Policy and Technology: Ensuring Ethics in the Submission and Access of Biomedical Research Data

Dina N. Paltoo Ph.D., M.P.H.
Assistant Director, Scientific Strategy and Innovation
National Heart, Lung, and Blood Institute

Federal College of Statistical Methods 2020 Fall Conference

Policy Session 4.3: Management of Ethical Issues across the Data Lifecycle

September 21, 2020

National Heart, Lung, and Blood Institute

The Benefits of Data Sharing

- **Preserves the scientific record**
  - Encourages better data management; not all results are published
- **Facilitates research integrity, transparency and trust**
  - Validates experiments and results
  - Engenders trust through transparency
- **Advances science and application**
  - Accelerates translation of results into practice
  - Suggests new hypotheses
  - Innovates through statistical methods, resources, and tools
- **Increases efficiency, fosters rigor and reproducibility**
  - Increases statistical power and value
  - Enables data generated from one study to be used by others to explore additional research questions
  - Decreases in spending on duplication of original studies

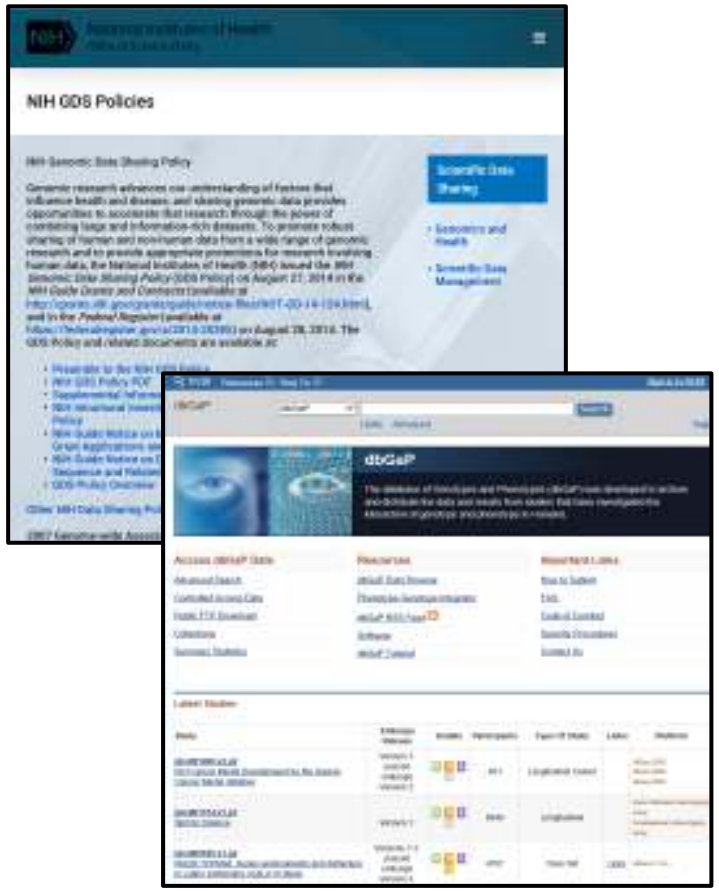
NIH has a longstanding commitment to data and resource sharing

Responsible Data Stewardship

Privacy and Trust are Key Components

- Participant protections and appropriate use of data
  - Health Insurance Portability and Accountability Act of 1996 (HIPAA)
  - Federal Policy for the Protection of Human Subjects (Common Rule – revised in 2018)
  - Certificates of Confidentiality
- Freedom of Information Act (FOIA)
- Privacy Act of 1974
- Genetic Information Nondiscrimination Act (GINA)
- Authentication and authorization of data users (e.g., for controlled-access data)

