



National Institute of
General Medical Sciences

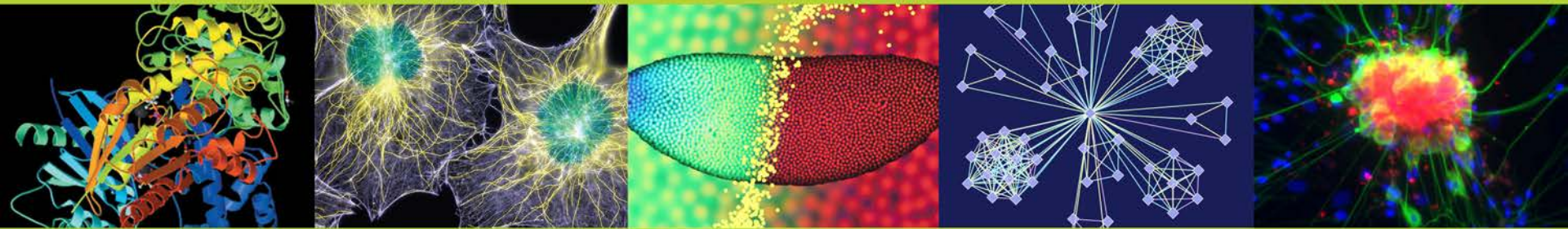


NLP, Deep learning and Data Science @ NIH

BioCreative, October 18th 2017

Susan K. Gregurick, Ph.D.

Biomedical Technology, Bioinformatics and Computational Biology



NIH Mission

To seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability

FY2018 Priorities:

1. Fundamental Science Enhanced by Technological Advances
2. Treatments and Cures
3. Health Promotion and Disease Prevention
4. Enhancing Stewardship

Brain Research through Advancing Innovative Neurotechnologies

The BRAIN Initiative:

Launched by the President in April 2013 as a bold new effort to revolutionize our understanding of the human brain

Goal: Produce a clearer, dynamic picture of the brain that can show, for the first time, how individual cells and complex neural circuits interact in both time and space

- Leverages new computational algorithms and statistical methods to simulate the complexities of neural circuitry
- Multi-agency initiative: NIH, DARPA, NSF, FDA, and IARPA

NIH BRAIN Initiative: Active Announcements*

A number of BRAIN Initiative awards are expected to begin to produce significant amounts of data in FY2018 and beyond.

- Experimental Standards FOA ([MH-17-256](#)) is to provide short term support (up to two years) to enable investigators to develop standards for new experimental protocols that are being used - or could be used - as part of the BRAIN Initiative. The application due dates for this FOA ([MH-17-256](#)) are January 10, 2017 and October 11, 2017.
- Data Archives - FOA ([MH-17-255](#)) support for the creation and management of data archives that will hold data related to the BRAIN Initiative. The data archive will make appropriate query tools and summary data easily available to allow the research community to check whether data of interest are held in the archive. The application due dates for this FOA ([MH-17-255](#)) are January 17, 2017 and October 19, 2017.
- Integration and Analysis FOA ([MH-17-257](#)) encourages the development of new data analysis and visualization tools, or to modify existing tools for BRAIN Initiative data. All tools must make use of relevant standards and will be built to enable easy integration with relevant data archives. The application due dates for this FOA ([MH-17-257](#)) are January 19, 2017 and October 26, 2017.

NIH Contact for all: **Greg Farber** (farberg@mail.nih.gov)

**There are many active announcements, <https://www.braininitiative.nih.gov/funding/initiatives.htm>*

Brain Initiative: Targeted BRAIN Circuits Projects

Develop and use innovative and methodologically-integrated approaches to understand how circuit activity gives rise to mental experience and behavior (RFA-NS-18-009, R01)

- Innovative approaches to understand network coding of sensory information in response to response to naturalistic inputs and perceptual contexts.
- Innovative approaches and new paradigms for identifying and understanding nociception and pain in the context of circuit mechanisms of the central nervous system.
- New paradigms to assess motor coding during quantifiable behaviors in real or virtual environments.
- Novel approaches to understand neural circuitry associated with well-defined social behaviors.
- Dynamic changes in functional circuit connectivity underlying the brain's ability to store information and to learn new behaviors.
- Distributed circuits that contribute to the coordination of motivational states and reward behavior.
- **New approaches to capture and assess information processing across brain regions during memory consolidation, memory retrieval, spatial/relational processing, attention, or planning.**
- **Approaches to assess distributed representations and the information processing underlying advanced mental processes such as decision making, numerical cognition, reasoning, and metacognition.**
- Empirical and analytical approaches to understand how behavioral states are emergent properties of the interaction of neurons, circuits, and networks.
- Research to advance principles of circuit function and neural systems in the central nervous system that regulate homeostasis, including biorhythms, and the balance of temperature, respiratory, energy and metabolic functions.

