Sharing Genomic Data :

Managing Variables & Uncertain Risks

Laura Lyman Rodriguez, Ph.D.

Office of Policy, Communications and Education

National Human Genome Research Institute

Data Sharing Strategies for
Environmental Health Sciences Research Workshop
February 6, 2012

NIH and Data Sharing

"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

Putting the Pieces Together

- **Scientific Design**
  - Research aims and objectives
  - Program priorities
- **Policies and Procedures**
  - Applicable laws and regulations
  - Guiding principles
- **Program Implementation**
  - Policy and ethics
  - Transparency
  - Shared responsibilities

Shifting Trends in the Science…

Single Project

Specific Consent

Defined Timeline

Many (unforeseen) Projects

Broader Consent

Governance Models

Undefined (unbounded) Timeline

What we want to avoid ...

The NIH GWAS Data Sharing Policy

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

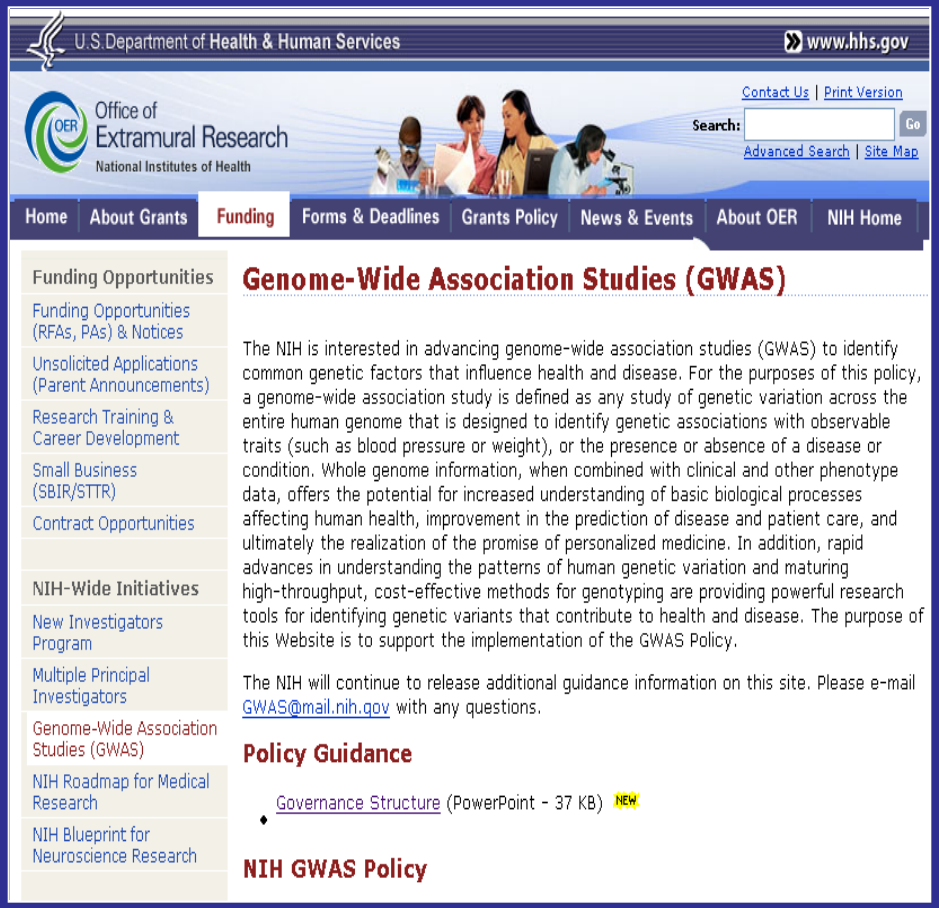
Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)

