

Data Science at NIH/NIDA

Susan N. Wright, PhD

Associate Director for Data Science at NIDA

Program Director for Big Data and Computational Science

Division of Neuroscience and Behavior (DNB)

National Institute on Drug Abuse (NIDA)

E-mail: susan.wright@nih.gov

Office: (301) 402-6683



**National Institute
on Drug Abuse**

Data Science Themes



NIH defines data science as

“the interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data.”

- Strategic Plan for Data Science (June 2018)

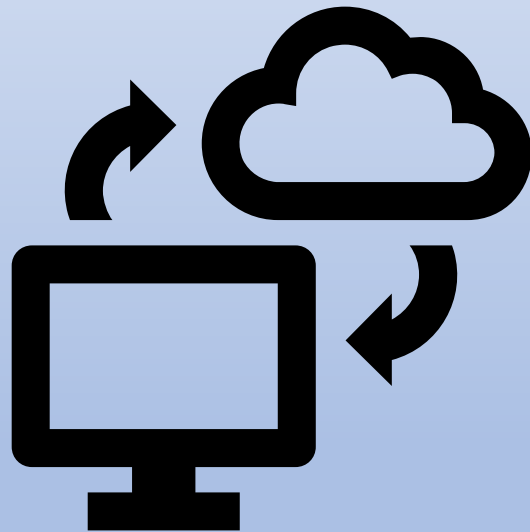
Data Science at NIDA

- The **complexity** and **volume** of basic, translational, and clinical research data generated by NIH-supported investigators continues to rapidly increase
- To take full advantage of these data, the NIH must integrate the collection, storage, analysis, use, and sharing of these data according to FAIR practices and foster a talented and **diverse data science workforce**
- **NIDA's mission** is to advance science on the causes and consequences of **drug use** and **addiction** and to apply that knowledge to improve individual and public health

Data Science Crosses All the Science Supported by NIDA

- Areas of Focus:
 - The **integration** of existing datasets and tools with those that are being newly developed
 - Making datasets **findable, accessible, interoperable, and reusable** (FAIR)
 - The **development** and/or **improvement** of statistical and analytical **methods** and **tools**
 - Data **storage** and **management**
 - Promoting **stewardship** and **sustainability**

Data Infrastructure



Goals:

- Optimize data storage and security
- Connect NIH data systems

Implementation:

- The Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative
- Researcher Auth Service Initiative

The STRIDES Initiative: Harnessing the power of the cloud in support of biomedical research

STRIDES Initiative Partnerships

- A series of public-private sector relationships to obtain cloud-based storage, computing, and related services at cost-effective rates
- Available to:
 - NIH
 - NIH-funded researchers