NIH Genomic Data Sharing Policy

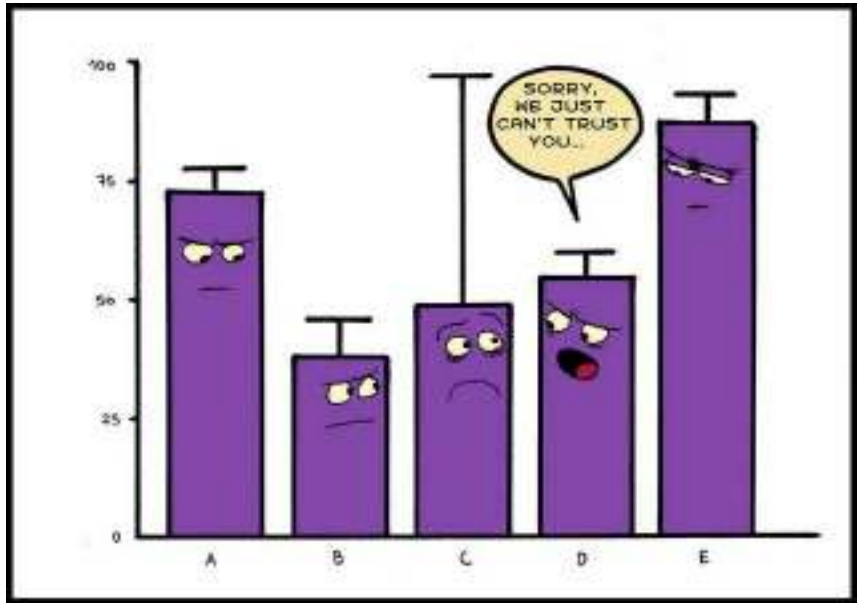


- Sets forth expectations and responsibilities to ensure the timely, broad, and responsible sharing and use of genomic data from NIH funded research
- Access to human data (de-identified) in **dbGaP** is tiered and based on informed consent of study participants
  - **Controlled-access**: individual-level genomic and phenotypic data (requires an application and approval by a Data Access Committee)
  - **Unrestricted-access**: Study descriptions; Genomic summary results (GSR, for most studies) under a new data management update
- Users, authenticated through eRA Commons, agree to terms of use and security practice
- **More than 1,200** studies available; **more than 57,000** Data Access Requests approved (cumulative)
- **More than 3,000** publications resulting from re-use of dbGaP data

dbGaP: database of Genotypes and Phenotypes

Access to “Genomic Summary Results (GSR)”

GSR Includes:
- Genotype counts
- Allele frequencies
- p-values
- Effect size estimates and standard errors



- GSR has been shared via “controlled-access” in dbGaP
  - Possibility of inferring group association of participants for some GSR
- Community to NIH - benefits of moving GSR to unrestricted-access outweigh potential risks
- GSR management update has been released!!!

