IMPLEMENTING THE NIH GENOMIC DATA SHARING (GDS) POLICY WITHIN IRP

PRESENTATION TO OHSRP LECTURE SERIES AUGUST 7, 2023

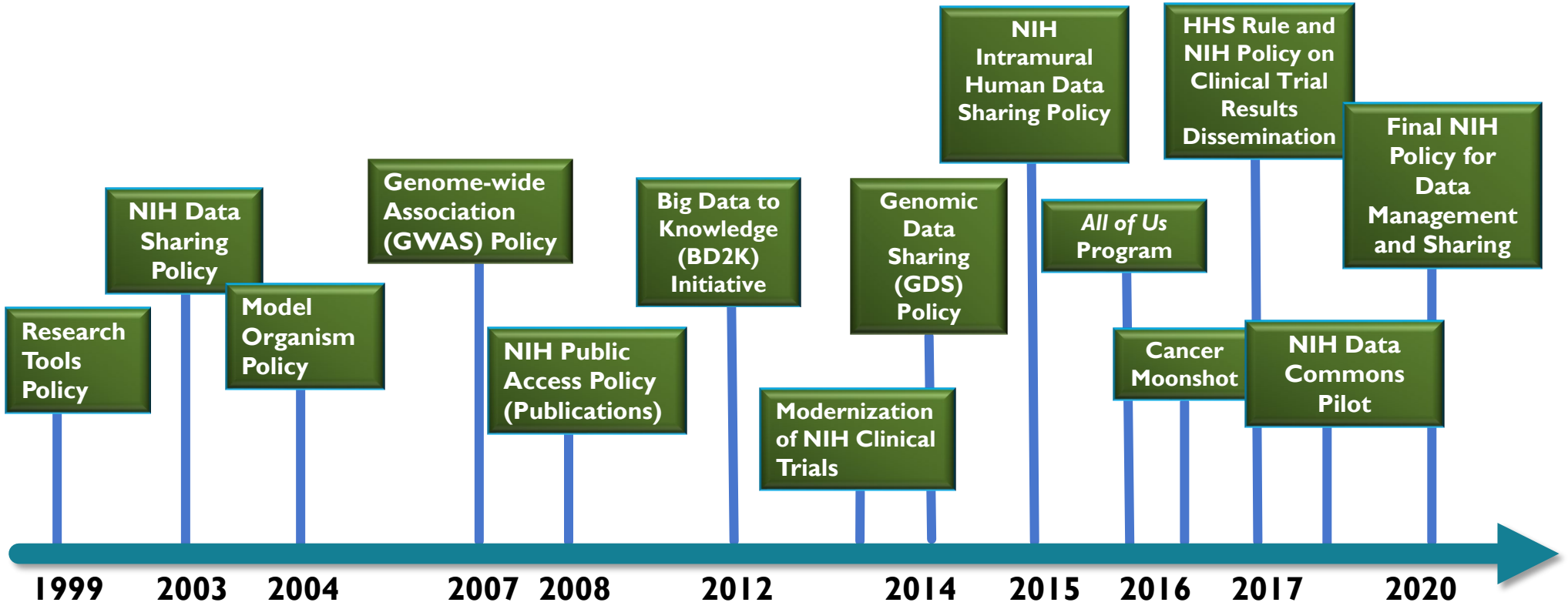
Cheryl Jacobs, PhD
Team Lead, Genomic Data Sharing Policy
Division of Scientific Data Sharing Policy
Office of Science Policy (OSP)

Julia Slutsman, PhD
Director, Genomic Data Sharing
Policy Implementation
Extramural Research (OER)