Google - July 2018

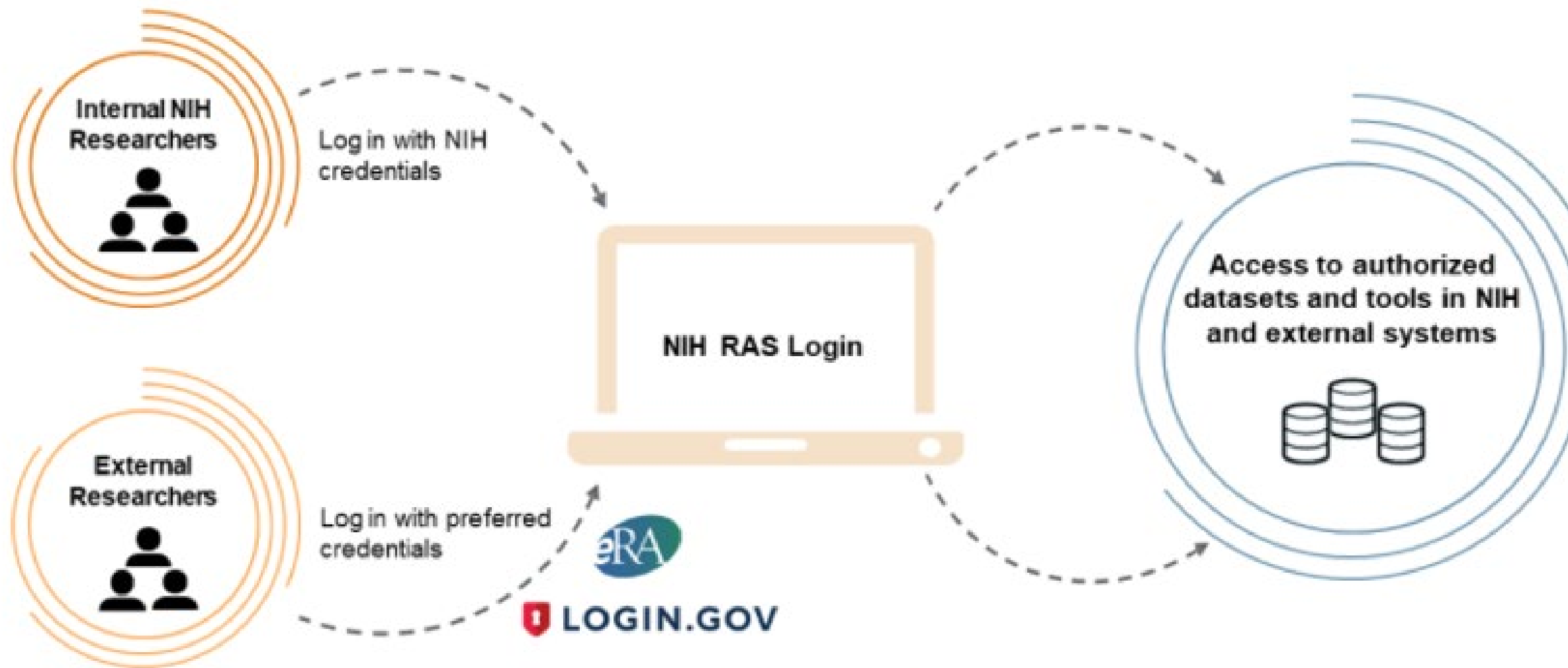
Amazon – September 2018



Azure – July 2021

Researcher Auth Service Initiative: Providing researchers a consistent and user-friendly way to access NIH's open and controlled data assets and repositories

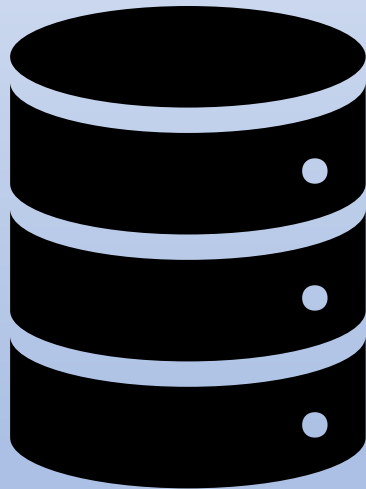
SERVICE	Authentication 	Authorization 	Auditing 
PROBLEM	Researchers are constantly challenged to provide multiple or duplicative credentials	Authorizations for access to NIH data are not communicated in a standardized way for researchers inside and outside NIH	Logging of data and events is not standardized across systems
PLANNED SOLUTION	No matter what preferred credentials researchers enter, their account identity and attributes are transferred	Comply with open standards to facilitate interoperability across the NIH data ecosystem, so researchers can securely access datasets they need and expect in ways they understand; facilitate a web of trust	Trace and log data in a standard way to protect staff and intellectual property



Integrate account information from multiple platforms

Federated Identity Broker to provide a unified, efficient and secure authentication, authorization and auditing mechanism

Data Ecosystem



Goals:

- Modernize the data repository ecosystem
- Support the storage and sharing of individual datasets
- Leverage ongoing initiatives to standardize data and adopt common data elements

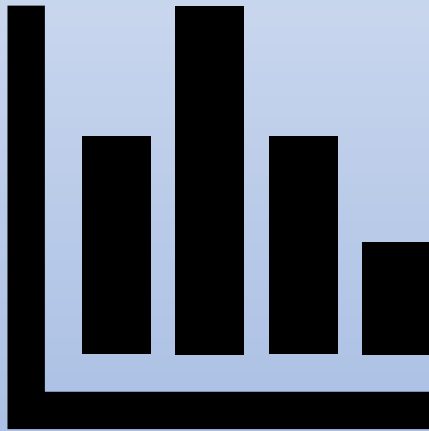
Implementation:

- Workshops supporting FAIR data principles
- Biomedical Data Repositories & Knowledgebases Initiatives

Funding Announcements of Interest

- [PAR-23-236](#): Early-stage Biomedical Data Repositories and Knowledgebases (R24 Clinical Trial Not Allowed)
 - Supports the development of early-stage or new data repositories and knowledgebases that could be valuable for the biomedical research community
- [PAR-23-237](#): Enhancement and Management of Established Biomedical Data Repositories and Knowledgebases (U24 Clinical Trial Not Allowed)
 - Designed to support established biomedical data repositories and knowledgebases as distinct and separate resources that have demonstrated impact and have potential for continued benefit to the community served