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



Policy Announced: August 28, 2007
Policy Effective: January 25, 2008

GWAS Homepage:
http://grants.nih.gov/grants/gwas/index.htm

The Scientific Program

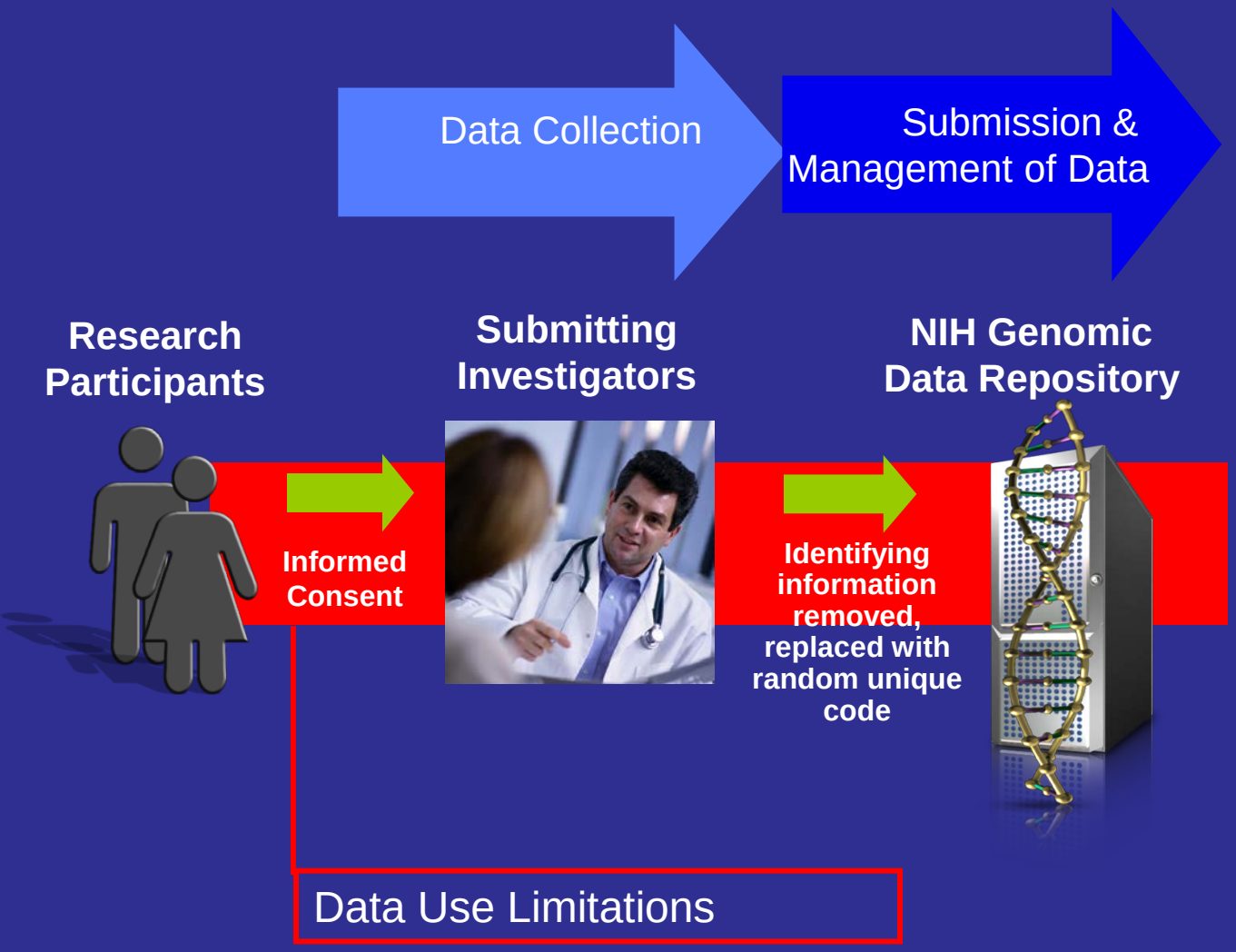
- Whole genome analysis
- Large datasets with lots of genotypes (or sequence), lots of phenotypes, lots of people
- Broad data sharing
- Government database

Guiding Principle for NIH Policies

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
- Data Sharing
- Freedom to Operate