GUIDING PRINCIPLE OF THE NIH GENOMIC DATA SHARING POLICY

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

NIH'S CULTURE OF DATA SHARING



EFFECTIVE JANUARY 25, 2015 *(NIH INTRAMURAL – AUGUST 31, 2015)*

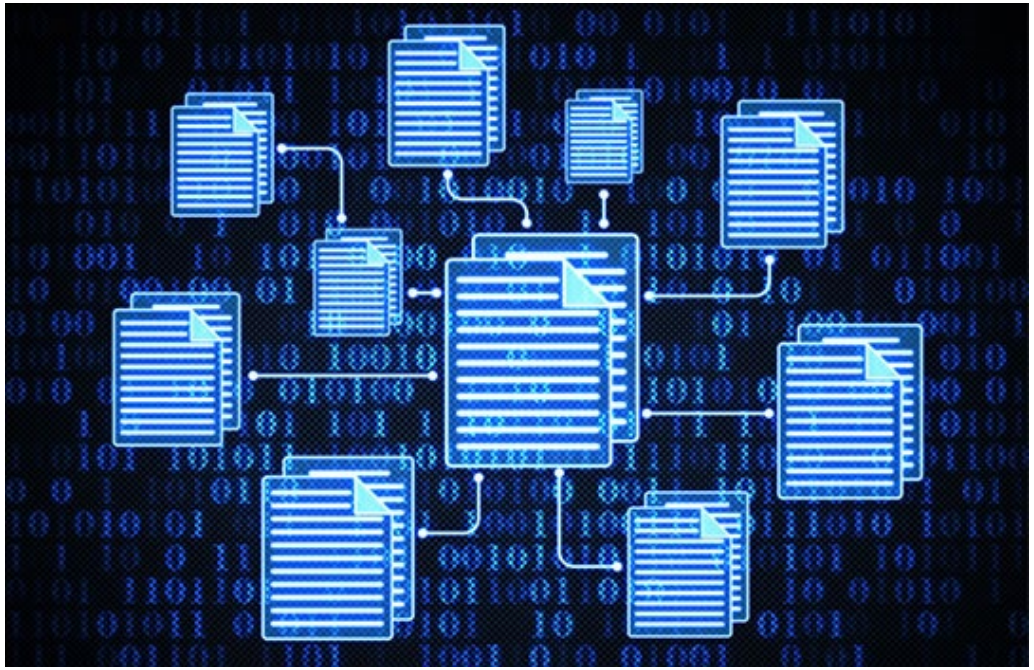
NIH GENOMIC DATA SHARING (GDS) POLICY OVERVIEW

Purpose

- Sets expectations and responsibilities to ensure broad, responsible, and timely sharing of genomic data

Scope

- Applies to all NIH-funded research generating large-scale human or non-human genomic data and secondary research using these data
- Applies to all funding mechanisms (grants, contracts, intramural support) regardless of cost



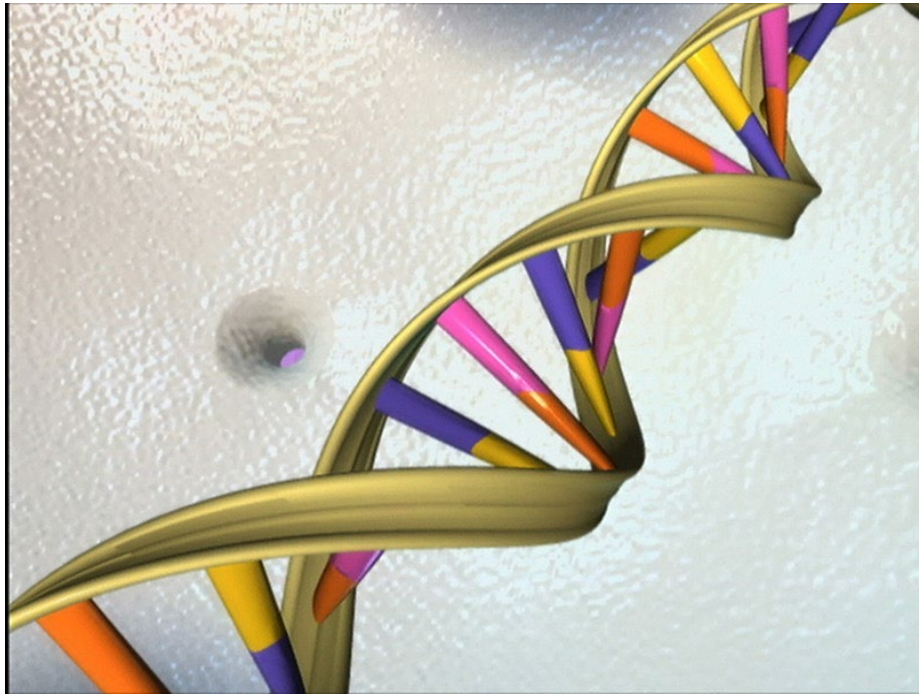
Genomic Data Sharing Policy

GDS POLICY APPLICATION

NIH FUNDED GENERATION OF LARGE-SCALE HUMAN OR NON-HUMAN DATA

Examples of such data (but not limited to)

- Sequence data from more than:
  - **One gene or region** of comparable size in the genomes of **more than 1,000** human research participants
  - **100 genes or region** of comparable size in the genomes of **more than 100** human research participants
  - **100 isolates** from infectious organismss
  - **100 metagenomes** of human or model organism microbiomes
- Catalog of **more than 100,000** single nucleotide polymorphisms (SNPs) from one or more model organism species or strains.