Tools and Analytics



Goals:

- Supporting useful, generalizable, and accessible tools and workflows
- Broadening utility, usability, and accessibility of specialized tools
- Improving discovery and cataloging resources

Implementation:

- Funding announcements
- Artificial Intelligence/Machine Learning

Funding Announcement of Interest

- [RFA-DA-23-039](#): The BRAIN Initiative: Theories, Models and Methods for Analysis of Complex Data from the Brain (TMM)

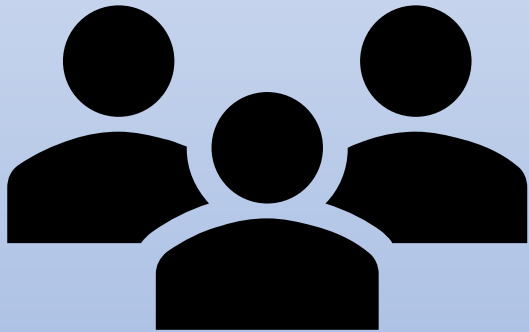
Overall Goals:

- To develop analytical tools for understanding brain function from complex neuroscience data
 - **Theories** to organize/unify data and infer general principles
 - Mathematical/statistical **models** to drive testable hypothesis
 - **Methods** to either support or refute a stated hypothesis about brain function
- Develop and deliver analytical tools to be used by the larger neuroscience community
- Focus on analytical tools for analyzing behavior & functional brain circuits that **include cellular & sub-second resolution**

Desired Outcomes:

1. Disseminatable Tools
 - Theories, models or methods for understanding brain circuits
 - Analytical tools to facilitate other projects
2. Research partnerships that strongly integrate truly diverse expertise
 - Theorists, modelers, data scientists, experimentalists

Community Engagement



Goals:

- Innovative collaborations
- Community outreach

Implementation:

- Meetings
- Showcasing NIH-funded research

Workforce Development



Goals:

- Enhance data science skills at NIH and the broad research community

Implementation:

- Funding Opportunities
- Civic Digital Fellowship
- DATA Scholar Program

Data Science Workforce Development at NIDA

- Increasing the capacity of experts in computation, data science, and related fields to move into the biomedical research space is an essential component of the NIH Strategic Plan for Data Science
- [NIDA's Strategic Plan](#) outlines how big data science can be leveraged to reveal new aspects of addiction biology and is closely aligned with the [NIH Strategic Plan for Data Science](#)
- Currently implementing multiple strategies to enhance the data science workforce at NIH, expand the national research workforce, and engage a broader community in the biomedical and clinical research fields

Funding Announcement of Interest

- [NOT-DA-21-013](#): Notice of Special Interest (NOSI): High-Priority Interest to Enhance Data Science Research Training in Addiction Research
- The purpose of this Notice is to highlight NIDA's high-priority interest in receiving applications that will support training and career development in Big Data and Computational Science (i.e., Data Science) within the overall field of addiction research.
- Grants supported under this Notice can be from individuals at formative career stages or from established investigators proposing training programs that will support early career scientists.
- Individuals from diverse backgrounds, including those from groups historically underrepresented in the STEM fields, who wish to pursue further studies or careers in Big Data and Computational Science as it relates to addiction research are highly encouraged to apply.

Benefits of Artificial Intelligence/Machine Learning (AI/ML) Technologies

- **There has been an increased focus recently on broadening the benefits of AI/ML technologies to:**
 - Reduce health disparities and inequities
 - Enhance the diversity of the AI/ML workforce

