



NIH Big Data

NIST Joint Cloud and Big Data Workshop

January 16, 2013

Peter M. Lyster, Ph.D.

Program Director,

*Biomedical Technology, Bioinformatics and
Computational Biology Division*

National Institute of General Medical Sciences

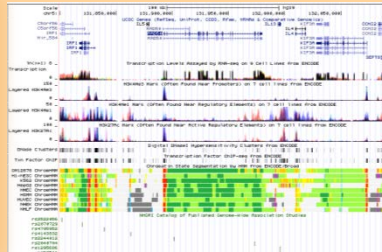
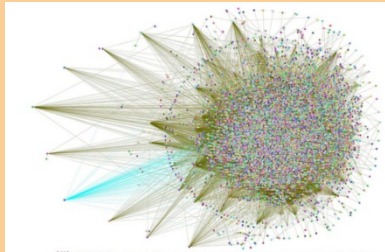


NIH Big Data

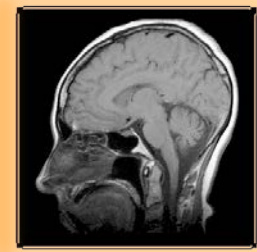
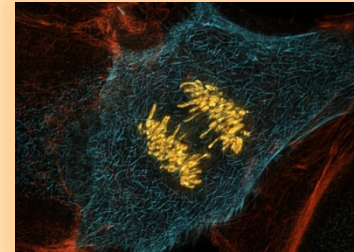
- NIH Investment in R&D related to (Big) Data
- NSF/NIH Solicitation
- BD2K: NIH plan for big data. Announcement of new position Associate Director for Data Science