FY18 Priorities for NIH

Treatments and Cures:

- Support for developing new cutting-edge techniques to advance treatment for **sickle cell disease**
- Focus on **Cancer Immunotherapy** to discover new immune targets and evaluate new immune-based approaches
- Intervening in the **Opioid Epidemic** to:
 - prevent opioid misuse and addiction
 - develop new and improved treatments for opioid addiction
 - improving the deployment of evidence based strategies for combatting overdose
 - preventing and treating addiction
 - develop more effective treatments for pain with reduced potential for addiction and misuse

The Application of Big Data Analytics to Drug Abuse Research

Application of Big Data analytics to reveal deeper or novel insights into the biological and behavioral processes associated with substance abuse and addiction (PA-16-199, R01)

- Translational integration between animal and human research data by using dimensionality reduction such as principal component or factor analysis;
- **Development of software able to analyze large, complex datasets commonly acquired during drug abuse research** (e.g. longitudinal analysis of calcium imaging data over the temporal course of self-administration; analysis of temporal geospatial data from mHealth studies);
- Dimensionality reduction allowing visualization of high-dimensional data;
- Considering the results and relationships of individual studies within the broader context of all work relevant to a particular knowledge base;
- Developing reference databases and atlases with utility for drug abuse research;
- Expanding on findings conducted with grand mean or within animal averaging by using single trial analyses or other high-resolution investigations of research data;
- Investigating individual variability on self-administration behavioral data to explore resilience and vulnerability factors;
- **Automated analysis and machine learning classification of "big behavioral data," such as multiple camera and long-term video monitoring of naturalistic behaviors (e.g. in the home cage setting), recording of ultrasonic vocalizations or other behavioral measures;**
- Analysis of electronic health record (EHR) data to identify patterns in health care data that could identify those at risk for developing substance misuse or substance use disorders or those at risk of relapsing (e.g. integration of EHRs with administrative data to examine the impact of the design or performance of the service delivery system on patient outcomes);
- Developing methods to integrate and analyze multiple sources of health data (i.e., EHR, mobile device, etc.)

FY18 Priorities for NIH

All of Us Research Program:

- Tailor treatments to individual patient's characteristics. Toward this goal, NIH is establishing a group of one million or more volunteer participants that reflect the diversity of the United States to contribute health information over many years to:
 - Link data to electronic health records
 - Provide a foundation for research studies leading to prevention strategies, novel therapeutics, and medical devices
 - Improve how drugs are prescribed (choice, dosage)

Making FAIR a Reality: NIH Data Commons Pilot Phase



- 1). Develop methods that support Global Unique Identifiers (GUID) for FAIR Biomedical Digital Objects
- 2). Promote indexing metadata through FAIR compliant APIs.



- 1). Create Workspaces for Computations
- 2). Develop Scientific Use Cases



- 1). Develop cloud agnostic platforms
- 2). Develop open API standards

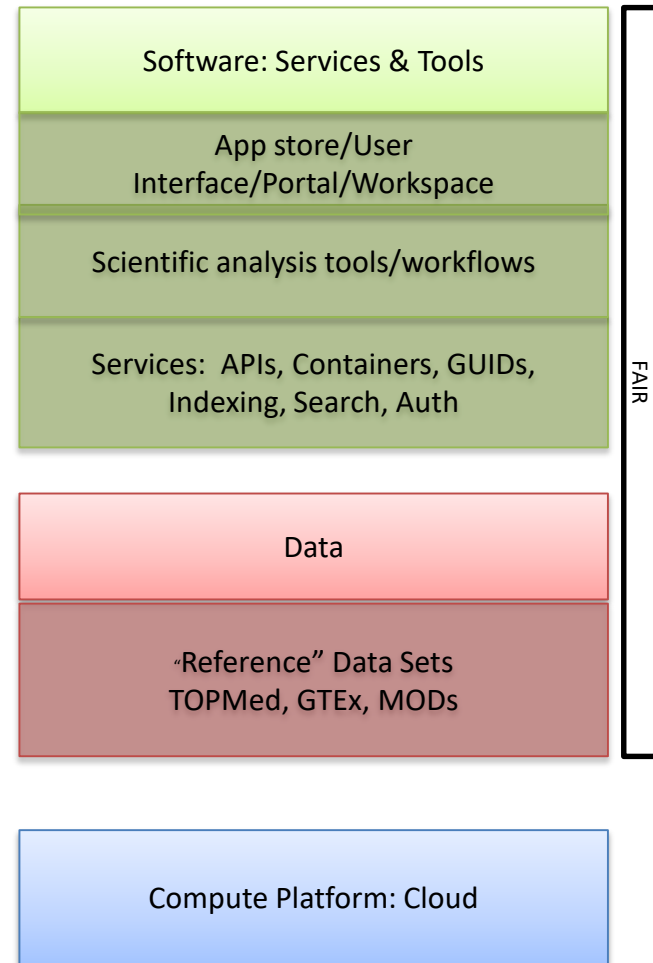


- 1). Develop FAIR guidelines and metrics
- 2). Develop approaches to address data ethics, privacy and security