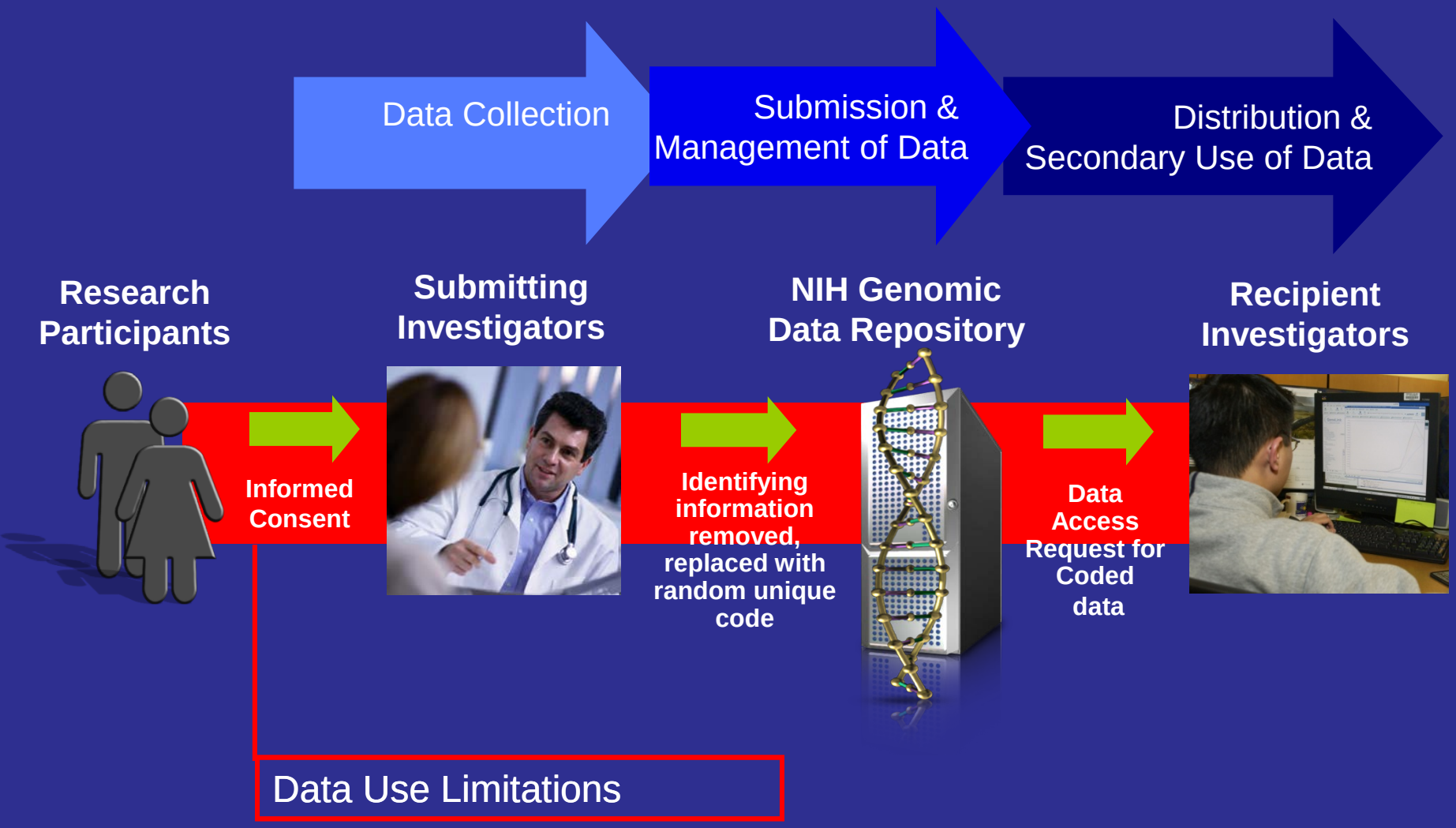
NIH Data Management Overview



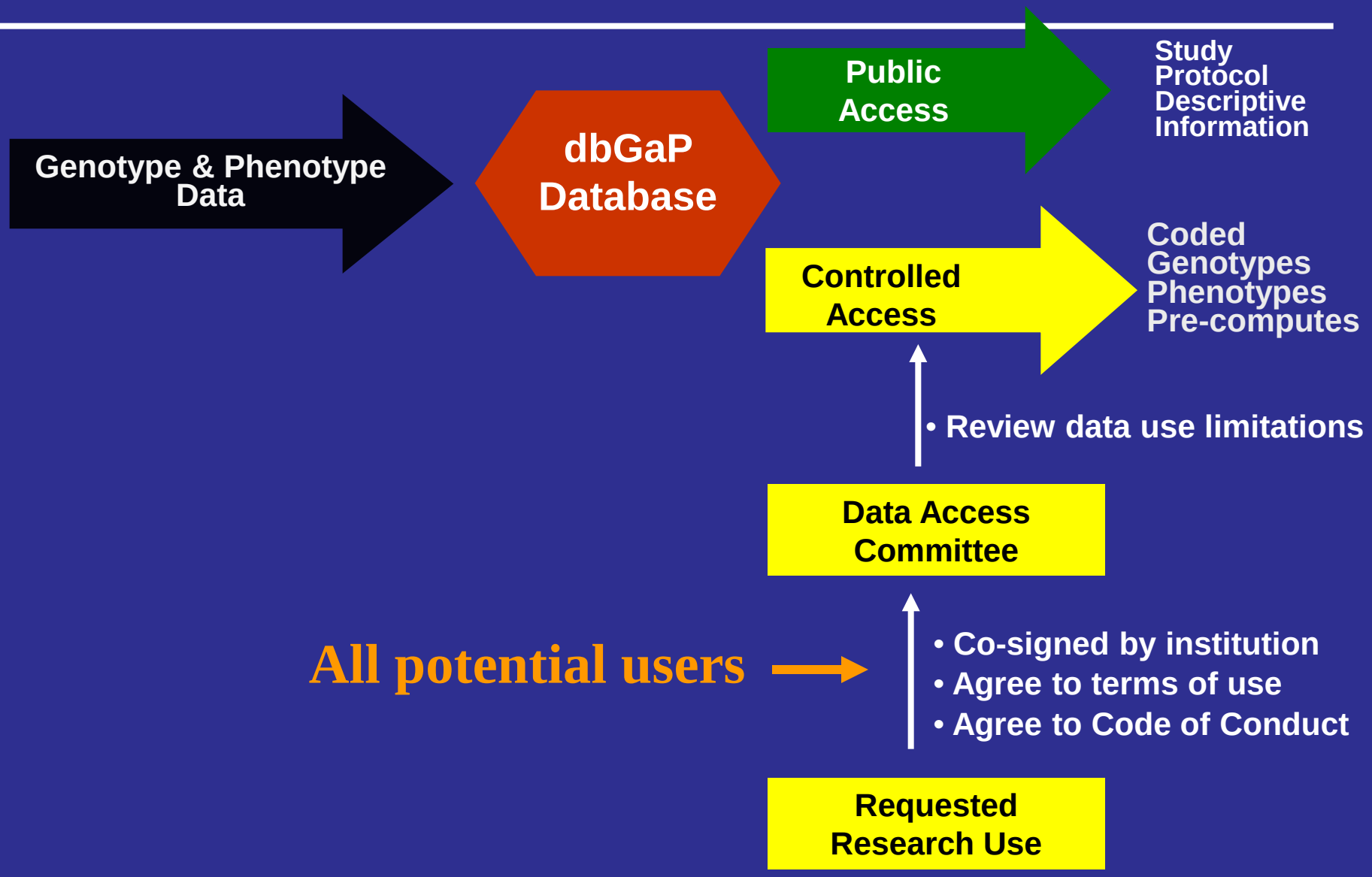
Data Submission

- **Local institution** will certify approval of submission to GWAS data repository, including statements that:
  - data are provided in accord with applicable laws and regulations
  - an **IRB** or Privacy Board has reviewed the submission plans
- The **PI** will remove HIPAA identifiers and retain the keycode to the data
- Any limitations on data use are requested at time of application (e.g., limitations imposed by existing informed consent).

NIH Data Management Overview



Data Access


Genotype & Phenotype Data
dbGaP Database
Public Access
Study Protocol Descriptive Information
Controlled Access
Coded Genotypes Phenotypes Pre-computes
• Review data use limitations
Data Access Committee
All potential users
• Co-signed by institution
• Agree to terms of use
• Agree to Code of Conduct
Requested Research Use

Data Access

- **Investigators and home institutions** responsible for compliance with federal, state, and local policies
  - Local institutional review – HIPAA – 45 CFR 46
- Secondary data users not conducting human subjects research
- Data Access Committees (DACs) will review requests for consistency with data use limitations
  - Federal staff with appropriate expertise
  - Also responsible for tracking and reviewing Approved User Annual Reports

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 other GWAS policies
  - abide by Code of Conduct: gwas.nih.gov/06researchers1.ht

Data Access Committees

- Data Access Committees (DACs) review all requests from the research community for access to dbGaP:
- 14 Active DACs across 14 Institutes (not 1:1)
  - GAIN, NHGRI, TCGA
  - CGEMS, eNCI (and TCGA)
  - NHLBI
  - JAAMH
  - NINDS
  - NIAID
  - NIDDK
  - JARDE, NIGMS, NEI, NICHD

Primary Question asked by DAC

- Is the proposed research use consistent with any data use limitations for the dataset?
- Data use limitations
  - Provided by the institution
  - Can be more than one “consent group” for a project