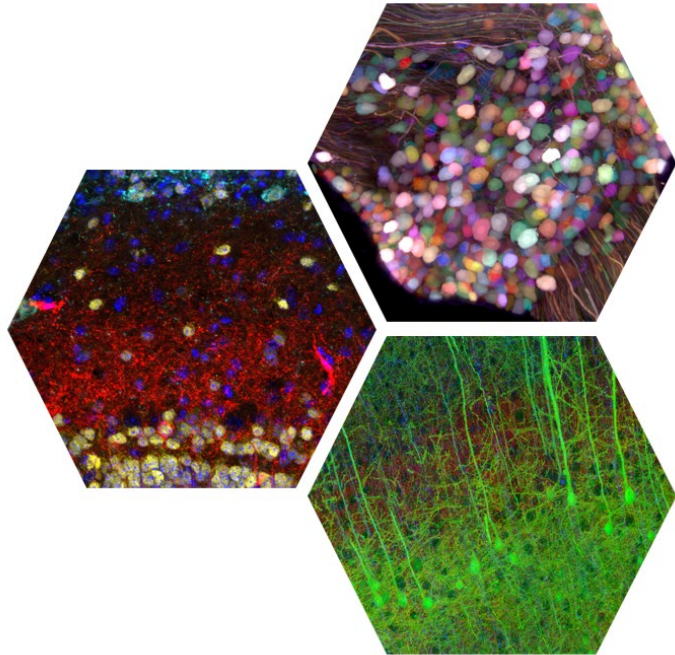
Data science doesn't require a massive wet-lab infrastructure and can provide opportunities for training future investigators in data science

Challenges of Artificial Intelligence/Machine Learning (AI/ML) Technologies

- **However, more widespread use of AI/ML technologies is hindered for several reasons:**
 - Cost
 - Capability for widespread application
 - Access to appropriate infrastructure, resources, and training
 - Lack of diversity in both data and researchers in the AI/ML field
 - Underrepresented groups often lack financial, infrastructural, and training capacity to apply data science approaches to research questions of interest to them

Additional Data Science Challenges/Considerations

- Statistical Bias
- Biases embedded in data or in machine learning algorithms
 - May inadvertently lead to racial disparities being perpetuated and exacerbated
 - Of particular importance for addiction, which carries stigma and potential discrimination
- Ethical, legal, and social implications



Overview of BRAIN Initiative Informatics


- **The BRAIN Initiative has made significant investments to build the infrastructure that is needed to effectively share and interpret data**
- The goals of the informatics program are to:
 - Build data science or informatics infrastructure that is useful to the research community
 - Make data and tools openly available to the research community
 - Help to enhance [FAIR principles](#) of data sharing and improve the [rigor and reproducibility](#) of BRAIN Initiative research
 - Enable or facilitate secondary analysis or data mining of BRAIN Initiative datasets

Currently, the Informatics Program Consists of Three Components

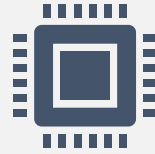
- 1) Data archives
- 2) Data standards
- 3) Software tools for data integration and analysis

Rather than building an all-encompassing infrastructure, the program is creating infrastructure for individual scientific areas:

- Integrated approaches to understanding circuit function in the nervous system
- Invasive devices for recording and modulation in the human central nervous system
- Non-invasive neuromodulation
- Next generation imaging
- Integrated approaches to cell census or atlas of the brain



Interaction and Interoperability



Data archiving, promoting data standards, and efforts around integration and analysis require interaction and interoperability



The data archives provide access and software tools, which enables users to analyze the archived data on a cloud environment without downloading the data to local machines



The data archives adopt and update data standards to re-enforce data sharing and facilitate continued rigor and reproducibility for experiments and analysis

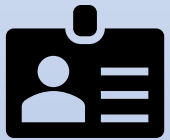
BRAIN Initiative Data Archives

- [Block and Object Storage Service & Database \(BossDB\)](#)
 - Electron microscopy and x-ray microtomography data
- [Brain Imaging Library \(BIL\)](#)
 - Confocal microscopy brain imaging data
- [Data Archive for the Brain Initiative \(DABI\)](#)
 - Human invasive neurophysiology data
- [Distributed Archives for Neurophysiology Data Integration \(DANDI\)](#)
 - Neurophysiology (ophys, ephys), behavioral time-series, and immunostaining imaging data
- [Neuroscience Multi-Omic Data Archive \(NeMO\)](#)
 - Multi-omics (genomics, epigenomics, transcriptomics) data
- [OpenNeuro](#)
 - Magnetic resonance imaging and other neuroimaging data
 - [OpenNeuroPET](#)
 - Positron emission tomography data
 - [Neuroelectromagnetic Data Archive and Tools Resource \(NEMAR\)](#)
 - Human EEG, iEEG, and MEG data

