proposals for contracts that are submitted to NIH on or after January 25, 2015; and
  - NIH intramural research projects that are generating genomic data on or after January 25, 2015.

Supplemental Information to the GDS Policy

- Provides examples of research within and outside the scope of the GDS Policy
- Refers to resources for data standards
- Provides timeline for data submission and release based on the type of data and processing level

Supplemental Information to the NIH GDS Policy:
Link to Supplemental Information to the GDS Policy

Examples of Large-Scale Genomic Data

Type of data	From	From
Sequence data	More than one gene or region of comparable size in the genome	More than 1,000 participants
Sequence data	More than 100 isolates	Infectious organisms
Sequence data	More than 100 metagenomes	Human or model organism microbiomes
Sequence data	More than 100 metatranscriptomes	Human or model organism microbiomes

Key Distinctions

	GWAS Policy	GDS Policy
Data Submission	Data submitted as soon as quality control procedures are completed	Human: Timelines vary by level of data processing, but generally, as soon quality control procedures are complete Non-Human: Generally, no later than time of publication
Data Release	Immediate data release 12 month publication embargo	Human: Generally, 6 month deferral of data release; no publication embargo Non-human: No later than time of publication

When Planning and Submitting Funding Applications

- Extramural investigators should:
  - Contact appropriate NIH Program/Project Officials to discuss expectations and timelines for sharing
    - Contact the Institute's Genomic Program Administrator with general policy questions Link to the Institute's Genomic Program Administrator
  - Submit genomic data sharing plans in resource sharing plan section of applications and proposals
    - Include resources necessary to support sharing in budget section
- NIH intramural investigators should contact their IC leadership or NIH Office of Intramural Research for guidance

Review of Genomic Data Sharing Plans: Grant and Contract Awards

- During peer review:
  - Reviewers will be asked to comment on the genomic data sharing plan
  - Review of the genomic data sharing plan will not be factored in the score, unless specified in the Funding Opportunity Announcement or solicitation.

- If the award is funded, the genomic data sharing plan will be referenced as a term and condition of award

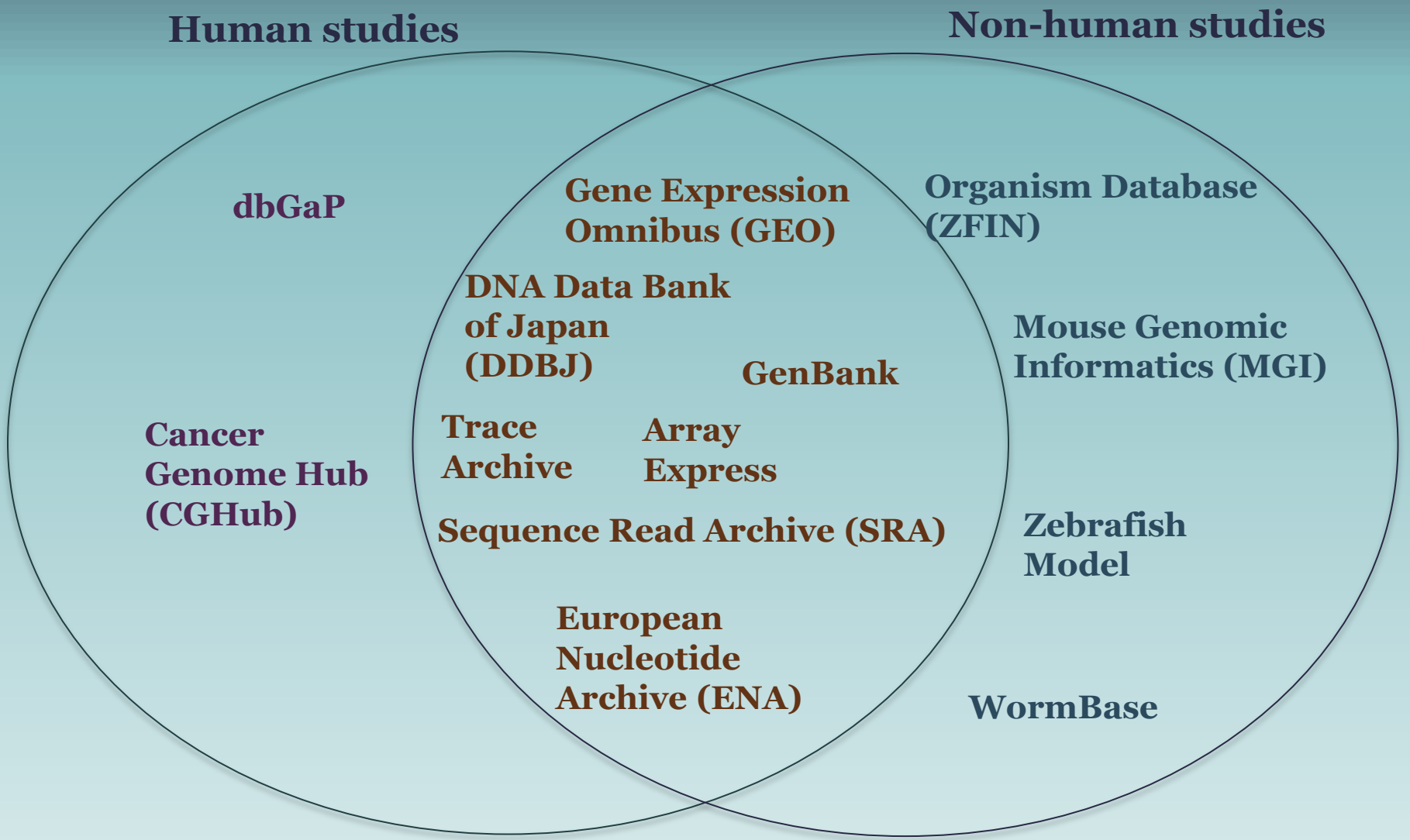
Responsibilities of Investigators Submitting Non-Human Genomic Data

