



*NIH Big Data to  
Knowledge (BD2K)*

# BD2K Update

Philip Bourne, PhD, FACMI  
Associate Director for Data Science

Advisory Committee to the NIH Director  
December 11, 2015

<http://datascience.nih.gov>

Slides: <http://www.slideshare.net/pebourne>



# BIG DATA<sup>to</sup> KNOWLEDGE

*All Hands Grantee Meeting*

Natcher Conference Center  
National Institutes of Health, Bethesda, MD



BD2K

November 12 - 13, 2015

## *One Year and Counting...*

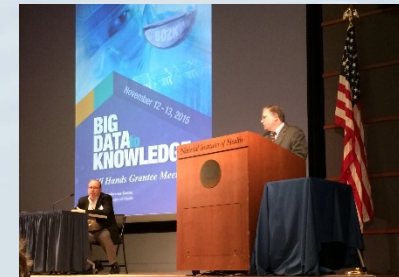
- ❑ 439 participants
- ❑ 167 remote viewers
- ❑ Breakout sessions
- ❑ 133 Posters
- ❑ 16 Demos
- ❑ 3 BOFs



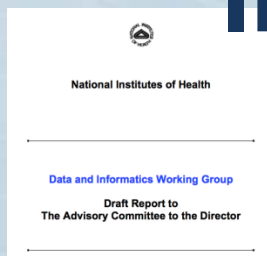




- “From the meeting, it was amply clear that NIH has the big data waterfront well-populated. From imaging, to molecular, to clinical, to mobile, BD2K has the A teams.” – **Zak Kohane, Harvard**
- "I have been involved in several national initiatives to bring advanced technology into biomedical research. I have never seen one with such an intense drive and uptake as the BD2K program. This stems not only from excellent leadership and vision, but also from the immediate impact of the centers.” - **Scott Delp, Stanford**
- 'BD2K has already changed the landscape of biomedical research in the USA. The All-hands meeting captured the excitement and change in culture that is happening across biomedical science, with the realisation that sharing data lies at the heart of biomedical research today and that establishing the international infrastructure to do so is critical. Great science too!!!' - **Janet Thornton, EBI**
- If people are the NIH's most valuable resource, then the BD2K centers are successfully addressing its second most valuable resource: data. - **David Haussler, UCSC**
- Amazing interest, support and excitement from the community – **Peipei Ping, UCLA**
- ‘We can now let the data lead to the discoveries and are able to do things we could not do before. Without the new scientific tools and strategies developed as part of BD2K we would remain anchored in our reductionistic past.’ - **Art Toga, USC**



# Implementing ACD Big Data Recommendations



## DIWG Recommendations

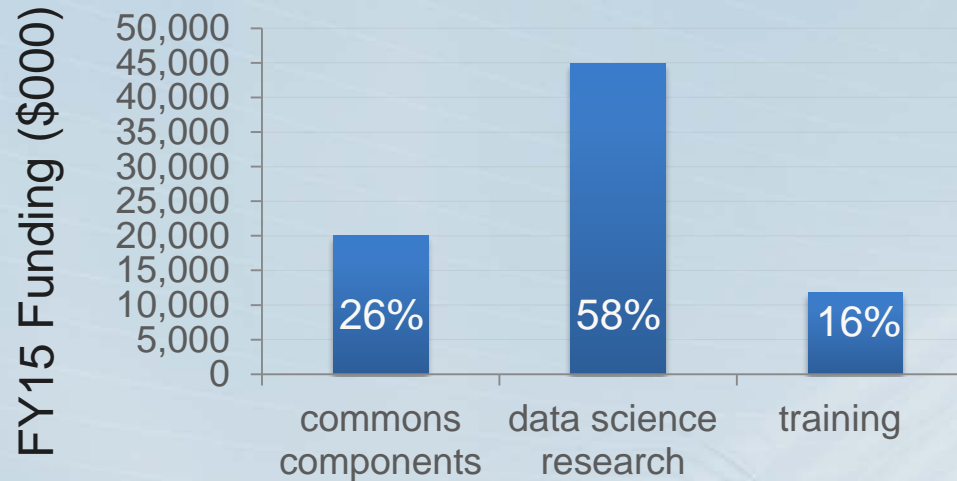
1. Sharing data & software through indexes
2. Advance big methods, tools & applications
3. Expand data science training
4. Continued support throughout the data & software lifecycle

## BD2K Implementation

1. Implement the Commons (indices, standards, etc.)
2. Data science research programs (Centers, U01s, etc.)
3. Training and workforce development programs
4. Addressing sustainability of science, technology, and funding mechanisms



# BD2K FY15 Funding for Sharing & Sustainability

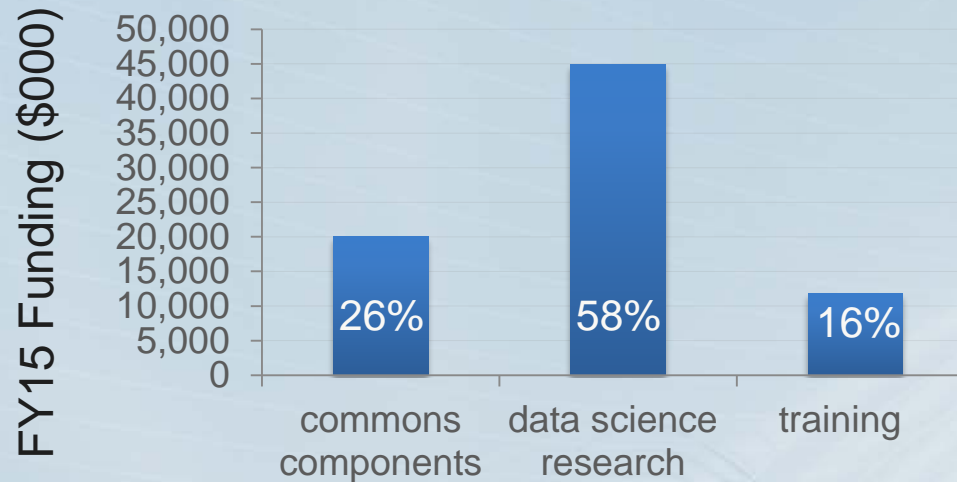


## Commons Components (\$20M)

- BioCADDIE (data discovery index prototype)
- Standards Coordinating Center contract
- Cloud Broker Model contract
- Supplements to support interoperability of NIH data repositories
- Supplements to MODs and BD2K awards to pilot Commons