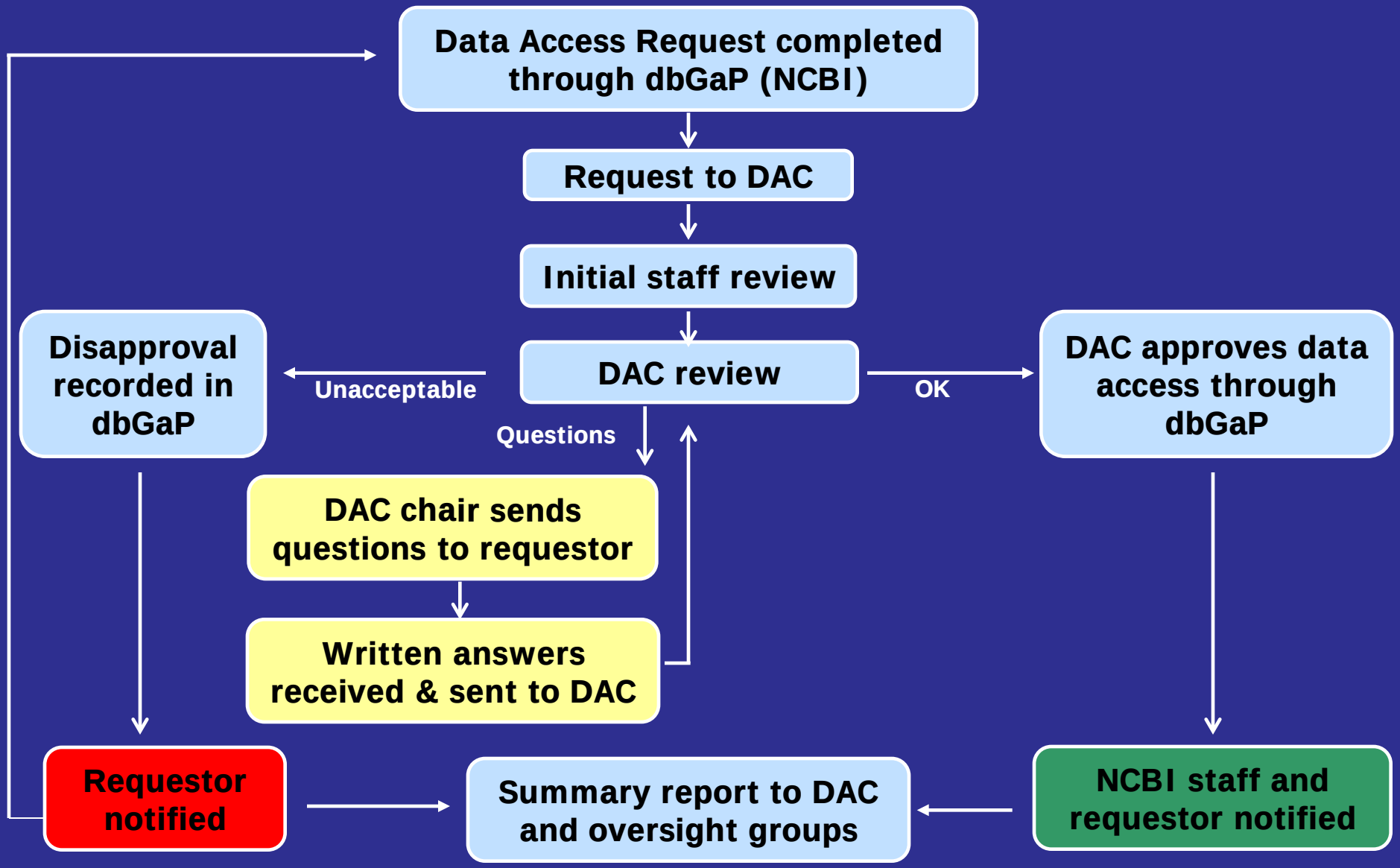
E.g. GAIN: Genome-Wide Association Study of Schizophrenia

General Research Use: May be used for any genetic studies

Schizophrenia and Related Disorders: Limited to genetic studies of Schizophrenia and related conditions. Related conditions include conditions with evidence of genetic relationships to schizophrenia or schizoaffective disorder, such as acute psychoses, bipolar disorder, MDD, or "Cluster A" personality disorders (schizotypal, schizoid, paranoid)

http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000021.v2.p1

Data Request Review Procedures



Trends in Access Requests

- Requestors come from across the research community, with most residing in academic institutions
- Proposed research uses include:
  - understanding the etiology of the target disease or related conditions/traits
  - testing new statistical methods to identify disease susceptibility genes or gene-environment interactions

DAC Productivity (aka workload)

From May 2011 – October 2011:

Per Project:

- 573 Research Statements Reviewed
- 459 Research Statements Approved
- 121 Research Statements Rejected
- 58 Research Statement Revisions Requested

Per Consent Group:

- 1671 Project Requests approved
- 474 Project Requests rejected
- 211 Project Request revisions requested

Public Disclosure under FOIA

- dbGaP GWAS data will be coded and deidentified.
- Policy concern remains that the extensive genotype data in dbGaP is intrinsically unique.
- NIH intends to deny FOIA requests for individual-level genomic data.
- ACD Working Group recommended seeking a legislative exception for genomic data