Confirm with your Genomic Program Administrator

GDS Policy Expectations

Top five things you need to know

To **comply** with the GDS policy, NIH expects that investigators and institutions:

1. Develop and provide a plan for sharing genomic data as a part of the Data Management and Sharing Plan
2. Provide an Institutional Certification form prior to starting a study, if working with human genomic data
3. Submit genomic data in a timely manner to an appropriate repository
4. Responsibly use controlled-access data
5. Appropriately cite controlled-access data in publications and presentations

Data Management and Sharing Plan

PROCESS FOR SUBMITTING OR ACCESSING HUMAN GENOMIC DATA

Investigators Must Submit:

- Data Management and Sharing Plan
- Institutional Certification prior to beginning a study
- Indication of any applicable data use limitations

Submit to NIH-Designated Data Repository or a widely used repository in their field (non-human data)

PROCESS FOR ACCESSING CONTROLLED HUMAN GENOMIC

Data Requesters Must Meet Following Criteria

- At a level equivalent to a tenure-track professor
- Senior scientist with responsibilities that may include a laboratory
- Research program administration and oversight

Data Requestors Submit a Data Access Request (DAR)

- Co-signed by institution
- Agree to terms of use in Data Use Certification
- PI agrees to Genomic Data User Code of Conduct

Data Access Committee (DAC) Reviews Request

- Review research proposal and ensure adherence to data use limitations and GDS Policy
- Verify PI credentials
- Consider the potential for group harm (e.g., stigmatization)

Data Requester Accesses Data

- PIs should safeguard the accessed datasets
- Access to the data for one year
- Adhere to the DUC Agreement and the Code of Conduct

SUPPLEMENTAL DMS POLICY RESOURCES

- Supplemental Information: Protecting Privacy When Sharing Human Research Participant Data (NOT-OD-22-213)
  - Provides a basic framework for considering how to protect privacy when sharing data from human participants
  - Not intended as a guide for regulatory compliance
  - ***Not intended to replace consent expectations of the GDS Policy***

- Supplemental Information: Responsible Management and Sharing of American Indian/ Alaska Native Participant Data (NOT-OD-22-214)
  - Respect for Tribal Sovereignty
  - Partnerships and mutual agreements
  - Building trust

HARMONIZATION OF THE GENOMIC DATA SHARING (GDS) POLICY WITH THE DATA MANAGEMENT AND SHARING (DMS) POLICY

WHAT HAS NOT CHANGED

- GDS Policy **consent expectations** for sharing human participant data
- Expectation that **Institutional Certifications** be submitted prior to starting a study
- GDS **timelines for data submission** and release
- GDS expectations for **amount and type of data** to share
- **When appropriate, designate a study as "sensitive"** for the purposes of access to Genomic Summary Results

WHAT HAS CHANGED

- **DMS Plan:** submit a single Plan that addresses GDS and DMS policies
- **DMS Plan Review:** conducted by IC Scientific Director or designee
- **GDS timelines** for data submission and release
- **Limitations on Sharing:** described in DMS Plans, no alternative data sharing plan
- **Compliance:** new and updated DMS plans must be entered by the time of the report process

ELEMENTS OF A DMS PLAN

Element 1: Data Type

Types and amount of scientific data expected to be generated in the project:

- *Summarize the types and estimated amount of scientific data expected to be generated in the project,*

Scientific data that will be preserved and shared, and the rationale for doing so:

- *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

Metadata, other relevant data, and associated documentation:

- *Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.*

Element 2: Related Tools, Software and/or Code:

- *State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.*

ELEMENTS OF A DMS PLAN

Element 3: Standards:

- ***State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.***

Element 4: Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived:

- *Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository).*

How scientific data will be findable and identifiable:

- *Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.*

When and how long the scientific data will be made available:

- *Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.*

ELEMENTS OF A DMS PLAN

Element 5: Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data:

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

Whether access to scientific data will be controlled:

State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).)