# BD2K FY15 Research Funding

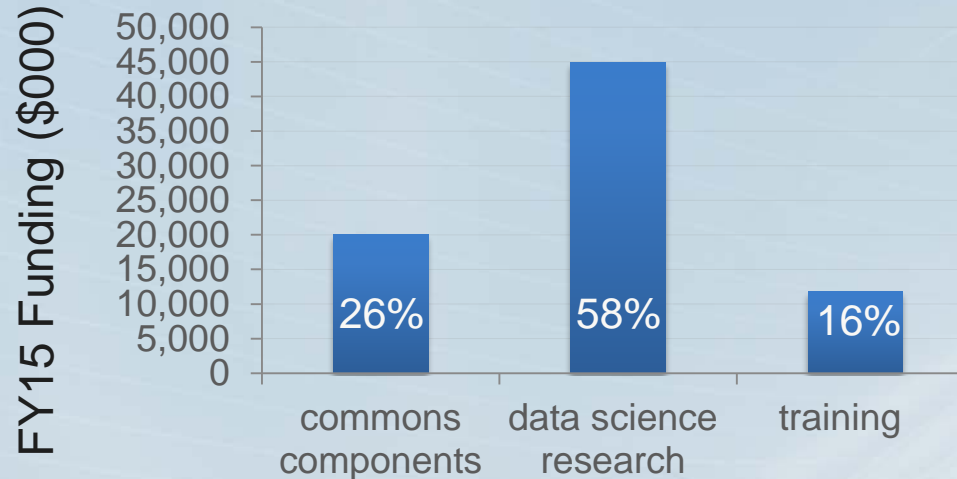


## Data Science Research (\$44.8M)

- 13 BD2K Centers awards, span scientific domains across NIH
- Targeted Software Awards on topics: data compression, visualization, provenance, wrangling.
- Innovations Lab to develop new biomedical-data science collaborative teams



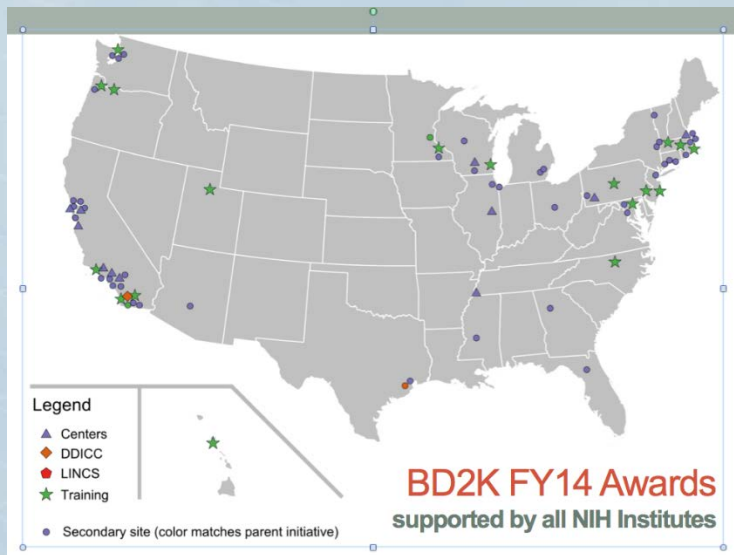
# BD2K FY15 Training Funding



## Training and Workforce Development (\$11.8M)

- Training Coordination Center
- R25 awards for MOOCS, short courses, open educational resources
- T32 training programs in data science
- K01 career development awards
- R25s MOOCS and online resources to libraries to support data management and curation
- R25 enhancing diversity in biomedical data science





# Research: Advance Big Methods Tools & Applications





# **mobilize**

Center for Mobility Data  
Integration to Insight

Scott Delp, PhD  
Department of Bioengineering  
Stanford University

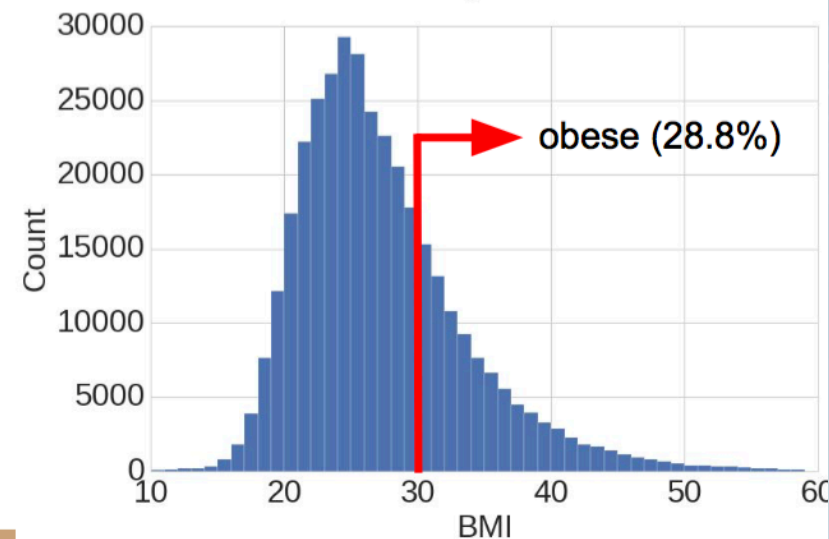
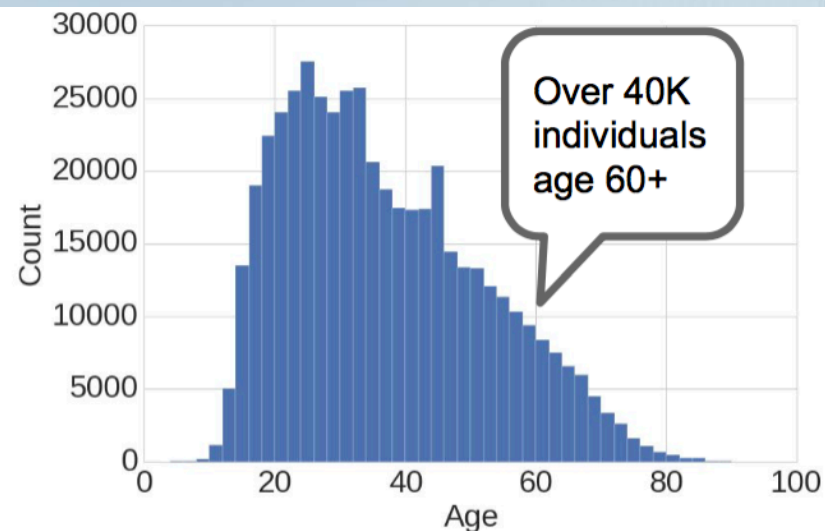
# Physical Activity from Personal to Planetary Scale