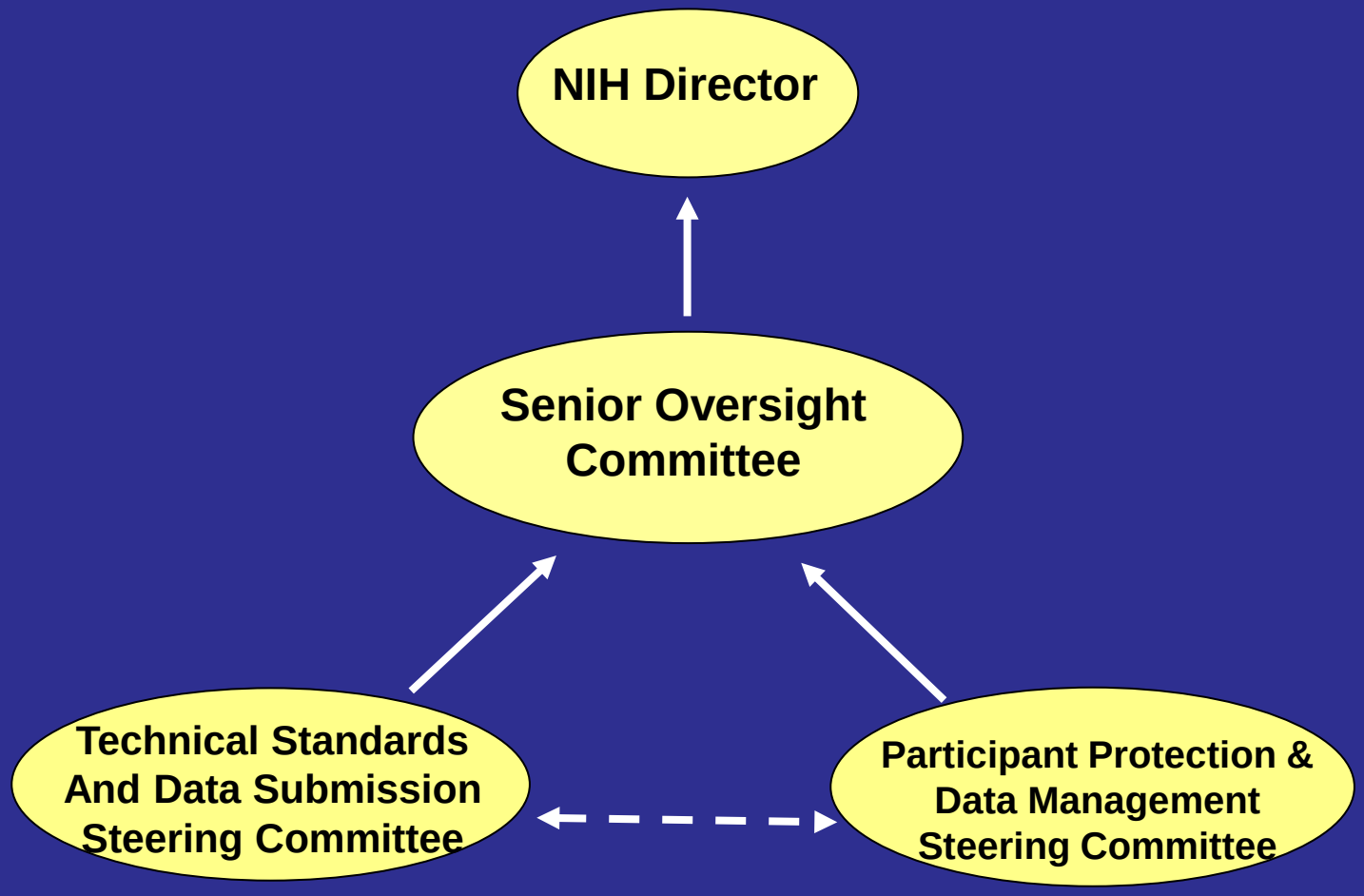
Compelled Disclosure

- Law enforcement and other judicial requests for data are handled under different procedures than FOIA requests.
- Standard protection is provided through the Secretary's "301(d)" authority *(aka Certificates of Confidentiality).*
  - Certificates are issued for a single project.
  - Submitting Investigators are encouraged to consider requesting a Certificate of Confidentiality.
  - dbGaP has received a Certificate.