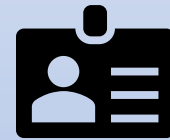
Current Funding Opportunities

-
- [RFA-MH-20-600](#): Data Archives for the BRAIN Initiative (R24 Clinical Trial Optional)
 - [RFA-MH-23-270](#): Integration and Analysis of BRAIN Initiative Data (R01 Clinical Trial Not Allowed)
 - [RFA-MH-22-145](#): Standards to Define Experiments Related to the BRAIN Initiative (R01 Clinical Trial Not Allowed)
 - [NOT-MH-23-115](#): Notice of Special Interest (NOSI): BRAIN Initiative: Developing Data Archive, Informatics Tools and Data Standards for Brain Behavior Quantification and Synchronization (BBQS)
 - [RFA-MH-23-130](#): Brain Behavior Quantification and Synchronization – Data Coordination and Artificial Intelligence Center (U24 Clinical Trial Optional)
 - [RFA-DA-23-039](#): The BRAIN Initiative: Theories, Models and Methods for Analysis of Complex Data from the Brain (TMM) (R01 Clinical Trials Not Allowed)

Contact Information



Susan Wright (NIDA)
susan.wright@nih.gov



Ming Zhan (NIMH)
ming.zhan@nih.gov

NIH/NSF Joint Initiatives

Collaborative Research in Computational Neuroscience (CRCNS)

- [NOT-MH-20-110](#): Collaborative Research in Computational Neuroscience (CRCNS) NSF Innovative Approaches to Science and Engineering Research on Brain Function
- At NIH, CRCNS is affiliated with the NIH Blueprint for Neuroscience Research
 - A collaborative framework that includes the NIH Office of the Director and 12 NIH Institutes and Centers that support research on the nervous system
- Through the CRCNS program, various agencies (within the U.S. and international) support collaborative activities that will advance the understanding of nervous system structure and function, mechanisms underlying nervous system disorders, and computational strategies used by the nervous system
 - The National Science Foundation (NSF)
 - National Institutes of Health (NIH)
 - Department of Energy (DOE)

Collaborative Research in Computational Neuroscience (CRCNS)

- Computational neuroscience provides a theoretical foundation and a rich set of technical approaches for understanding complex neurobiological systems, building on the theory, methods, and findings of computer science, neuroscience, and numerous other disciplines.
- Two classes of proposals:
 - **Research Proposals** describing collaborative research projects
 - **Data Sharing Proposals** to enable sharing of data and other resources
- NSF coordinates and manages the review of proposals jointly with participating domestic and foreign funding organizations, through a joint panel review process used by all participating funders

Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science

- [NOT-OD-23-165](#): Notice of NIH Participation in the National Science Foundation Solicitation NSF 23-614: Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science
- The Smart Health program supports **innovative, high-risk/high-reward** research with the **promise of disruptive transformations** in biomedical and public health research
- Proposals submitted must make fundamental contributions to **two or more disciplines** to improve the fundamental understanding of biomedical and health related processes and address a key health problem.
 - computer or information sciences, engineering, social, behavioral, biomedical, cognitive and/or economic sciences

Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science

- Out of Scope:
 - Traditional disease-centric medical, clinical, pharmacological, biological, or physiological studies and evaluations
 - Fundamental biological research with humans that also does not advance other fundamental science or engineering areas

Smart Health and Biomedical Research in the Era of Artificial Intelligence and Advanced Data Science

- A pre-application webinar for prospective applicants will be held via Zoom on **September 25, 2023 at 3pm E.T. Registration is required.**
- During the webinar NIH and NSF staff will provide an overview of the NSF-NIH Smart Health solicitation ([NSF-23-614](#)) and NIH Notice ([NOT-OD-23-165](#)) unique program features and requirements, the application review process, and answer questions from prospective applicants.
- Prospective applicants are invited to attend and are encouraged to submit questions in advance to maximize the use of time.
- Questions should be sent via email to sch-correspondence@nsf.gov by September 22, 2023 at 5pm E.T.

Incorporating Human Behavior in Epidemiological Models (IHBEM)

- [NOT-DA-23-017](#): Notice of Joint NSF/NIDA Initiative to Support Research in the Incorporating Human Behavior into Epidemiological Models Program
- The goal of the IHBEM joint initiative is to support interdisciplinary collaborations that integrate research on behavioral and/or social processes in mathematical epidemiological models.
- Projects should be collaborative in nature and have coordinated interaction of two or more PIs/co-PIs, with balanced participation from the mathematical sciences **and** the social, behavioral, or economic sciences.
 - Additional participants from other disciplines, especially the biological sciences, are also welcome.