*Physical activity helps prevent heart disease, stroke, diabetes, and weight gain, but inactivity remains a worldwide public health issue.*



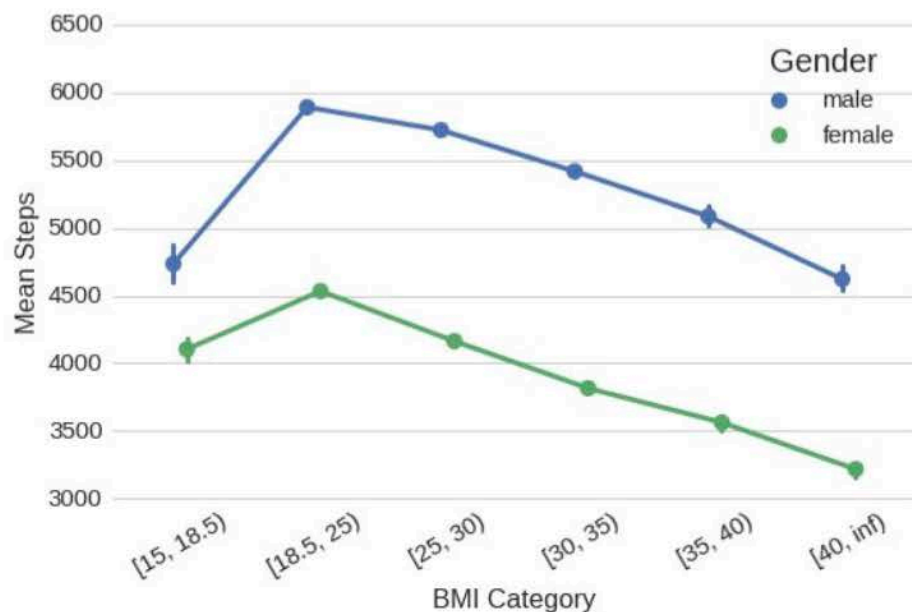
## Azumio Smartphone App

- 2M subjects worldwide and 74M days of activity
- 100B data points, which is ~1000X more than NHANES



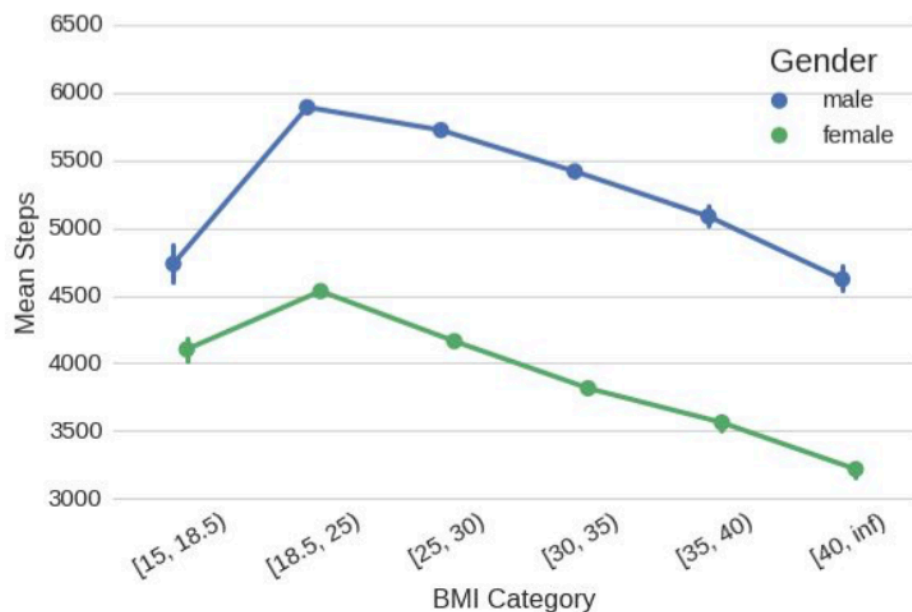
# Personal Scale: Physical Activity & BMI