Data Commons Pilot Phase Consortium (DCPPC)

Key Capabilities (KC)

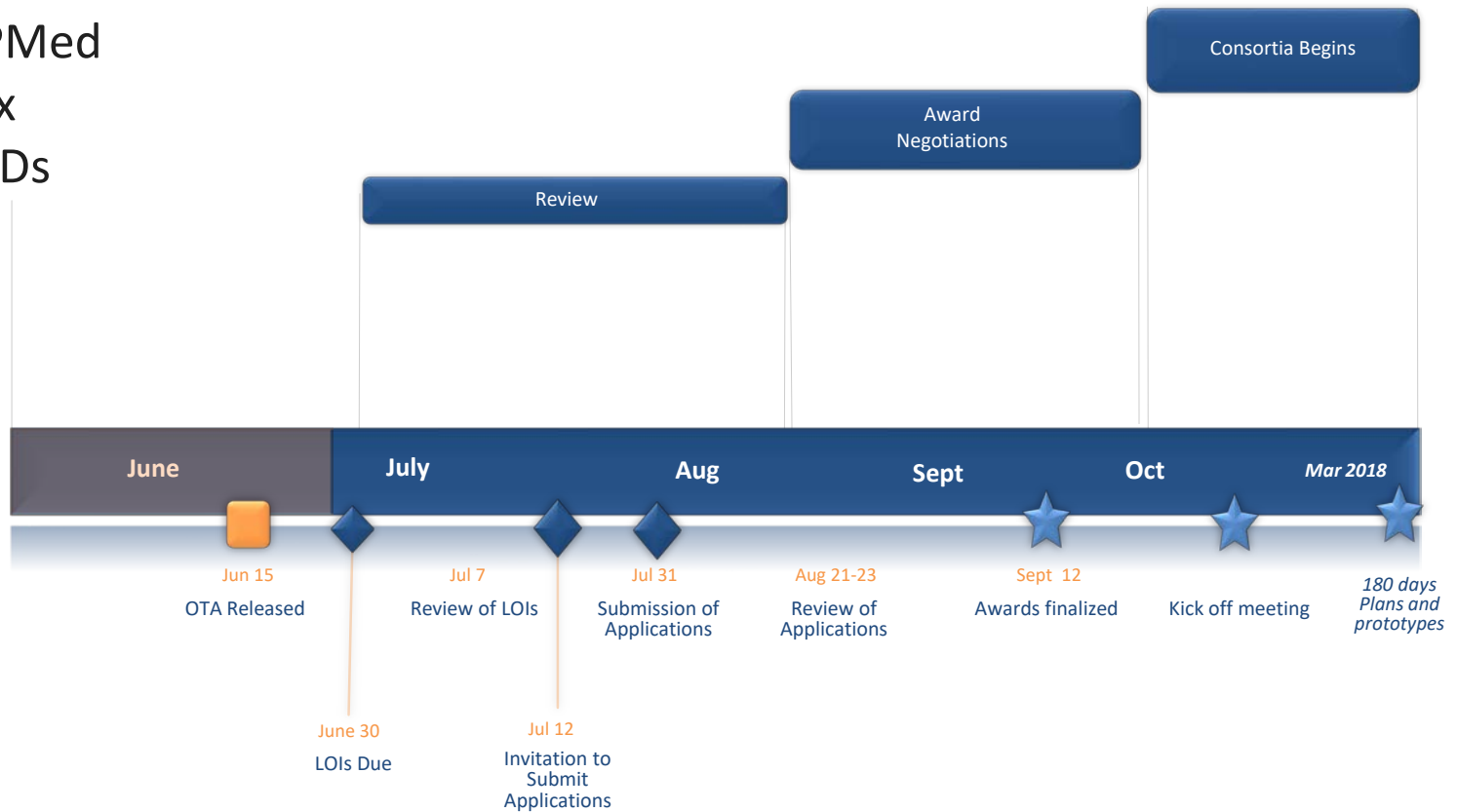
1. Development and Implementation Plan for Community Supported FAIR Guidelines and Metrics
2. Global Unique Identifiers (GUID) for FAIR Biomedical Digital Objects
3. Open Standard APIs
4. Cloud Agnostic Architecture and Frameworks
5. Workspaces for Computation
6. Research Ethics, Privacy, and Security (AUTH)
7. Indexing and Search
8. Scientific Use cases
9. Training, Outreach, Coordination