Myriad Biomedical Data Types

Complex, High Volume, High Velocity



'Omic



Imaging

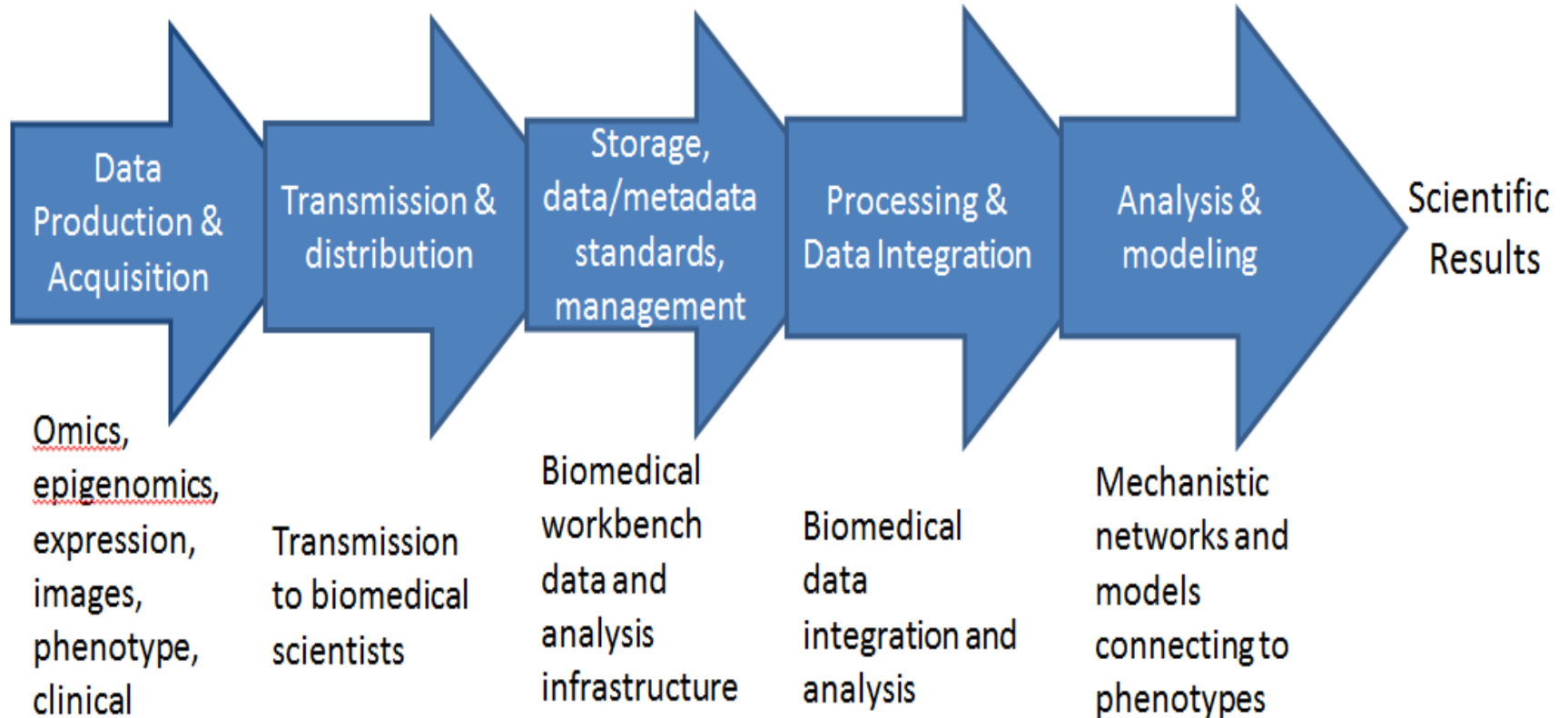


Phenotypic

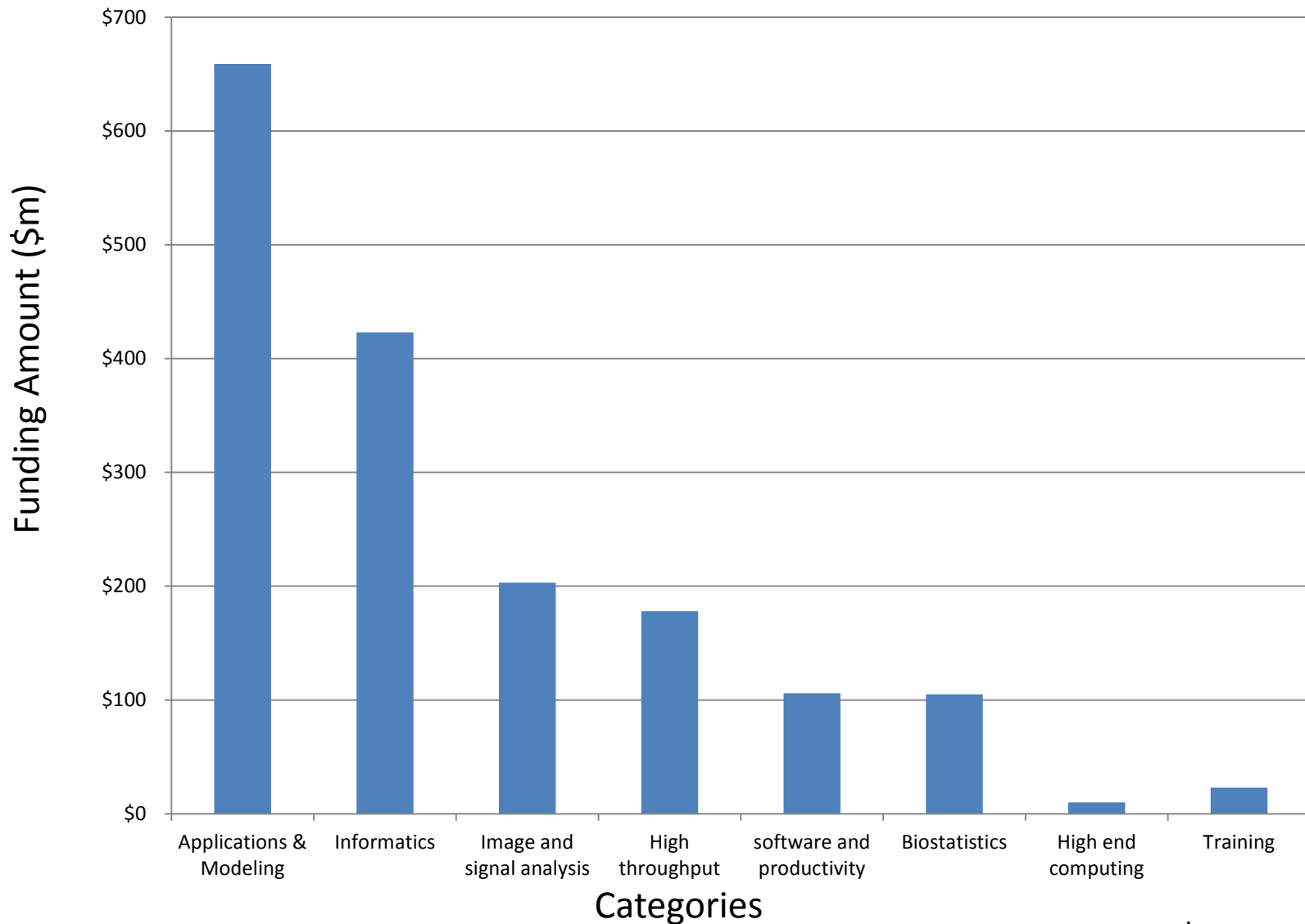


Clinical

Big Data Pipeline



Spectrum of NIH Funding of Biomedical Informatics and Computational Biology*, FY2011



*Lyster, NIGMS, spring 2012



Biomedical Information Science and Technology Initiative (BISTI)

- Trans-NIH group that coordinates research grants, training opportunities, and scientific symposia associated with biomedical computing
- <http://www.bisti.nih.gov>
- Broad-based and Institute-specific Funding Opportunity Announcements, National Centers for Computational Biology

NSF/NIH Initiative – Spring 2012

Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)

PROGRAM SOLICITATION

NSF 12-499



National Science Foundation

Directorate for Biological Sciences
Directorate for Computer & Information Science & Engineering
Directorate for Education & Human Resources
Directorate for Engineering
Directorate for Geosciences
Directorate for Mathematical & Physical Sciences
Directorate for Social, Behavioral & Economic Sciences
Office of Cyberinfrastructure
Office of Polar Programs



National Institutes of Health

National Cancer Institute
National Institute of Biomedical Imaging and Bioengineering
National Institute on Drug Abuse
National Institute of General Medical Sciences
National Institute of Neurological Disorders and Stroke
National Library of Medicine
National Human Genome Research Institute

Full Proposal Deadline(s) (due by 5 p.m. proposer's local time):