**Activity** decreases with increasing **BMI** and activity is lower in females (e.g., Tudor-Locke et al., 2010).

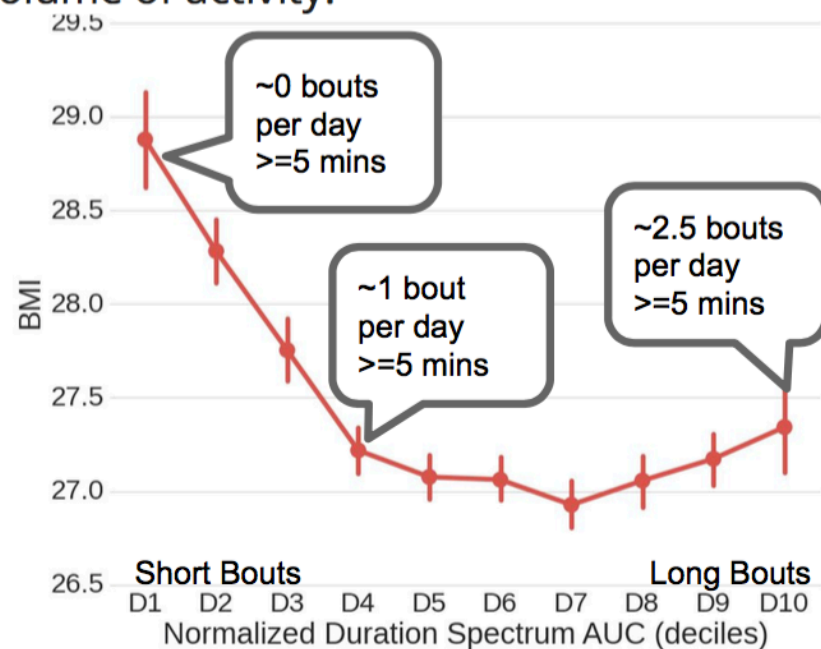


# Personal Scale: Physical Activity & BMI

**Activity** decreases with increasing **BMI** and activity is lower in females (e.g., Tudor-Locke et al., 2010).



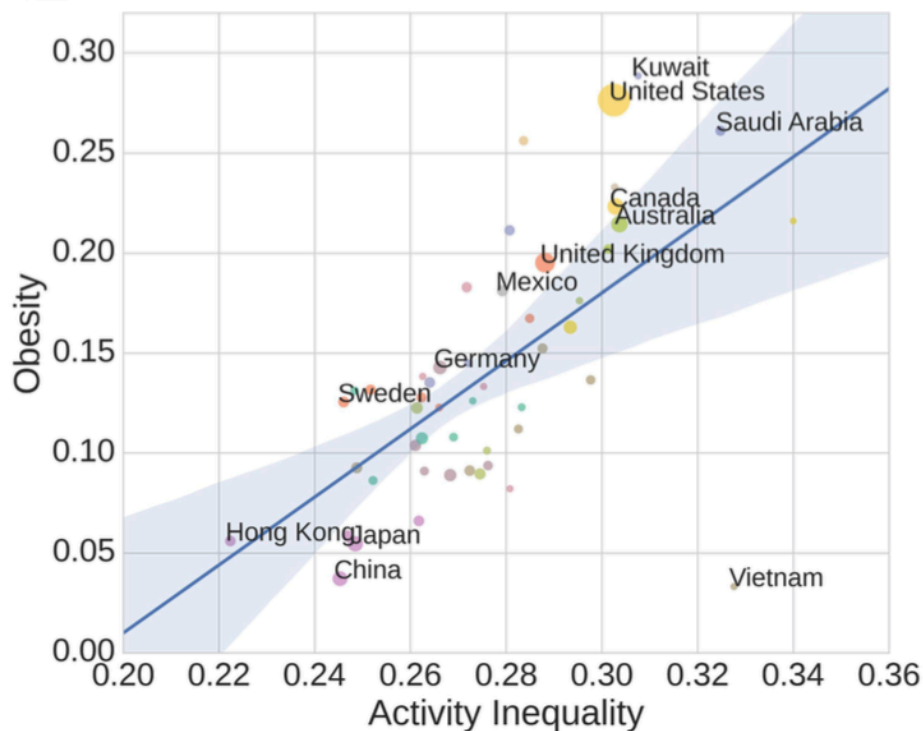
How is **bout duration** related to **BMI**?  
Control for gender, age, daily wear time, and volume of activity.





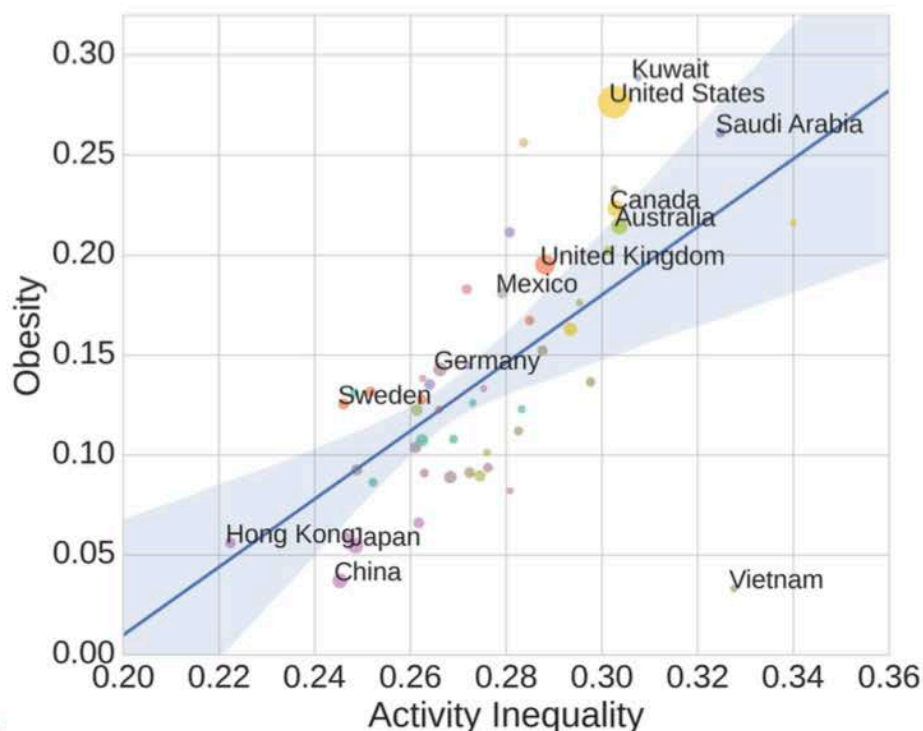
# Planetary Scale: Environment, Activity, & Obesity

How is obesity related to **activity inequality** (Gini Coefficient)?

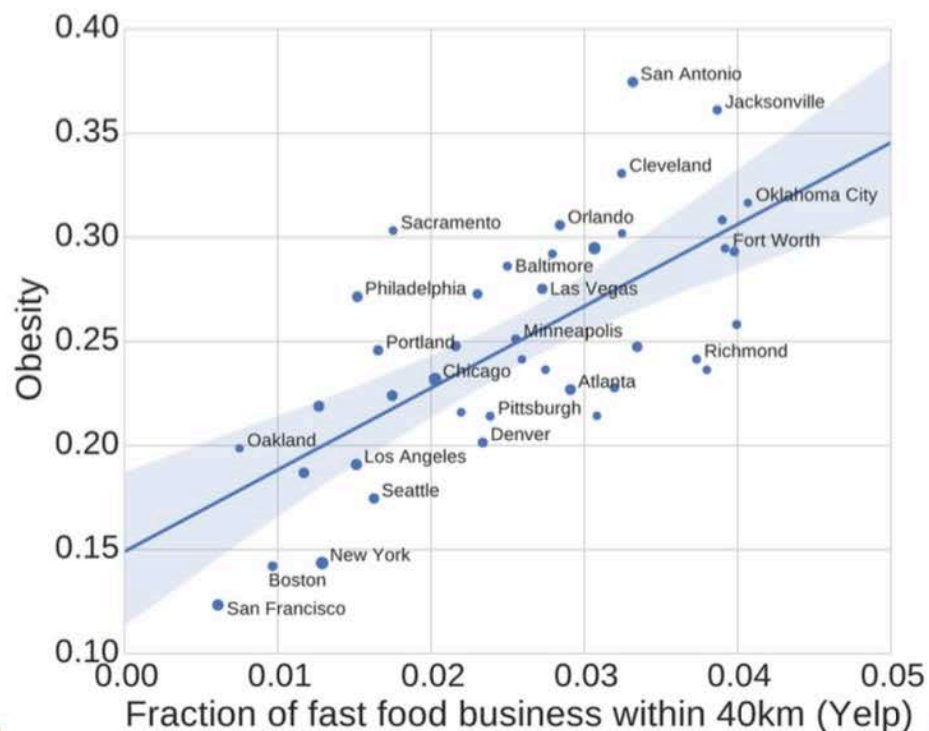


# Planetary Scale: Environment, Activity, & Obesity

How is obesity related to **activity inequality** (Gini Coefficient)?