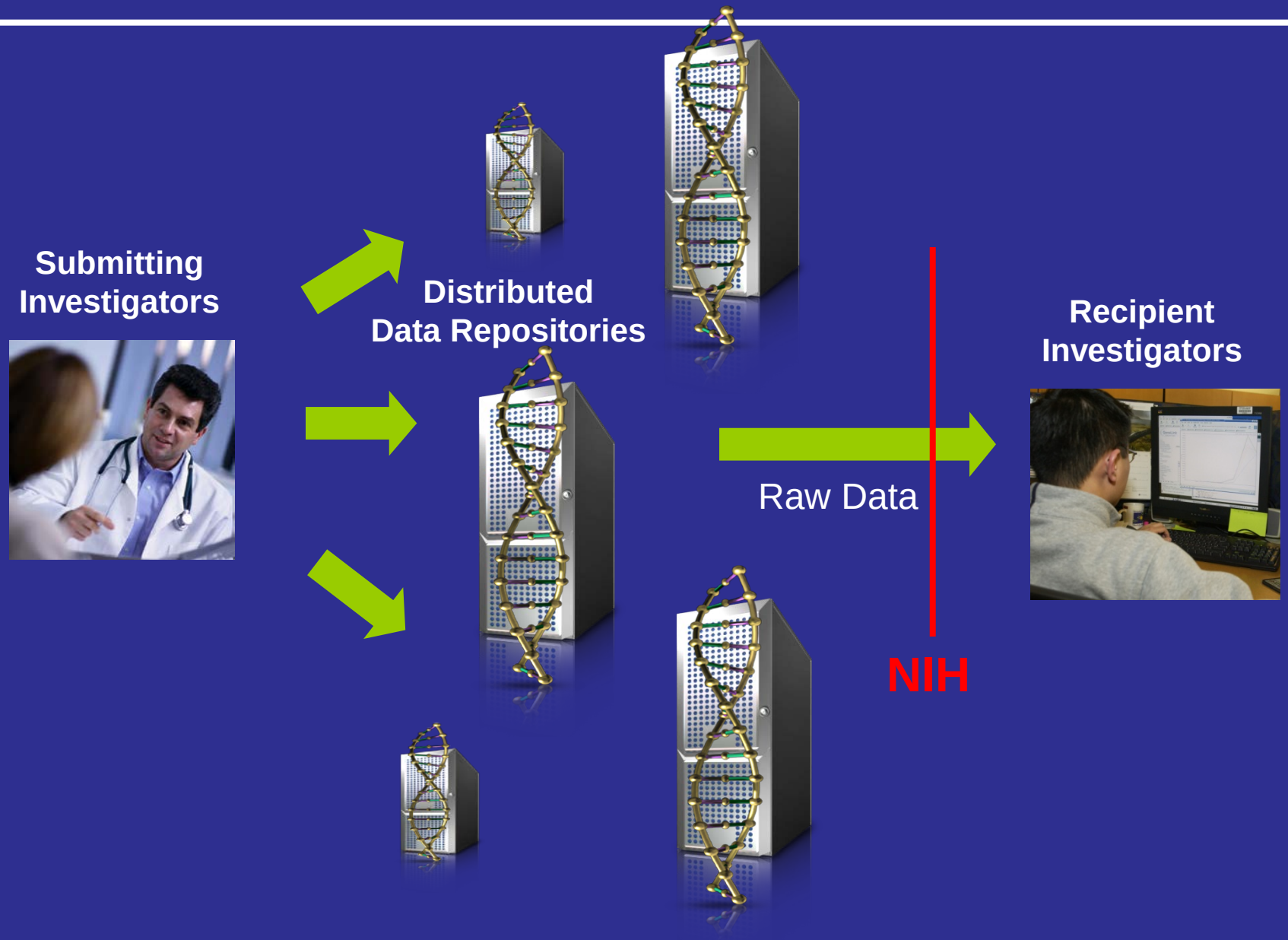
Exceptions to Data Deposition

- 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
  - Point of negotiation with Program staff
  - Agreed upon plan before funding
- To date: 7 requests granted
  - Limited consent
  - Legal restrictions
  - Localized geographic representation
- Points to Consider under development

Governance & Oversight



Trusted Partner Model(s)



NEW: Genomic Data Sharing (GDS) Policy

Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data

Notice Number: NOT-HG-10-006

Key Dates
Release Date: October 19, 2009

Issued by
National Human Genome Research Institute (NHGRI), (http://www.genome.gov/)

Purpose

The purpose of this Notice is to inform the research community of plans by the National Institutes of Health (NIH) to:

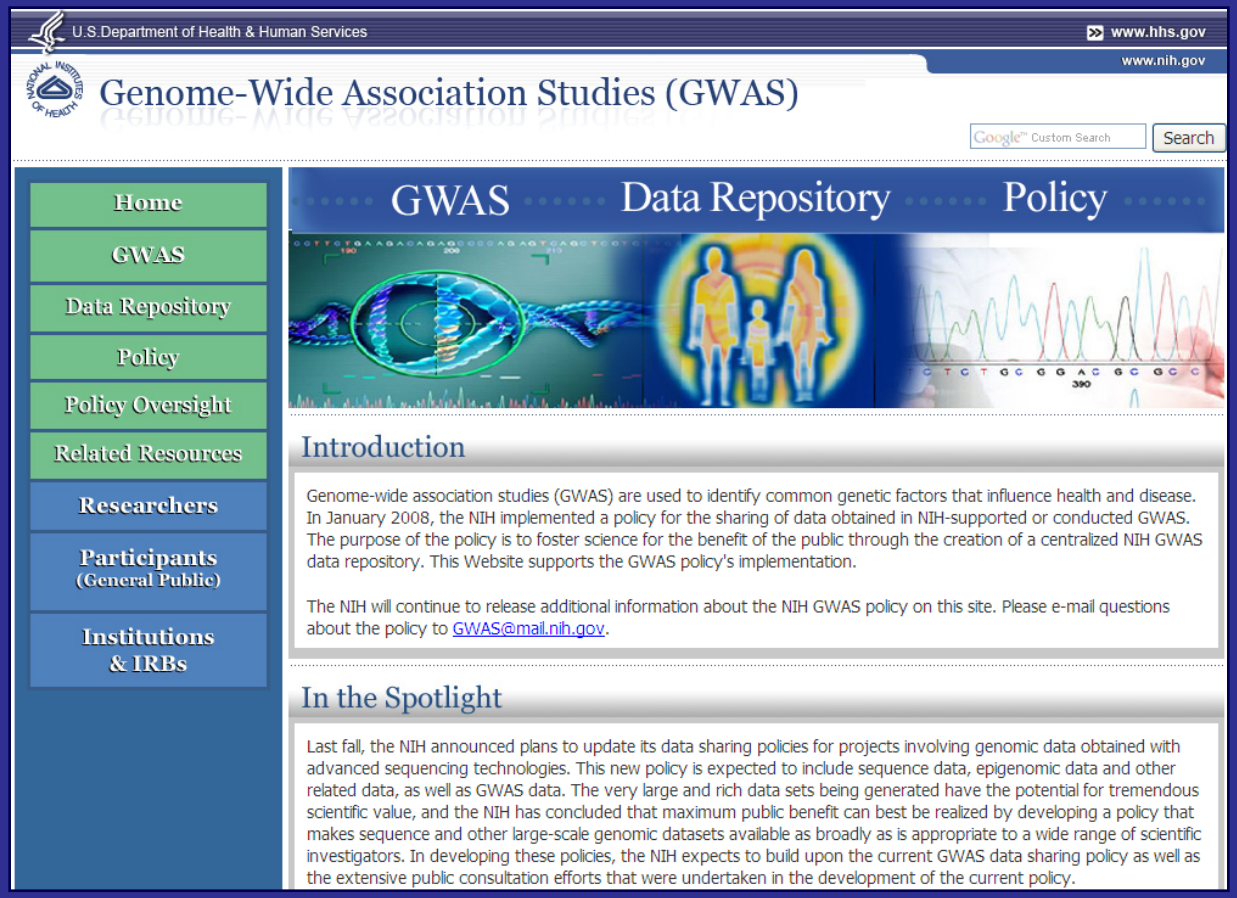
1. Update data sharing policies for NIH supported research, including extramural and intramural projects, involving sequence and related genomic data obtained with advanced sequencing technology (*e.g.*, medical resequencing data, sequence data from non-human species, including microorganisms, transcriptomic and epigenomic data, as well as data needed for interpretation, including associated clinical, other phenotype and metadata, such as supporting study documents and methodologies);
2. Encourage investigators and IRBs to consider the potential for broad sharing of sequence and related genomic data in developing informed consent processes and documents for such studies involving human sequence data; and,
3. Communicate the agency's intent and current underlying considerations related to developing a policy pertaining to the deposition of these large datasets into centralized databases, such as the GenBank Short Read Archive (SRA) or the Database of Genotypes and Phenotypes (dbGaP), so that they are available as broadly and rapidly as possible to a wide range of scientific investigators.

- Extension of the rationale for data sharing policies
- Accessibility of whole genome sequencing and other genomic technologies advancing rapidly
- Consistent approach for investigators, institutions secondary users, and participants
- Complexity of data and project types much greater

Questions for GDS …

- What data and project types should be included
- Data submission process and timeline
- Are there options for Open Access data release?
- Data release process and timeline
- Participant protection issues:
  - - Informed consent - Privacy & Confidentiality
  - - IRB review - Exceptions to deposition
- Updates/lessons learned from GWAS experience
- NIH implementation & oversight mechanisms

For more Information & Updates



http://gwas.nih.gov

or email: gwas@mail.nih.gov or laura.rodriguez@nih.gov