NIH Data Commons Pilot Phase

Targeted Data Sets in Pilot Phase:

- TOPMed
- Gtex
- MODs



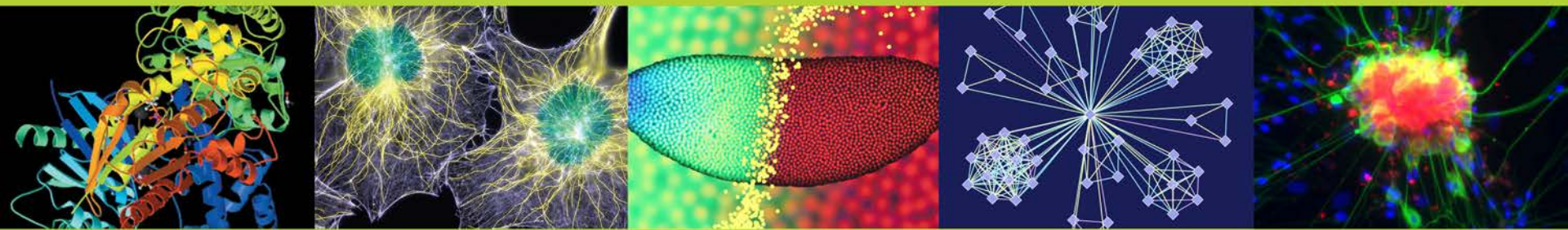
Initiatives that Support Computational and Mathematical Sciences

- **Biomedical Information Science and Technology Initiative (BISTI)**
Promote the optimal use of computer science and technology to address problems in biology and medicine by fostering collaborations and interdisciplinary initiatives (**bisti.nih.gov**)
- **Big Data to Knowledge Initiative (BD2K)**
Develop new approaches, standards, methods, tools, software and competencies that will enhance the use of biomedical data & support initial efforts toward making data sets “FAIR” Findable, Accessible, Interoperable, and Reusable(**datascience.nih.gov/bd2k**)
- **Interagency Modeling and Analysis Group (IMAG)**
Provide an open forum for communication among government representatives for trans-agency activities that have a broad impact in science (**imagwiki.nibib.nih.gov**)
- **NSF/NIH Joint program in Mathematical Biology**
Bring mathematics and statistics into the core of biological and biomedical research and to broaden the use of innovative mathematics in understanding life processes.

Computing, Informatics and Data Sciences across NIH

Networking and Information Technology Research and Development Program (NITRD)	\$1.04 billion/year in 2016
High End Computing	\$35 million/year
Large Scale Data Management & Analysis	\$148 million/year
High Capability Computing Systems Infrastructure and Applications	\$210 million/year
Software Design & Productivity	\$179 million/year
Education & Workforce	\$69 million/year

Thank You!





NIGMS Predoctoral Training Programs

- Behavioral-Biomedical Sciences Interface
- **Bioinformatics and Computational Biology**
- **Biostatistics**
- Biotechnology
- Cellular, Biochemical, and Molecular Sciences
- Chemistry-Biology Interface
- Genetics
- Medical Scientist Training Program (M.D.-Ph.D.)
- Molecular Biophysics
- Molecular Medicine
- **Pharmacological Sciences**
- **Systems and Integrative Biology**

Ruth L. Kirschstein National Research Service Award (NRSA)

- Awards honor Dr. Ruth L. Kirschstein, former Director of the National Institute for General Medical Sciences. Aside from Dr. Kirschsteins scientific accomplishments in polio vaccine development, she was a champion of research training and a strong advocate for the inclusion of underrepresented individuals in the scientific workforce
- **Individual Predoctoral MD/PhD or Other Dual-Doctoral Degree Fellowship**
PA-16-306
- **Individual Predoctoral Fellowship** (PA-16-309)
- **Individual Predoctoral Fellowship to Promote Diversity in Health-Related Research** (PA-16-308)
- **Individual Senior Fellowship** (PA-16-310)

Modeling of Infectious Disease Agents

Project Period: 3-5 years, renewable

Budget: no cap (see Parent R01);

FOA Type: PAR-17-267

Review: CSR special emphasis panel or cluster in study sections

- Focus on developing and improving mathematical, computational, and statistical methods for studying infectious disease outbreaks
- No data collection allowed
- Model verification and validation required
- PIs join network of MIDAS Investigators
 - May dynamically respond to current disease outbreaks
 - Annual PI meeting and working groups