

# “Big Data”

- The Advisory Committee to the Director (ACD) Data and Informatics Working Group
- Trans-NIH Big Data Working Groups
- Interagency Big Data Group under Networking and IT Research and Development Program (NITRD)

# Initial Charge to the ACD Working Group

- Provide the ACD and the NIH Director with expert advice on the management, integration, and analysis of large biomedical datasets. The DIWG will address the following areas:
  - Research data spanning basic science through clinical and population research
  - Administrative data related to grant applications, reviews, and management
  - Management of IT at the NIH
- Began in August 2011 and Final Report due in June 2012

# Research Data Spanning Basic Science through Clinical and Population Research

- The connection and integration of large volumes of “omics” data with other large data sets including clinical and phenotypic data
- The management and curation of these large data sets, including the use of new and emerging technologies
- The analysis of these integrated data sets to facilitate the development of more sophisticated predictive models of disease susceptibility and pathobiology
- Focusing on data domain areas and common themes
  - Molecular Profiling
  - Phenotypic
  - Imaging
  - Grant Administrative Information and IT Management

# Possible Questions to be Answered

- Define the problem space (in terms of scope and the research information lifecycle)
- What issues/challenges is the extramural community faced with when working with this type of data and/or at this point in its life cycle?
- Are the issues/challenges solvable?
- What research benefits are not being realized, because of each of the issues/challenges?

# Possible Questions to be Answered (cont.)

- What concrete recommendations can the committee make to NIH for NIH action?
  - Policy
  - Investment in standards
  - Investment in science
  - Investment in infrastructure
  - New collaborations
  - Updates to current policies, standards, investments, and collaborations
  - Others

# Request for Information

- Responses due by March 12, 2012
  - Scope of challenges/issues
  - Standards development
  - Secondary/future use of data
  - Data accessibility
  - Incentives for data sharing
  - Support needs
- Critical issue(s) and impact(s), importance, ways issues would or should affect NIH policies and processes
- [http://grants.nih.gov/grants/guide/rfi\\_files/nih\\_di/add.cfm](http://grants.nih.gov/grants/guide/rfi_files/nih_di/add.cfm)



# Membership

Name	Institution	Title
David DeMets, Co-Chair	University Wisconsin - Madison	Professor, Dept. Biostatistics & Medical Informatics
Lawrence Tabak, Co-Chair	NIH	Principal Deputy Director
Russ Altman	Stanford University	Chair, Department of Bioengineering
David Botstein	Princeton University	Director, Lewis-Sigler Institute
Andrea Califano	Columbia University	Chief of Biomedical Informatics
David Ginsburg, ACD Member; Chair, NCBI Needs-Assessment Panel	University of Michigan	Professor, Internal Medicine; HHMI
Patricia Hurn	University of Texas	Associate Vice Chancellor Health Science Research The University of Texas System
Dan Masys	University of Washington	Affiliate Professor, Department of Medical Education and Biomedical Informatics
Jill Mesirov, Ad Hoc Member, NCBI Needs-Assessment Panel	Broad Institute	Associate Director and Chief Information Officer
Shawn Murphy	Harvard University	Associate Director of Laboratory of Computer Science and Assistant Professor , Neurology
Lucila Ohno-Machado	University of San Diego	Chief Division of Biomedical Informatics



# Common Themes

- Policies and Standards
- Frameworks for Metadata
  - Data
  - Methods
- Supporting Technologies and Tools
- Biomedical Computing and Bioinformatics Workforce
- The Human Dimension