- Standard for sharing non-human data is maintained under the GDS Policy
- Current resources and databases will remain the standard mechanism for sharing (e.g., GEO, SRA, WormBase, GenBank), but data can be submitted to any widely used repository
- Microbiome data – considered non-human data, once the human variant data has been removed

Responsibilities of Investigators Submitting Human Genomic Data

- After the Policy effective date, samples or cell lines should be consented for future research purposes and broad data sharing
- Institutional Certification should be submitted prior to award
  - Letter of assurance from the institution that also includes IRB/privacy board assurance
- Investigators should register studies in dbGaP; data should be submitted to the relevant NIH-designated repository (e.g., dbGaP, GEO, SRA)
  - Data should be de-identified according to HHS Regulations for the Protection of Human Subjects and HIPAA

Examples of Human and Non-Human Genomic Data Repositories



Repositories for Human Genomic Data

- **Unrestricted-access**: data are accessible to anyone via public website (previously referred to as "open access")
  - Example: 1000 Genomes project
- **Controlled-access**: data are available to an investigator for a specific project only if certain stipulations are met
  - Example: dbGaP, Cancer Genomics Hub

Participant Protections

- Respect for and protection of the interests of research participants are fundamental to the NIH's stewardship of human genomic data
- Submitting Institutions and IRBs determine the appropriateness for data to be submitted and shared for secondary research use

Informed Consent Expectations for Studies Initiated *After* January 2015

- Investigators should ask participant's consent for genomic and phenotypic data to be used for future research purposes and to be shared broadly
  - Explanation as to whether the data will be shared via unrestricted- or controlled-access repositories
  - If participant does not consent to broad sharing of data, he or she may still be enrolled in the study, but the data may not be shared

- For studies using cell lines or clinical specimens created or collected after the effective date:
  - Informed consent for future research use and broad data sharing should have been obtained, even if samples are de-identified
  - Exception to consent expectation can be granted, if there are compelling scientific reasons

Expectations for Studies Initiated *Before* January 2015

- Existing studies vary considerably in how sharing is addressed in informed consent documents
  - IRBs should ensure that the proposal for data submission is not inconsistent with the informed consent provided by the research participant