How is obesity related to **fast food access**?

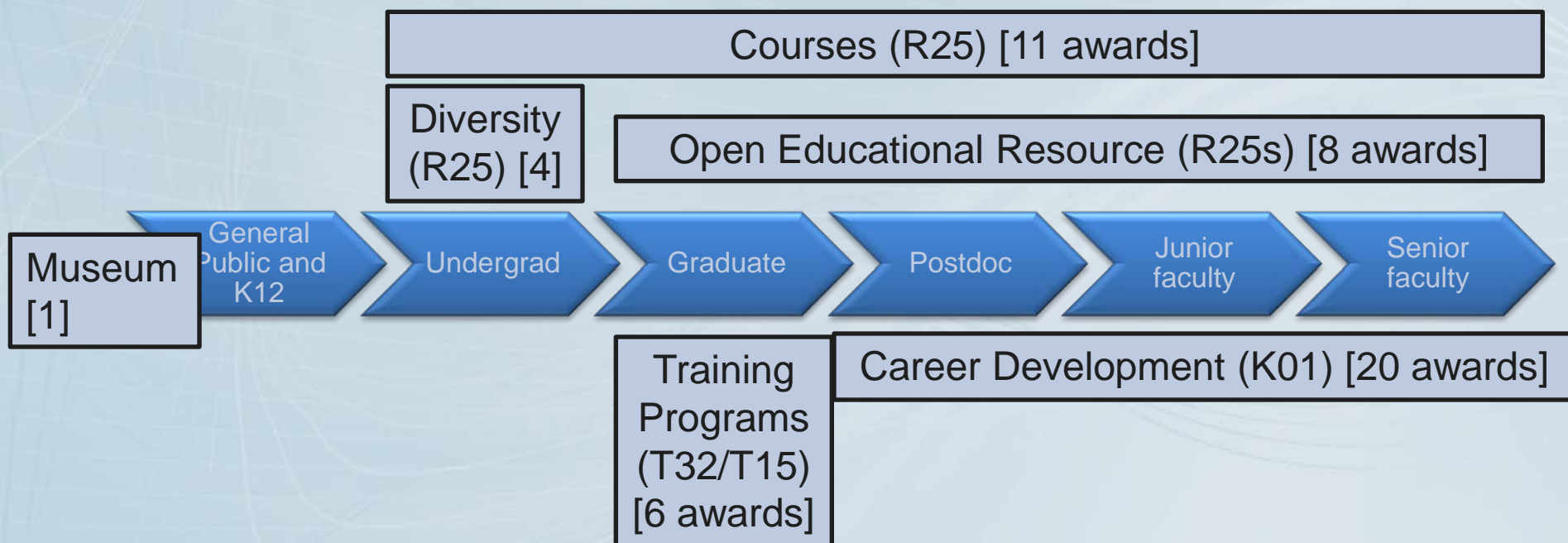


# Advance Data Science Training



# Training Programs Initiated FY14-15

## Biomedical Science Specialists



## Data Science Specialists





3.1415

MARCH 13, 2015

Porter Neuroscience Building  
(Building 35A), Room 620/630

PI DAY  
@NIH

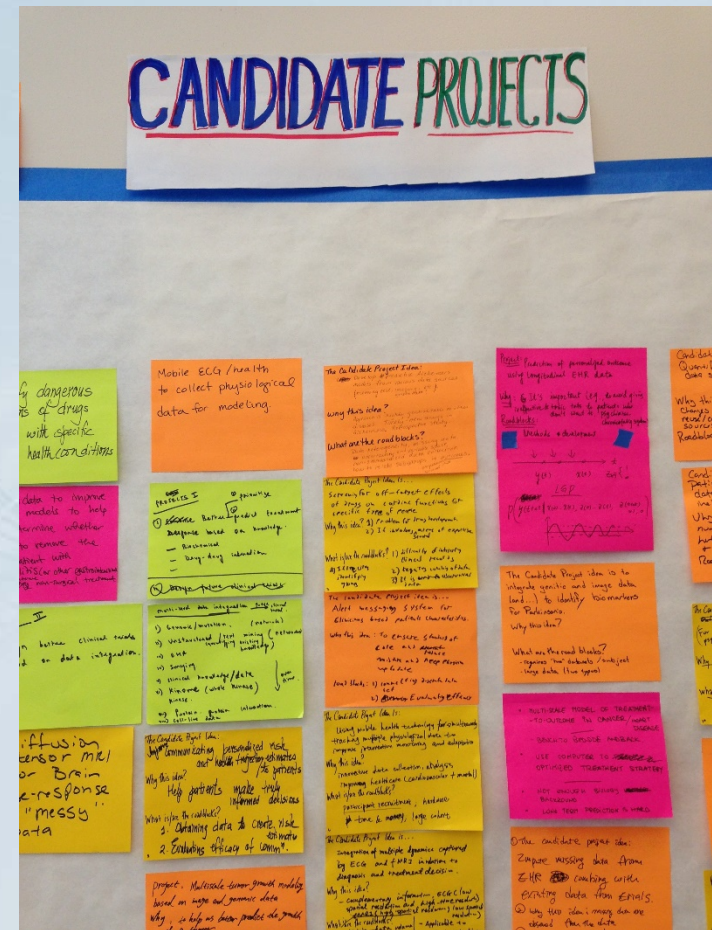
- Distinguished Lecture Series
- Frontiers in Data Science Lecture Series
- Software carpentry
- Hackathons