Protections for privacy, rights, and confidentiality of human research participants:

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

Element 6: Oversight of Data Management and Sharing:

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

ELEMENTS TO INCLUDE IN A DMS PLAN WHEN SUBJECT TO GDS POLICY

Instructions for all scientific data

Additional expectations for genomic data

Data Type

Briefly describe the scientific data to be managed and shared:

- Summarize the types (for example, 256-channel EEG data and fMRI images) and amount (for example, from 50 research participants) of scientific data to be generated and/or used in the research. Descriptions may include the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing.
- Describe which scientific data from the project will be preserved and shared. NIH does not anticipate that researchers will preserve and share all scientific data generated in a study. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors. The plan should provide the reasoning for these decisions.
- A brief listing of the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data

For data subject to the GDS Policy:

- Data types expected to be shared under the GDS Policy should be described in this element. Note that the GDS Policy expects certain types of data to be shared that may not be covered by the DMS Policy's definition of "scientific data". For more information on the data types to be shared under the GDS Policy, consult Data Submission and Release Expectations.

Ex: Sharing GWAS data (GDS Policy) versus only sharing data from GWAS dataset of sufficient quality to validate and replicate findings, per the DMS Policy

ELEMENTS TO INCLUDE IN A DMS PLAN WHEN SUBJECT TO GDS POLICY

Instructions for all scientific data

Data Preservation, Access, and Associated Timelines

Give plans and timelines for data preservation and access, including:

- The name of the repository(ies) where scientific data and metadata arising from the project will be archived. See Selecting a Data Repository for information on selecting an appropriate repository.
- How the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.
- When the scientific data will be made available to other users and for how long. Identify any differences in timelines for different subsets of scientific data to be shared.
    - Note that NIH encourages scientific data to be shared as soon as possible, and no later than the time of an associated publication or end of the performance period, whichever comes first. NIH also encourages researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, institutions, and/or the broader public.

Additional expectations for genomic data

For data subject to the GDS Policy:

- For human genomic data:
    - Investigators are expected to submit data to a repository acceptable under the Genomic Data Sharing Policy. See Where to Submit Genomic Data.
    - Human genomic data is expected to be shared according to NIH's Data Submission and Release Expectations, but no later than the end of the performance period, whichever comes first.
- For Non-human genomic data:
    - Investigators may submit data to any widely used repository.
    - Non-human genomic data is expected to be shared as soon as possible, but no later than the time of an associated publication, or end of the performance period, whichever is first.

ELEMENTS TO INCLUDE IN A DMS PLAN WHEN SUBJECT TO GDS POLICY

Instructions for all scientific data

Access, Distribution, or Reuse Considerations

Describe any applicable factors affecting subsequent access, distribution, or reuse of scientific data related to:

- Informed consent
- Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local laws, regulations, and policies
- Whether access to scientific data derived from humans will be controlled
- Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements
- Any other considerations that may limit the extent of data sharing. Any potential limitations on subsequent data use should be communicated to the individuals or entities (for example, data repository managers) that will preserve and share the scientific data. The NIH ICO will assess whether an applicant's DMS plan appropriately considers and describes these factors. For more examples, see Frequently Asked Questions for examples of justifiable reasons for limiting sharing of

Additional expectations for genomic data

Expectations for human genomic data subject to the GDS Policy:

- Informed Consent Expectations:
  - For research involving the generation of large-scale human genomic data from cell lines or clinical specimens that were created or collected **AFTER the effective date of the GDS Policy (January 25, 2015):**
    - NIH expects that informed consent for future research use and broad data sharing will have been obtained. This expectation applies to de-identified cell lines or clinical specimens regardless of whether the data meet technical and/or legal definitions of de-identified (i.e. the research does not meet the definition of "human subjects research" under the Common Rule).
  - For research involving the generation of large-scale human genomic data from cell lines or clinical specimens that were created or collected **BEFORE the effective date of the GDS Policy:**
    - There may or may not have been consent for research use and broad data sharing. NIH will accept data derived from de-identified cell lines or clinical specimens lacking consent for research use that were created or collected before the effective date of this Policy.