- NIH will accept data derived from de-identified cell lines or clinical specimens created before effective the date that lack consent for research use

When Broad Sharing of Data is Not Consistent with Informed Consent

- Institutions should:
  - Note any data use limitations in the funding application or proposal
  - Specify the data use limitations in the Institutional Certification

Institutional Certification

- The Institutional Certification assures that:
  - The data submission is consistent with applicable national, tribal, and state laws and regulations, and institutional policies
  - Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated
  - The identities of research participants will not be disclosed
  - The IRB has reviewed the investigator's proposal for data submission

Assurance from the IRB in the Institutional Certification

- The protocol for the collection of genomic and phenotypic data is consistent with DHHS regulations on protection of human subjects
- Data submission and sharing are consistent with the informed consent of study participants
- IRB has considered risks to individual participants and their families
- IRB has considered risks to groups or populations associated with the submission
- The investigator's plan for de-identifying datasets is consistent with the standards outlined in the GDS Policy

Provisional Institutional Certification

- Intended to be used as needed on a case by case basis
- When the submitting institution, at the time of Just In Time, is unable to provide assurance to the elements of the formal Institutional Certification
  - E.g., a prospective study where the IRB has not completed its review of the protocol and therefore the institution cannot attest to all of the elements of the formal Institutional Certification.
- The provisional Institutional Certification is submitted until a formal Institutional Certification can be provided

Exceptions to Human Genomic Data Sharing Expectations

- If any of the criteria of the Institutional Certification cannot be met, investigators should provide a justification for an exception to the data sharing expectation in the funding application or proposal
- The funding IC may grant an exception to the expectation for submission of genomic data, and the investigator would be expected to develop an alternate plan to share data through other mechanisms
- The study will still be registered in dbGaP and the exception publicly explained there

Responsibilities of PIs Accessing and Using Human Genomic Data

- PIs must request access in order to download or use controlled-access data
- PIs approved to access data agree to a Data Use Certification cosigned by their Institutional Signing Official
- PIs also agree to abide by the NIH Genomic Data User Code of Conduct
- Approved users of controlled-access data are encouraged to consider obtaining a Certificate of Confidentiality
- GDS Policy also includes conditions for use of unrestricted-access Data

Additional Protections for Human Genomic Data

- To further protect participant's privacy, submitters and users may consider obtaining a Certificate of Confidentiality
  - Issued by NIH to protect identifiable research information from compelled disclosure of any genomic data submitted

Intellectual Property

- U.S. Supreme Court found that naturally occurring DNA sequences are not patentable in the U.S.
- NIH GDS Policy considers basic sequence data (e.g., genotypes, haplotypes, p-values, allele frequencies) to be "pre-competitive"
- These data and any data and any conclusions derived from them should remain freely available without licensing requirements
- Discourages the use of patents to block access to genotype-phenotype data

Transition Provisions for Research Funded Before the GDS Policy's Effective Date

- NIH strongly encourages investigators to comply with the expectations outlined in the Policy
- Investigators should provide an updated genomic data sharing plan to the funding IC in the submission of the research performance progress report
- For human studies, investigators should plan to transition to a consent for future research uses and broad sharing, if possible, particularly for new or additional collections of specimens
- The goal is to bring long term projects into alignment with NIH's expectations for consent and data sharing, to the extent possible, in a reasonable timeframe

For More Information

- NIH Genomic Data Sharing Website
  - NIH Genomic Data Sharing Policy
  - Information for researchers
  - Information for institutions and IRBs
  - Examples of NIH Data Repositories, NIH-Funded Databases, and NIH Database Collaborations
  - GDS Policy Frequently Asked Questions (FAQs)
- NIH database of Genotypes and Phenotypes (dbGaP)
  - dbGaP Submission System
  - dbGaP Access System
  - YouTube Tutorials for dbGaP:
    - Data Submission: dbGaP: Complete a Study Registration
    - Data Access: dbGaP: Apply for Controlled Access Data
- Email: 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
301-496-9838