June 13, 2012

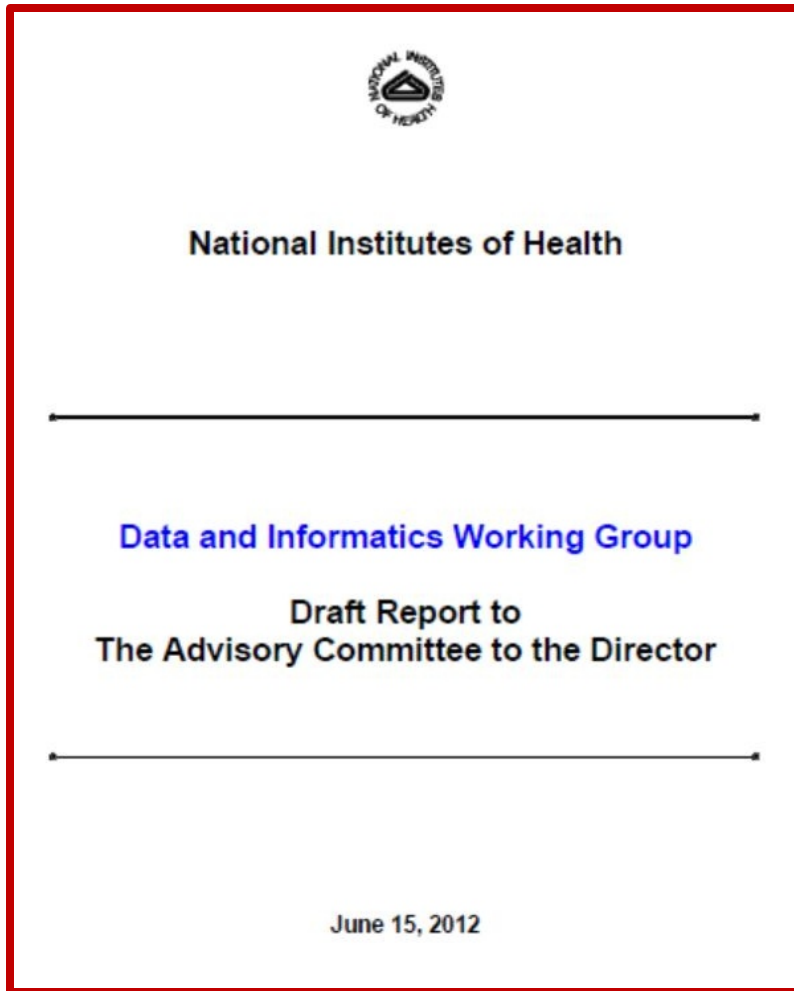
Mid-Scale Projects

July 11, 2012

Small Projects

FY12: \$15 M to develop new tools and methods to extract and use knowledge from collections of large data sets

NIH Big Data to Knowledge BD2K



- Advisory Committee to the NIH Director Data and Informatics Working Group (DIWG*) provided expert advice on the management, integration, and analysis of large biomedical datasets
- The DIWG was asked to address:
 - Research data spanning basic science through clinical and population research
 - Administrative data related to grant applications, reviews, and management
 - Management of IT at NIH

*David DeMets & Larry Tabak, Co-Chairs



Six Big Problems to Solve

1. Locating the data
2. Getting access to the data
3. Extending policies and practices for data sharing
4. Organizing, managing, and processing biomedical Big Data
5. Developing new methods for analyzing biomedical Big Data
6. Training researchers who can use biomedical Big Data effectively



Overarching Strategy and Goals

- Two initiatives being proposed to overcome roadblocks
 - *Big Data to Knowledge (BD2K)* – enable the biomedical research enterprise to maximize the value of biomedical data
 - *InfrastructurePlus* – create an adaptive environment at NIH to sustain world-class biomedical research
- Both led by Trans-NIH Advisory Data Councils

BD2K

I. Facilitating Broad Use of Biomedical Big Data

- New Policies to Encourage Data & Software Sharing
- Catalog of Research Datasets to Facilitate Data Location & Citation
- Community-based Development of Data & Metadata Standards

BD2K

II. Developing and Disseminating Analysis Methods and Software

- Software to Meet Needs of the Biomedical Research Community
- Facilitating Data Analysis: Access to Large-scale Computing
- Dynamic Community Engagement of Users and Developers

BD2K

III. Enhancing Training for Biomedical Big Data

- Increase Number of Computationally Skilled Trainees
- Strengthen the Quantitative Skills of All Researchers
- Enhance NIH Review and Program Oversight

BD2K

IV. Establishing Centers of Excellence for Biomedical Big Data

- 5-10 Investigator-initiated Centers
- 2-5 NIH-specified Centers

Sources

- NIH Advisory Committee to the Director
Working Group on Data and Informatics and
<http://acd.od.nih.gov/diwing.htm>
- Data and Informatics Implementation Team
Report
<http://acd.od.nih.gov/meetings.htm>

Questions?

Contact Peter Lyster: lysterp@mail.nih.gov