SAMPLE NIH DMS PLANS AVAILABLE

- 13+ sample NIH DMS Plans available for educational purposes, including:
  - Human clinical and/or MRI data (NIMH)
  - Human genomic data (NIMH, NHGRI, NIDDK)
  - Human & non-human genomic data (NIMH)
  - Secondary data analysis (NIMH, NIDDK)
  - Human clinical and genomics data (NICHD)
  - Human survey data (NICHD)
  - Model organism (Zebrafish) data (NICHD)
  - Technology development (NHGRI)
  - Clinical data (NIDDK)
  - Non-human basic research (NIDDK)

DATA MANAGEMENT AND SHARING PLAN

An example from an application proposing to collect single cell genomic data from mice and humans.

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted (but this has not been done in the sample below). There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:

Summarize the types and estimated amount of scientific data expected to be generated in the project.

As detailed in the Research Strategy Section, we propose the generation of a spatially mapped single-cell atlas of the developing mouse brain and include specific deliverables. Our primary deliverable for each modality will be a matrix of cells × (counts in peaks for ATAC, UMIs in genes for RNA, or methylation status for DNAm) along with a dense metadata table with information for each cell. This includes the animal sex, developmental time point, punch of origin with x,y,z coordinates, assigned cluster and inferred cell type, assigned subcluster and inferred cell type, as well as a number of QC metrics (total reads, passing reads, reads in peaks, TSS enrichment, cell barcode combination, date of preparation for each stage, sequencing platform, likelihood of being a doublet, and any other relevant metrics that arise during the project).

The amount and type of data from human cells will depend on the results from the mouse studies. Data sharing plans will be updated when appropriate (likely at the start of year 4 of the grant award).

See Writing a Data Management & Sharing Plan for details

IRP PI RESPONSIBILITIES & GDS POLICY RESOURCES

NIH IRP PI RESPONSIBILITIES UNDER THE NIH GENOMIC DATA SHARING POLICY FOR HUMAN DATA

1. Develop a Data Management & Sharing (DMS) Plan and obtain approval of the Plan from the IC Scientific Director (SD)/designee

2. Complete and submit the applicable IRP Institutional Certification(s).
   i. Ensure that the type of access and data limitations you choose in the Institutional Certification(s) are consistent with protocol and consent form(s)
   ii. ii. Share your DMS Plan, protocol, consent form(s) and Institutional Certification(s) with the IRB (or equivalent)
   iii. Obtain SD or designee approval (and signature) of the Institutional Certification(s)
   iv. Provide the completed and signed Certification(s) to your IC GPA

3. Before or at the time that data cleaning and quality control measures begin, register the study in dbGaP

4. Make a copy of the genomic data and deidentify data per GDS Policy expectations, code the data and maintain a code key

5. Submit the data to an NIH-designated repository per the applicable submission requirements

INTRAMURAL INSTITUTIONAL CERTIFICATION FORMS

- Institutional Certification must be submitted before a study begins
- Must be reviewed by Scientific Director and an IRB
- Multiple options for Certification
  - Pre-2015 with consent
  - Pre-2015 without consent
  - Post-2015 (informed consent expected)
  - Provisional (prior to IRB review)

PROCESS FOR INSTITUTIONAL CERTIFICATION REVIEW

- IRB must attest to certain elements of the Institutional Certification
  - PI should email IRB@od.nih.gov with required content to request review of IC for each protocol that generates data to be submitted to repository
- As part of this process, IRB must review:
  - Approved DMS Plan
  - All versions of informed consent documents signed by participants
  - Data sharing information included in the Institutional Certification

CONSULT GENOMIC PROGRAM ADMINISTRATORS

- ICO experts on the GDS Policy:
  - **Genomic Program Administrators (GPAs)** – responsible for submission of data to NIH-designated repositories
  - **Genomic Data Submission and Management (GDSM) Taskforce representatives** – trans-NIH oversight of GDS Policy