- 2016 Lecture by Carlos Bustamante, Ph.D.
- Posters
- PiCo Lightning Talks
- Event for High School Students
- Workshop on Reproducible Research
- Pies



# Innovation Lab

- **Description:**
  - 5-day mentored workshop facilitated by KnowInnovation
  - Joint initiative of NSF and NIH
- **Purpose:**
  - To build interdisciplinary (biomedical and data science) teams
  - To develop teams' research programs
- **Outcome:**
  - New teams formed and competed for funding
  - Innovation lab teams had a higher than average success rate





The  
**Open  
Science**  
Prize



**wellcome**trust



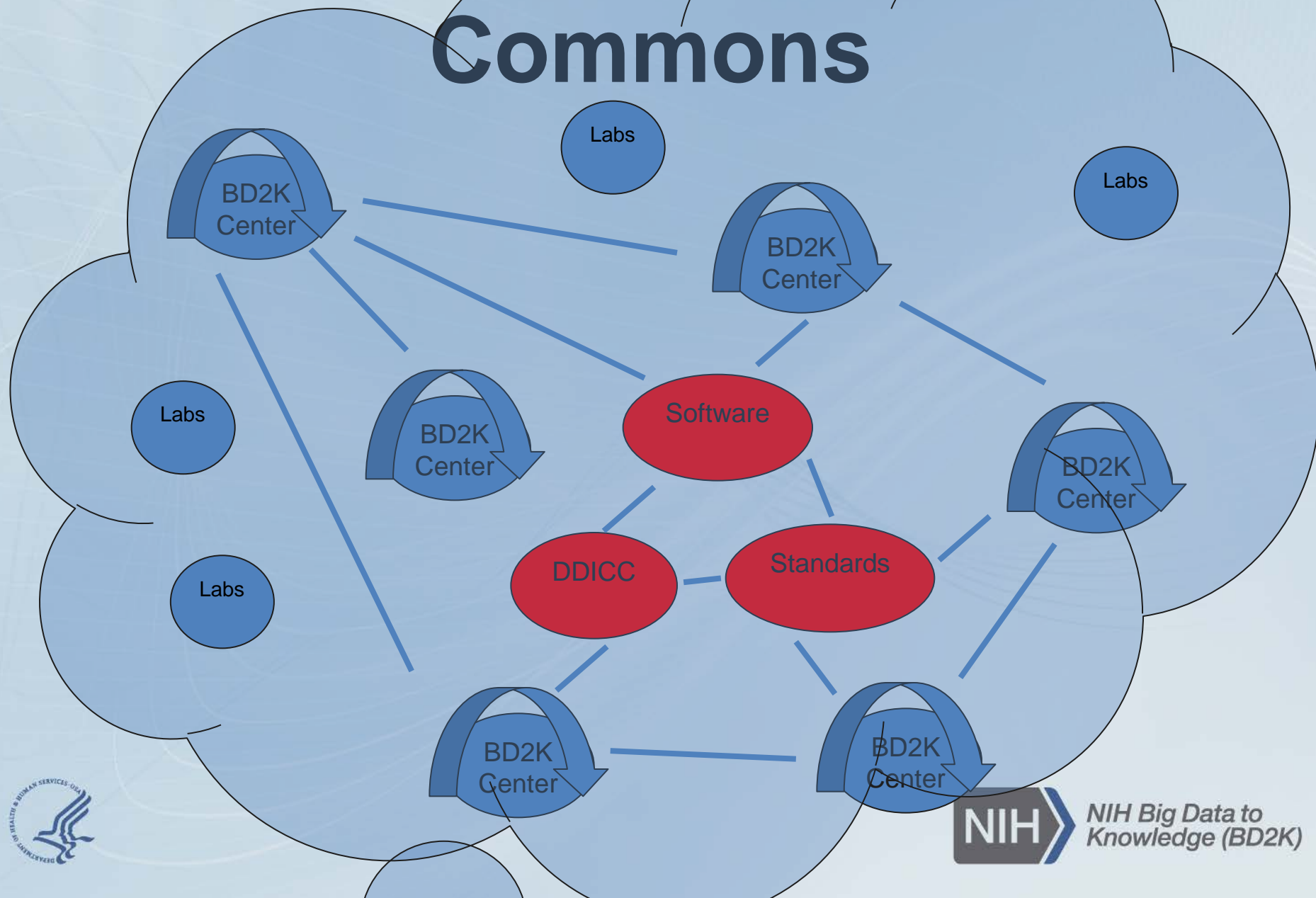