Protections and Safeguards for dbGaP Data Access and Use

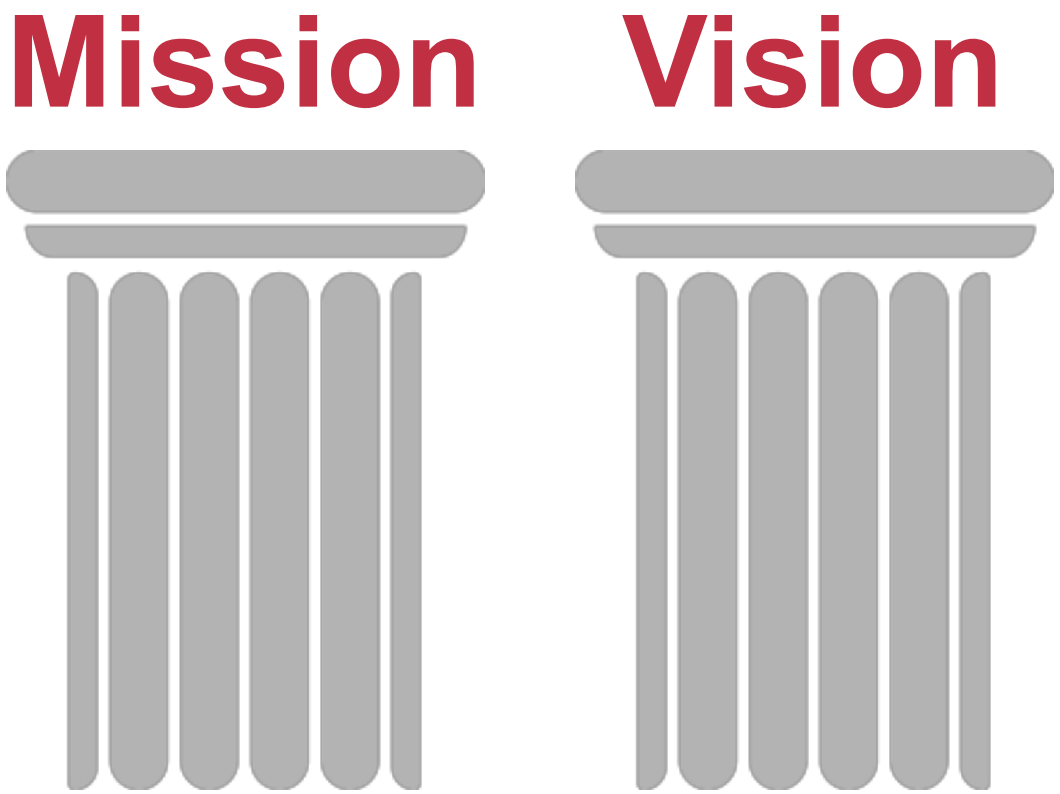
- **Protections:**
  - Institutional Certification for data submitters
    - Appropriateness of access level
    - Informed consent and any exceptions for data submission
    - Awareness of and respect for cultural and/or community-based concerns
    - Institutional certification and IRB determinations of consent applicability & data protection
  - Data Use Certification for data users
  - NIH Data Access Committees review of Data Access Request
  - Certificate of Confidentiality protects "identifiable, sensitive information"
- **Safeguards:**
  - Prevent identification of individual participants without appropriate approvals
  - Only authorized individuals can gain access to data
  - Use requested datasets solely in connection with project approved for data use
  - Approved investigators follow guidance on security best practices
  - Report any inadvertent data release, breach of data security, or other data management incidents

Connecting NIH Data Systems and Resources

- Align dbGaP access management with standardized NIH research commons Identity and Access Management specifications

- Develop a framework for resource management within cloud environments
  - Modeling of authority and the permissions required to access resources in order to support cooperative computing amongst the internal NIH ICOs and future federated partners
  - Allows for a federated login for researchers