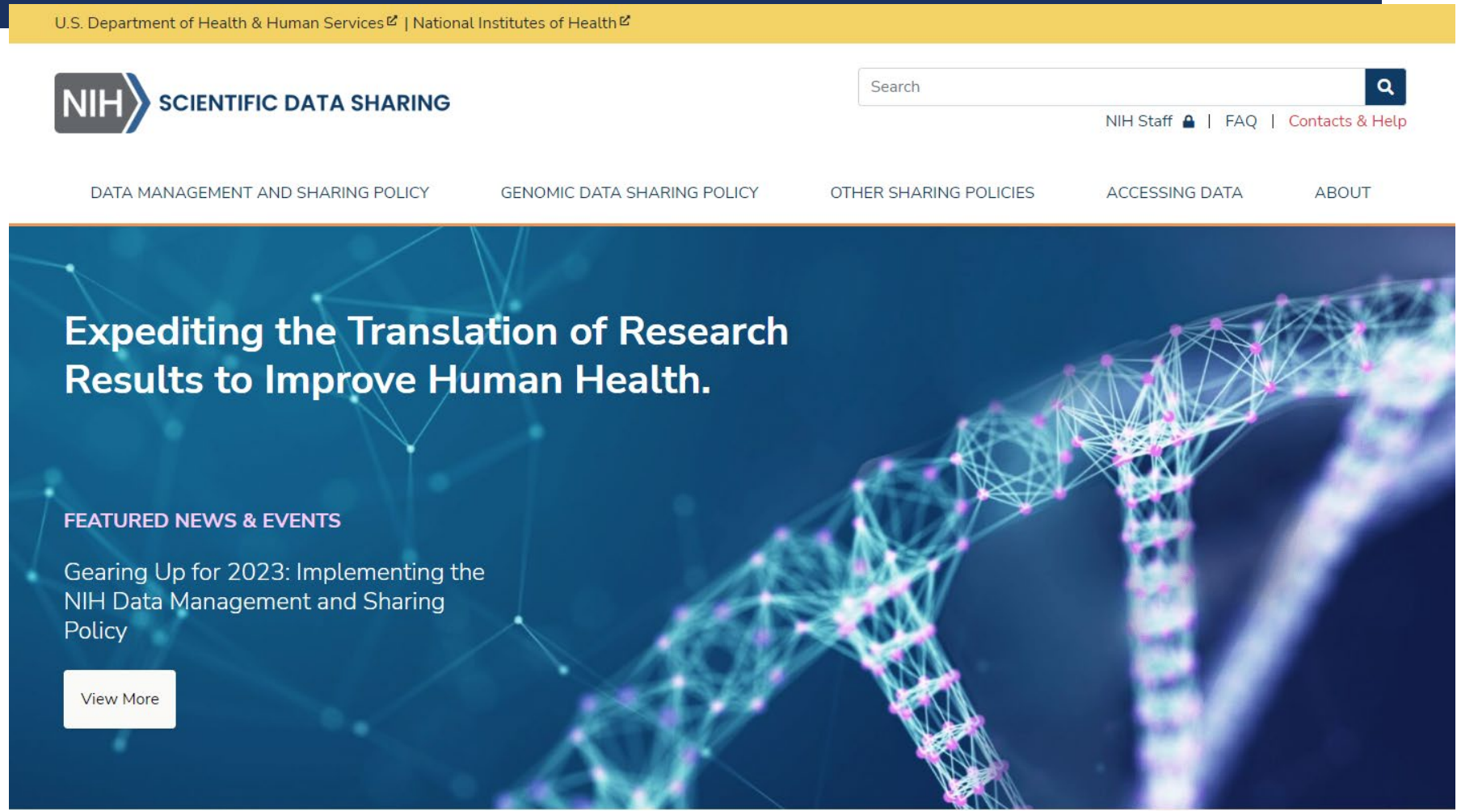
- They are the primary resources to help decide:
  - Whether research is subject to the GDS Policy
  - Any additional IC Specific GDS Policy Expectations
  - What should be included in a DMS Plan
  - Whether research can meet the expectations of the Institutional Certification
  - How to submit genomic data to repositories

RESOURCES

- NIH resources:
  - GDS Policy inbox, GDS@mail.nih.gov
  - GPAs, https://sharing.nih.gov/contacts-and-help
  - NIH data sharing website, https://sharing.nih.gov/
- OIR Genomic Data Sharing: https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/genomic-data-sharing
- OIR Data Management and Sharing: https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy
- OHSRP Consent Templates and Guidance: https://irbo.nih.gov/confluence/display/ohsrp/Consent+Templates+and+Guidance

RESOURCES: SHARING.NIH.GOV

- Provides a central source of guidance related to multiple NIH data sharing policies
- Covers Data Management and Sharing, Genomic Data Sharing, Model Organisms, and Research Tools policies
- Content updated regularly



QUESTIONS?

THANK YOU!