# Sharing Data & Software Through Indexes





# Infrastructure - The Commons



# Community Engagement In the Commons: Beacon

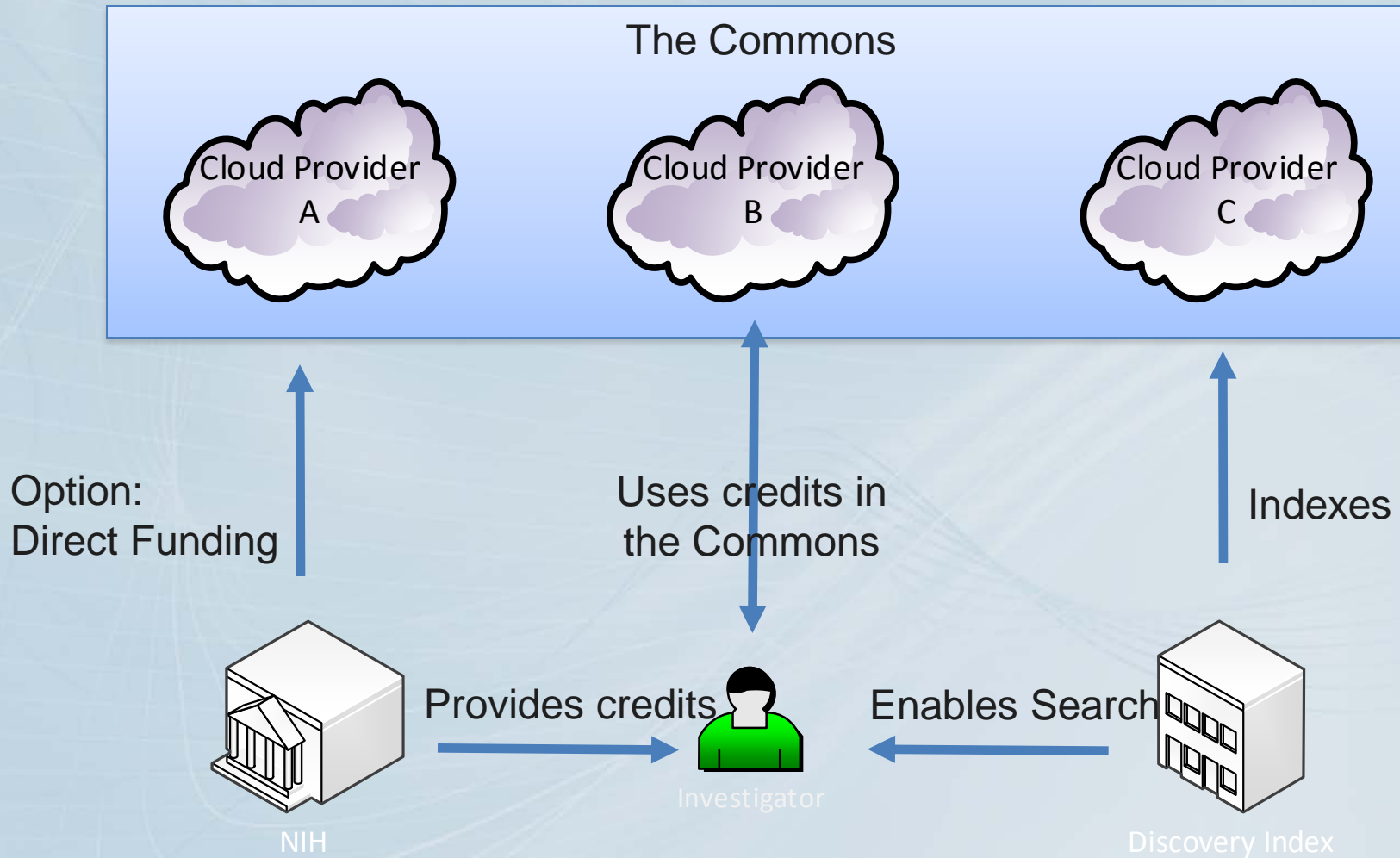
A beacon answers the simple question, have you observed a genome with a given mutation?

You can ask “Do you have a genome with an A at position 100,000 on chr1?”

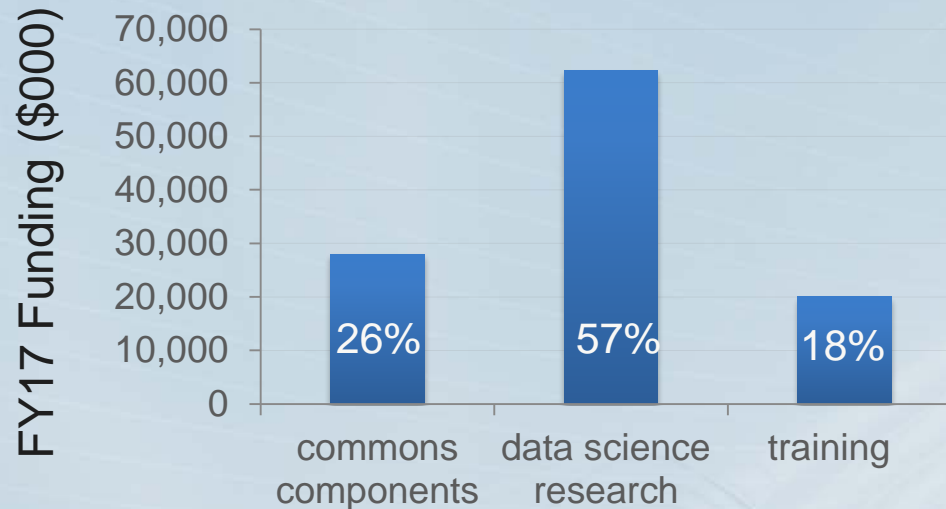


YES

# Commons Credits Model



# BD2K FY17 Funding for Sharing & Sustainability



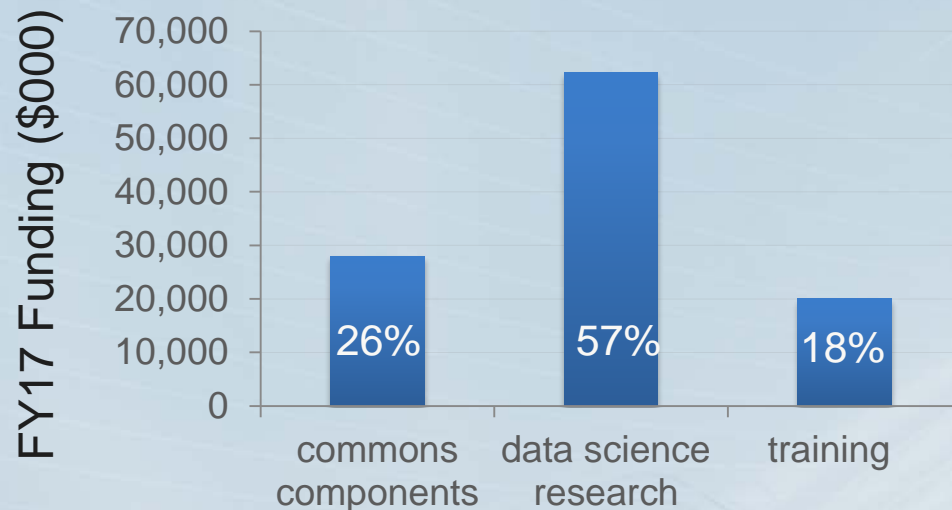
## Commons Components (\$28M)

- Resource Indexing (data, software...)
- Standards coordination and community-based development
- Cloud Broker Model contract
- Reference data sets to the cloud
- Innovations in curation RFA