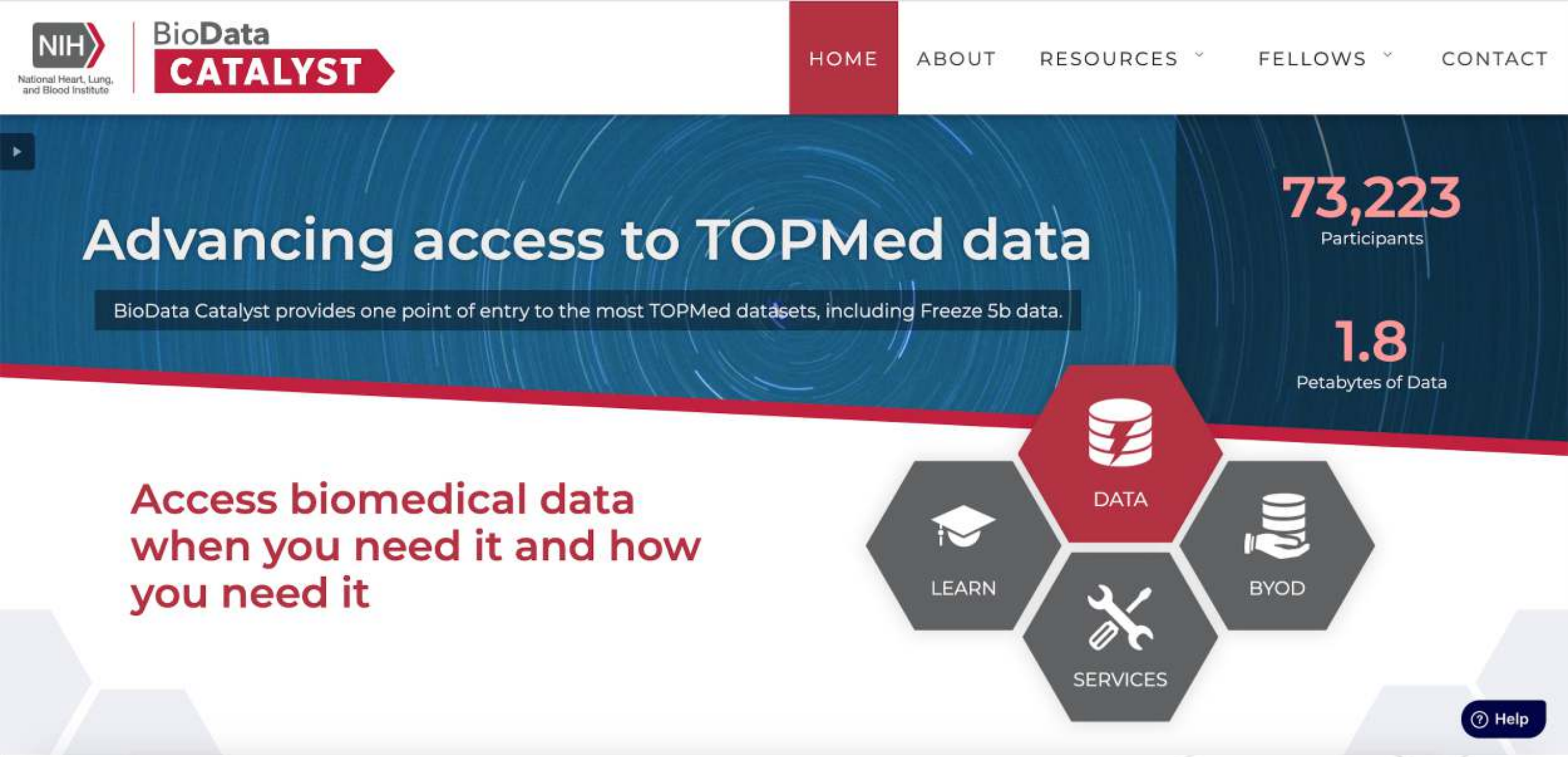
NHLBI BioData Catalyst



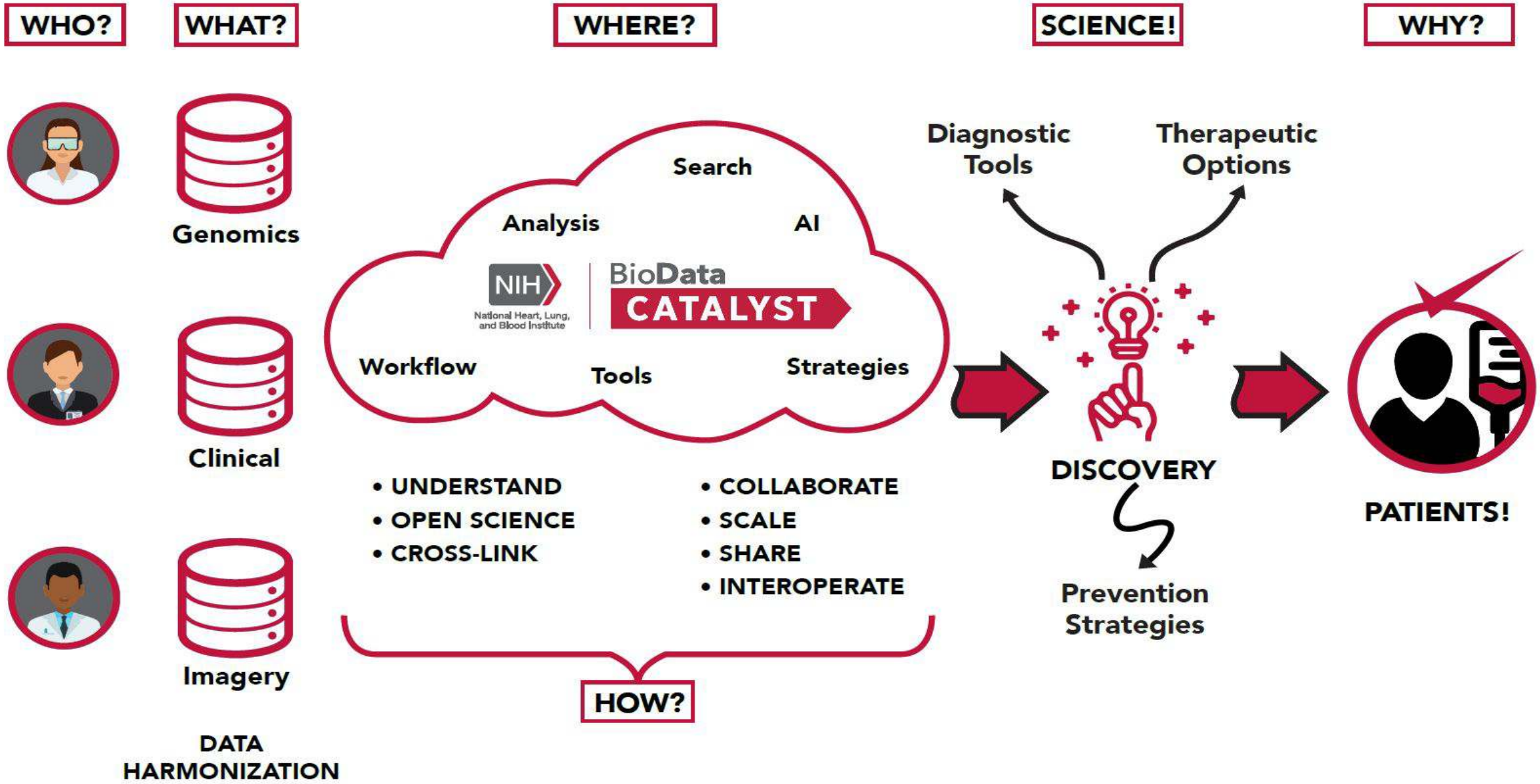
The ***mission*** is to develop and integrate advanced cyberinfrastructure, leading edge tools, and FAIR data to support the NHLBI research community.

