

*Advancing Health and Discovery through Big Data*



# **NIH Big Data to Knowledge** *activities in FY14 and beyond*

Jennie Larkin, PhD  
ADDs Office

ADDs Data Science Meeting  
September 3, 2014



# BD2K Vision

*To enable biomedical research as a  
**sustainable digital research enterprise**  
to facilitate discovery and support new knowledge  
and maximize community engagement.*

# Major Data Science Problems to Solve

1. Locating and citing the digital assets. **data and software discovery indices**
2. Ensuring digital assets are useful and usable. **BD2K standards activities**
3. Extending policies and practices for data sharing. **working across NIH and agencies**
4. Developing new methods to analyze and manage biomedical Big Data (computing across data types, data integration). **new data science research**
5. Training researchers who can use biomedical Big Data effectively. **workforce development**

# BD2K Status: September 2014

- The first round (FY14) of BD2K FOAs are getting ready to be paid.
  - Strong responses to BD2K RFAs for Centers, Targeted Software Development, and Training.
  - The first round of funding (almost \$32M) will be paid in September for Centers, Training, Data Indexing.
  - Truly trans-NIH management of these diverse awards.



*BD2K Executive Committee: representatives from each IC*

# BD2K Status: looking forward

- New areas for FY15 include:
  - Piloting the Data Commons
  - More diverse training activities
  - Increased focus on clinical research and standards activities
  - Increased collaboration with NIH policy experts.
- Increased Communication and Outreach
  - Within NIH, other agencies, the larger community
  - New web site, social media, blog, *what else?*

# Developing the digital research enterprise: *a community endeavor*

- Not just NIH and federal mandates
- Not only innovation from the extramural community.
- A collaboration between NIH and stakeholders in the biomedical research community

*research institutions, publishers, societies, researchers, libraries, industry, etc.*



# A Brief Review of BD2K Activities

- Establishing/Piloting the Data Commons.
- Facilitating the Broad Use of Biomedical Digital Assets: indexing and standards.
- Developing and Disseminating Analysis Methods and Software for Big Data.
- BD2K Centers of Excellence: PI and NIH-initiated.
- Training

# Result of BD2K?

- Enable a new digital enterprise that will:
  - include researchers, clinicians, computer scientists, and others who work with digital research assets.
  - recognize and support the importance of publications, data, software, and analyses.
  - ensure that knowledge and resources coming from biomedical research can be more informative and reusable.
  - Promote cultural changes in the scientific community



*More details on BD2K activities*

# Facilitating Broad Use of Biomedical Digital Assets: indexing

- NIH is developing strategy to make digital assets (data, software, etc) **discoverable** and **citable** through indexes:
  - Workshop on the Data Catalog (August 2013)
  - Data Discovery Index Coordination Consortium RFA (funding in September 2014)
  - Workshop on Data and Metadata Standards and Frameworks (September 2013)
  - Workshop on Software Indexing (April 2014)

# BD2K Data Discovery Index Coordination Consortium

FOA Name	FOA #	BD2K activity	FY
DDI-Coordinating Consortium	RFA-HL-14-031	sharing	2014
DDI-CC Admin Supplements	NA	sharing	2014

- Engage the diverse stakeholders to address the challenge of tagging digital assets with unique identifiers, to make them findable and citable.
- DDI-CC pilot activities should collaborate with ADDS data commons and cloud pilots.
- The DDI will help support an incentive/reward system for data sharing and to support development of new metrics.

# Facilitating Broad Use of Biomedical Digital Assets: standards

- FY15: NIH Standards Information Resource
  - RFI: Input on Information Resources for Data-Related Standards Widely Used in Biomedical Science
  - <http://grants.nih.gov/grants/guide/notice-files/NOT-CA-14-054.html>
  - Contact: Sherri De Coronado, NCI
- FY15: Workshop for Community-Based Framework for Data and Metadata Standards Development
  - Contact: Allen Dearry, NIEHS

# Federal Science Policy Changes

- Federal Agencies are working to make digital assets from federally funded research available.
  - *Public Access to Data Memo:*  
[http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)
  - Applies to publications and digital scientific data
  - Agencies must develop a strategy to:
    - leverage existing archives (where appropriate)
    - foster public-private partnerships with scientific journals relevant to the agency's research
- Other policy changes being considered to support data sharing (genomic data sharing, dbGaP, clinical trials, etc.)

# **Developing and Disseminating Analysis Methods and Software for Biomedical Big Data**

- BD2K has released Targeted Software Development FOA. Jennifer Couch and Dave Miller, NCI.
- Planning a workshop on gaming and community-based software development for big data. Jennifer Couch and Dave Miller, NCI.
- Piloting instances of Data Commons on public cloud providers. Vivien Bonazzi ADDS, and George Komatsoulis, NLM.

# BD2K Centers

FOA Name	FOA #	BD2K activity	FY
PI-initiated BD2K Centers of Excellence	RFA-HG-13-009	centers	2014
BD2K-LINCS Perturbation DCIC	RFA-HG-14-001	centers	2014

- Will bring innovation and expertise from the community to critical Data Science challenges
- Accessing, handling, integrating, and analyzing big data
- Cloud-based activities will help pilot the Data Commons
- Will foster workforce development with critical data science skills in a research-based setting.

# BD2K Centers

- Developing a coordinated, trans-NIH plan for administration:
  - Administered by NHGRI, NIGMS, NIBIB, NIAID.
  - Will have a Science Officer from diverse ICs assigned to each Center.
  - Will manage the entire program coherently to ensure coordination.
- No new PI-initiated centers in FY15. Will assess development of BD2K, Commons, and the Centers and identify new opportunities.