Incorporating Human Behavior in Epidemiological Models (IHBEM)

NIDA is specifically interested in applications that support scientific research on drug use and its health and social consequences across the spectrum, from occasional use to problematic use and substance use disorders (SUDs). Some examples of areas of interest include:

- Using technology and advanced statistical methods to inform our understanding of both social, behavioral and neurobiological components of drug use that are strongly influenced by diverse environmental and social factors
- The development and validation of technologies, analytics, and models to help individuals gather, manage, and use data and information related to drug use and their personal health
- Methods and algorithms for aggregation of data including, but not limited to, electronic health records (EHRs), laboratory generated data, environmental, and/or behavioral data
- Diagnostic/monitoring tools and technology platforms to optimize drug use interventions and delivery

NIH Data Management and Sharing Policy

General Information

- January 25, 2023: the new NIH Data Management and Sharing (DMS) Policy replaced the 2003 NIH Data Sharing policy
 - Applies to competing grant applications that are submitted to NIH for the January 25, 2023, and subsequent receipt dates
- Under this new policy, NIH requires researchers to **submit** a Data Management and Sharing Plan, and **comply** with the Data Management and Sharing Plan approved by the funding Institute or Center
- During a research project's funding period, compliance with the DMS Plan is determined by the NIH Institute or Center
- After the end of the funding period, non-compliance with the DMS Plan may be taken into account by NIH for future funding decisions for the recipient institution

Applicability

- The DMS Policy applies to all NIH research, funded, or conducted in whole or in part by NIH, that results in the generation of scientific data
- The DMS Policy does not apply to research and other activities that do not generate scientific data (e.g., research training, fellowships, infrastructure development, and non-research activities).
- NIH has published a comprehensive listing of NIH activity codes that generally require applicants to submit a DMS Plan
- If an activity code does not appear on this list, that activity code does not require submission of a Data Management and Sharing Plan

Application Submission Guidance

- When a DMS Plan is required under a NOFO, the DMS Plan must be submitted with the application as an “Other Plan(s) attachment” (see application instructions)
- **Applications that do not submit a required DMS Plan will be deemed incomplete and will not undergo any further reviews**
- Applications subject to the Genomic Data Sharing (GDS) Policy should also address GDS-specific considerations within the DMS Plan
- DMS Plans must include, at minimum, the following:
 - How data will be managed (required)
 - How data will be shared (e.g., repository name) (if applicable), taking into account any restrictions or limitations for sharing (if applicable, after negotiations between the PO and PD/PI)
 - The timeline for data sharing
 - Any additional ICO- or program-specific requirements outlined in the NOFO (if applicable)

Six Elements That Should Be Included in DMS Plans

-
1. Data Type
 2. Related Tools, Software and/or Code
 3. Standards
 4. Data Preservation, Access, and Associated Timelines
 5. Access, Distribution, or Reuse Considerations
 6. Oversight of Data Management and Sharing

Genomic Data Sharing and Other Policies

-
- [Genomic Data Sharing Policy | Data Sharing \(nih.gov\)](#)
 - [Model Organism Sharing Policy | Data Sharing \(nih.gov\)](#)
 - [Human Data Sharing | NIH Office of Intramural Research](#)
 - [Research Tools Policy | Data Sharing \(nih.gov\)](#)
 - [Requirements for Registering & Reporting NIH-funded Clinical Trials in ClinicalTrials.gov | grants.nih.gov](#)
 - [NIH Institute and Center Data Sharing Policies | Data Sharing](#)
 - [When and How to Comply | Public Access \(nih.gov\)](#)



GREI Collaborative Webinar: Metadata Recommendations

September 15 | 2 pm ET

Register Now →



Description: Researchers and academic staff who support them, representatives from data repositories, and NIH staff interested in how metadata can make NIH-funded research more findable are invited to the GREI Metadata and Search subcommittee's Collaborative Webinar: Metadata Recommendations on September 15 at 2pm US EDT.

- At this webinar, attendees will learn about the metadata recommendations from the GREI metadata and search subcommittee, including how the recommendations came about, what we hope to achieve, and next steps. Attendees will also have a chance to share their feedback.
- Registration is free and open to all who are interested. We look forward to seeing you there!
- Webinar registration: https://cos-io.zoom.us/webinar/register/WN_0lvxKj71RXqTlwwpqS6_6Q#/registration

<https://www.drugabuse.gov/about-nida/organization/divisions/division-neuroscience-behavior-dnb/data-science-research>

Thank you!
Questions?

Contact Info:

E-mail: susan.wright@nih.gov

Office: (301) 402-6683