The ***vision*** is to be a community-driven ecosystem implementing data science solutions to democratize data and computational access to advance Heart, Lung, Blood, and Sleep science.

FAIR – findable, accessible, interoperable, reusable



[To view a video-only demonstration, click here.]


WHO?
WHAT?
WHERE?
SCIENCE!
WHY?
Genomics
Clinical
Imagery
DATA HARMONIZATION
Search
Analysis
AI
NIH
National Heart, Lung, and Blood Institute
BioData CATALYST
Workflow
Tools
Strategies
• UNDERSTAND
• OPEN SCIENCE
• CROSS-LINK
• COLLABORATE
• SCALE
• SHARE
• INTEROPERATE
HOW?
Diagnostic Tools
Therapeutic Options
DISCOVERY
Prevention Strategies
PATIENTS!

Using Deep Learning with the BDCatalyst Ecosystem

What are the **key features** of subjects that makes them **more likely** to develop **severe symptoms of COVID-19?**

BDCatalyst User Community

1. Researcher enters the NHLBI BDCatalyst portal
2. Researcher authenticates with Gen3.
3. Data search and cohort creation occurs with PIC-SURE
4. Data is exported to a workspace for analysis

Gen3 Data Commons

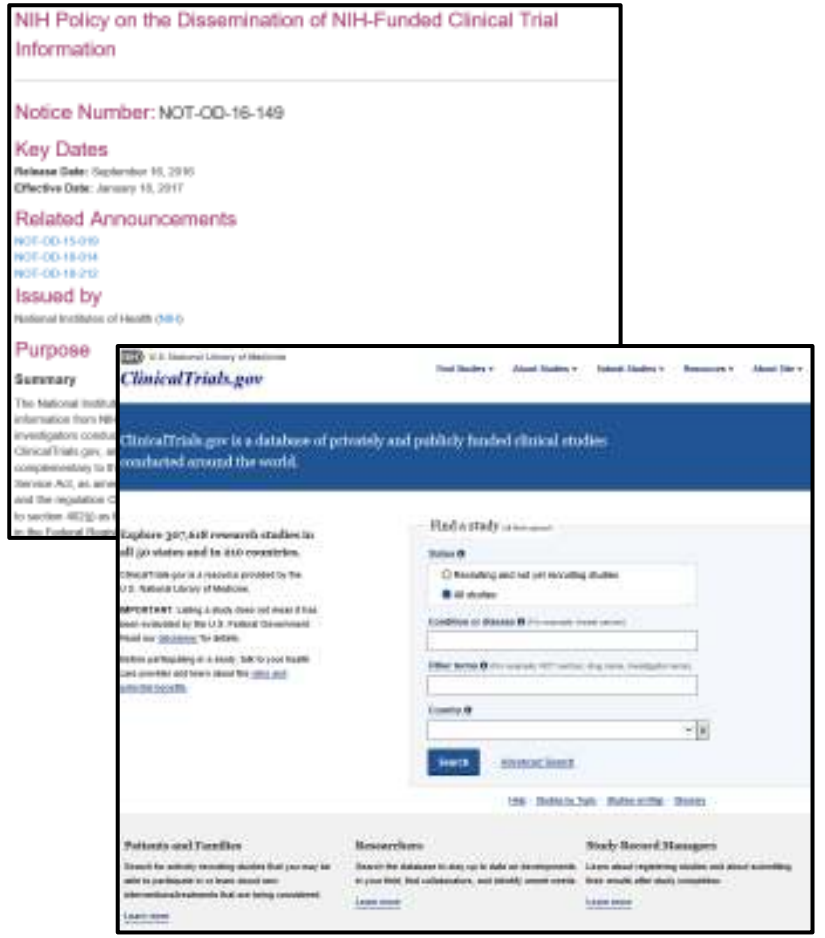
PIC-SURE

SevenBridges

Terra

Safeguards for Clinical Trial Data: Dissemination of Results

- Enhance transparency into NIH-funded and other clinical trials
- Registration of study objectives, design, etc. at **ClinicalTrials.gov**
- Summary results of clinical trial and participant characteristics
- **No** participant level data
- Linked to related information – possibly participant level data
- Full study protocol
- **More than 350,000** registered studies; **more than 39,000** summary results; **more than 115,000** users/ day



National Heart, Lung, and Blood Institute

NIH Data Management and Sharing Policy Development

Released for Public Comment: Proposed Policy Provisions for a Draft NIH Data Management and Sharing Policy (**October 2018**)

Analysis of public comments, and considerations for Policy Guidance and Implementation for:

- **Data Management and Sharing Plans** (elements, costs, collecting and evaluating, ensuring compliance)
- **Infrastructure** (e.g., NIH Figshare Pilot program, characteristics of repositories)
- **Timing** of Policy implementation
- **NIH Policy** that is **reasonable** and **achievable** for NIH-funded research

Released for public comment: draft NIH Data Management and Sharing Policy and supplemental guidance (**November 2019**)

Release final NIH Policy

Image from: http://www.whistlercentre.ca/2010/03/draft-icsp-framework-nearing-completion-for-williams-lake/people-working-together-small/

Application of Novel Tools and Technologies: Future Considerations

For approaches such as artificial intelligence (AI):

- Purpose, type of data, source of data
- Ethics and security
  - Governance
  - Risk and responsibility
  - Potential biases
- Future use and unintended use
- Transparency, validation, trust

AI is integrated into numerous technologies that people use every day. *Credit: iStock-metamorworks*

Lessons Learned and Take-away Messages

- **Considerable value** from enhancing access to (biomedical) data
  - Accelerates and improves **science**
  - Builds **trust** in the research enterprise
- **Sharing not binary (yes/no) – Effective models exist** for sharing data while providing necessary protections, e.g.,
  - Summary statistics from clinical trials, genomics research
  - Controlled access to deidentified participant level data
- **Consistency with informed consent** of great importance
- **Approaches**
  - Conservative initially
  - Listen to the community
  - Risk/benefit analysis
  - Balanced approach with solution that accommodates sensitive studies/populations
- **Technology and policy** go hand-in-hand

THANK YOU!

Dina.Paltoo@nih.gov

National Heart, Lung, and Blood Institute