# BD2K FY17 Research Funding

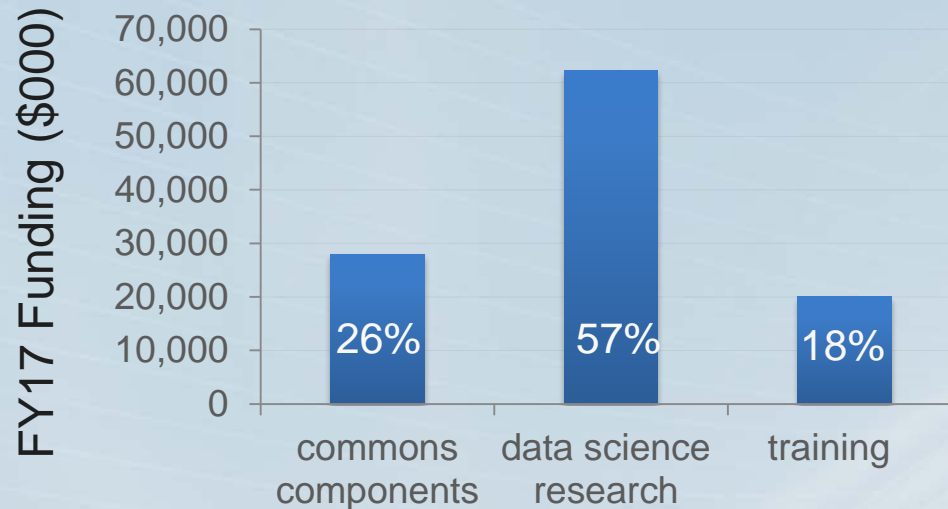


## Data Science Research (62.3\$M)

- 13 BD2K Centers awards, span scientific domains across NIH
- Targeted Software Awards on topics: data privacy, repurposing, applying metadata, interactive digital media
- Innovations Lab to develop new biomedical-data science collaborative teams
- Professional-grade software support and services in the Commons
- CDE harmonization



# BD2K FY17 Training Funding

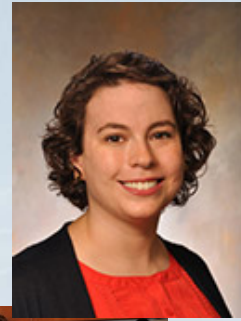
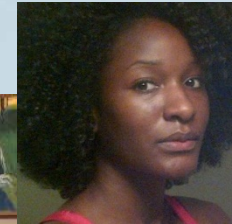


## Training and Workforce Development (20.1\$M)

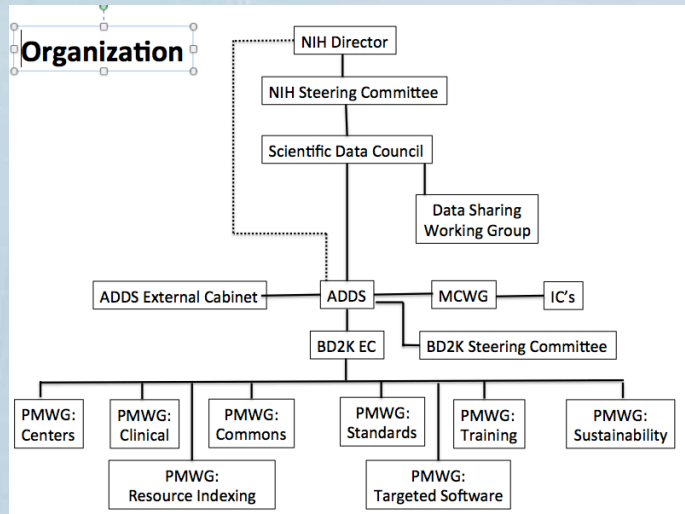
- Training Coordination Center
- R25 awards for MOOCS, short courses, open educational resources
- T32 training programs in data science
- K01 career development awards
- R25s MOOCS and online resources to strengthen data science curriculum in biomedical courses
- R25 enhancing diversity in biomedical data science



# ADDS Team



## Leadership



## IC Representatives



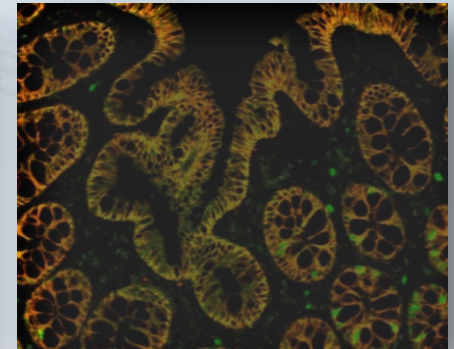
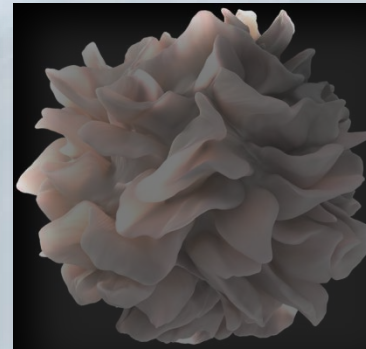
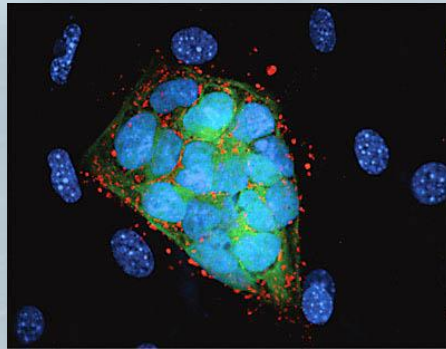
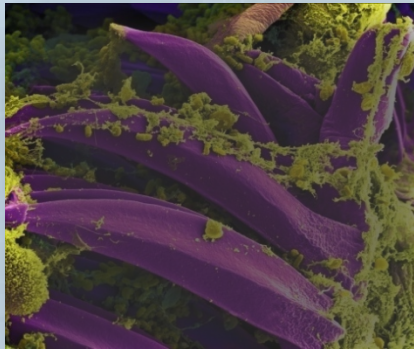




# NIH...

*philip.bourne@nih.gov*  
<https://datascience.nih.gov/>

## Turning Discovery Into Health





# Timeline Through 2021

- Advanced Tools & Applications

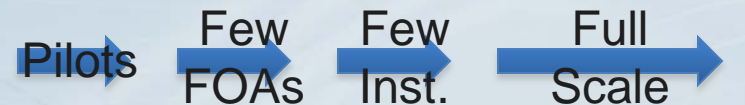
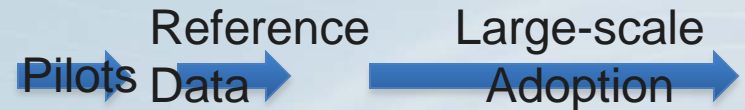
- Centers
- Software
- Other

FY 15 16 17 18 19 20 21



- Sharing Data & Software

- Commons
- Credits
- Indexing



- Training



- Sustainability

