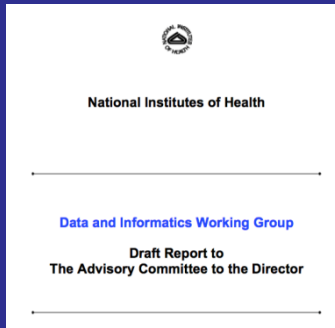


# NIH As A Digital Enterprise

Philip E. Bourne Ph.D.  
Associate Director for Data Science  
National Institutes of Health



# Data Science Timeline

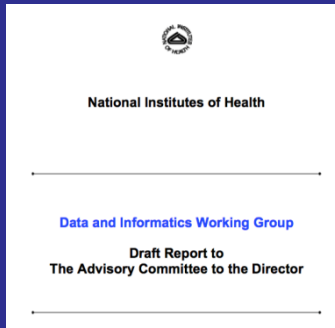


6/12

- Findings:
  - Sharing data & software through catalogs
  - Support methods and applications development
  - Need more training
  - Need campus-wide IT strategy
  - Hire CSIO
  - Continued support throughout the lifecycle



# Data Science Timeline



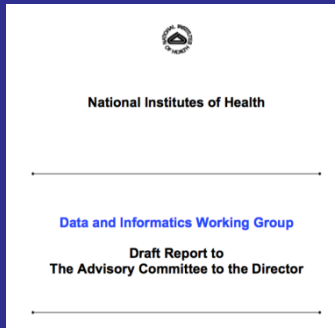
6/12

2/14

- U54 Centers of Excellence – awarded 10/14
- U54 BD2K-LINCS– awarded 10/14
- U24 Data Discovery Index– awarded 10/14
- R01, R41, R42, R43, R44, U01 software and analysis methods grants – awarded 10/14
- T32, T15, K01, R25 and R26 training awards – awarded 10/14



# Data Science Timeline



6/12

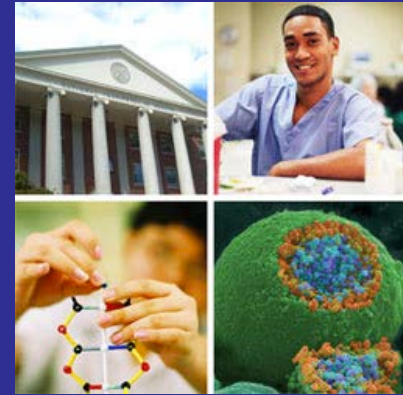
2/14

3/14

- U54 Centers of Excellence – awarded 10/14
- U54 BD2K-LINCS– awarded 10/14
- U24 Data Discovery Index– awarded 10/14
- R01, R41, R42, R43, R44, U01 software and analysis methods grants – awarded 10/14
- T32, T15, K01, R25 and R26 training awards – awarded 10/14



# Mission Statement



To foster an ecosystem that enables biomedical research to be conducted as a digital enterprise that *enhances health, lengthens life and reduces illness and disability*



# Example Components of the Ecosystem

- NIH
  - 20/27 ICs
- Agencies
  - NSF
  - DOE
  - DARPA
  - NIST
- Government
  - OSTP
  - HHS HDI
  - ONC
  - CDC
  - FDA
- Private sector
  - Phrma
  - Google
  - Amazon
- Organizations
  - PCORI
  - RDA, ELIXIR
  - CCC
  - CATS
  - FASEB
  - Biophysical Society
  - Sloan Foundation
  - Moore Foundation



# Goals & *Examples* of the Digital Enterprise

- Sustainability – *50% business model*
- Efficiency – *sharing best practices in longitudinal clinical studies*
- Collaboration - *identification of collaborators at the point of data collection not publication*
- Reproducibility – *data accessible with publication*
- Integration – *phenotype homogenization*
- Accessibility – *clinical trials registration*
- Quality – *sharing CDEs across institutes*
- Training – *keeping trainees in the ecosystem*



# Raw Materials to Build the Digital Enterprise

- NIH mandate & support
- ADDS team of 8 people
- Intramural participation of over 100 team members across ICs
- Funding through BD2K:
  - ~\$30M in FY14
  - ~\$80M in FY15
  - ....





**we have organized ourselves around  
5 thematic areas ...**



# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## *Programmatic Theme*

Sustainability

Education

Innovation

Process

Collaboration

## *Deliverable*

Commons

Training

BD2K

Efficiency

Partnerships

## *Example Features*

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists



# The Biomedical Research Digital Enterprise

# Associate Director for Data Science

Scientific Data Council

External Advisory Board

*Programmatic Theme*



Education



Innovation

F



Collaboration

*Deliverable*

Exa



NIH

# The Biomedical Research Digital Enterprise

# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## *Programmatic Theme*

Sustainability

Education

Innovation

Process

Collaboration

### *Deliverable*

Commons

Training Center

BD2K

Modified Review

Communication

### *Example Features*

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists

\* Hires made

# The Biomedical Research Digital Enterprise

# The Commons (Vivien Bonnazi & George Komatsoulis (NCBI))



- Public/private partnership
- Work with IC's, NCBI and CIT to identify and run pilots – cloud, HPC centers
- Port DbGAP to the cloud
- ? Experiment with new funding strategies
- Evaluate



# Sustainability and Sharing: The Commons

Commons == Extramural NCBI == Research Object Sandbox == Collaborative Environment

*The Why:*  
*Data*  
*Data Sharing Plans*

*The How:*

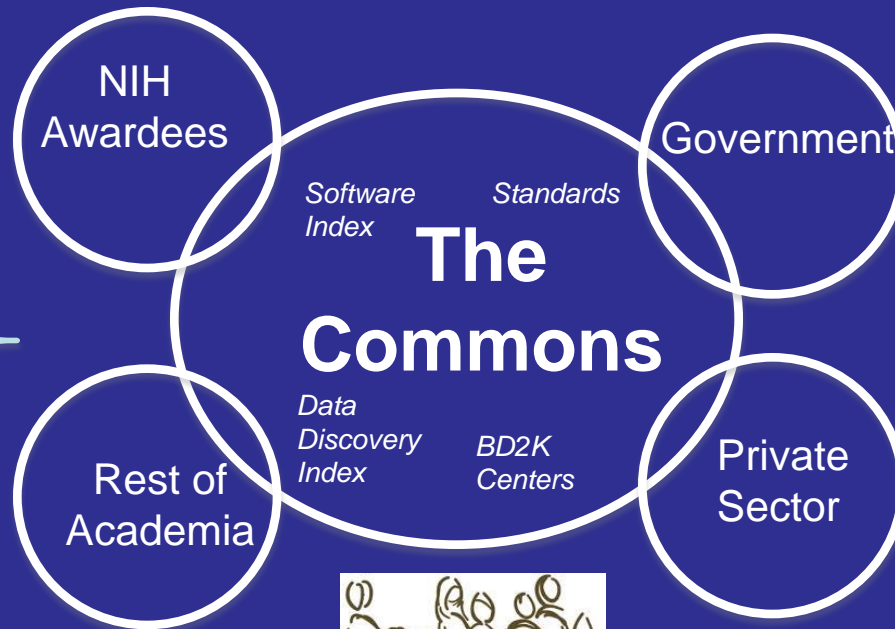
*The End Game:*



The Long Tail



Core Facilities/HS Centers



- Scientific Discovery
- Knowledge
- Usability
- Quality
- Security/Privacy
- Metrics/Standards
- Sustainable Storage

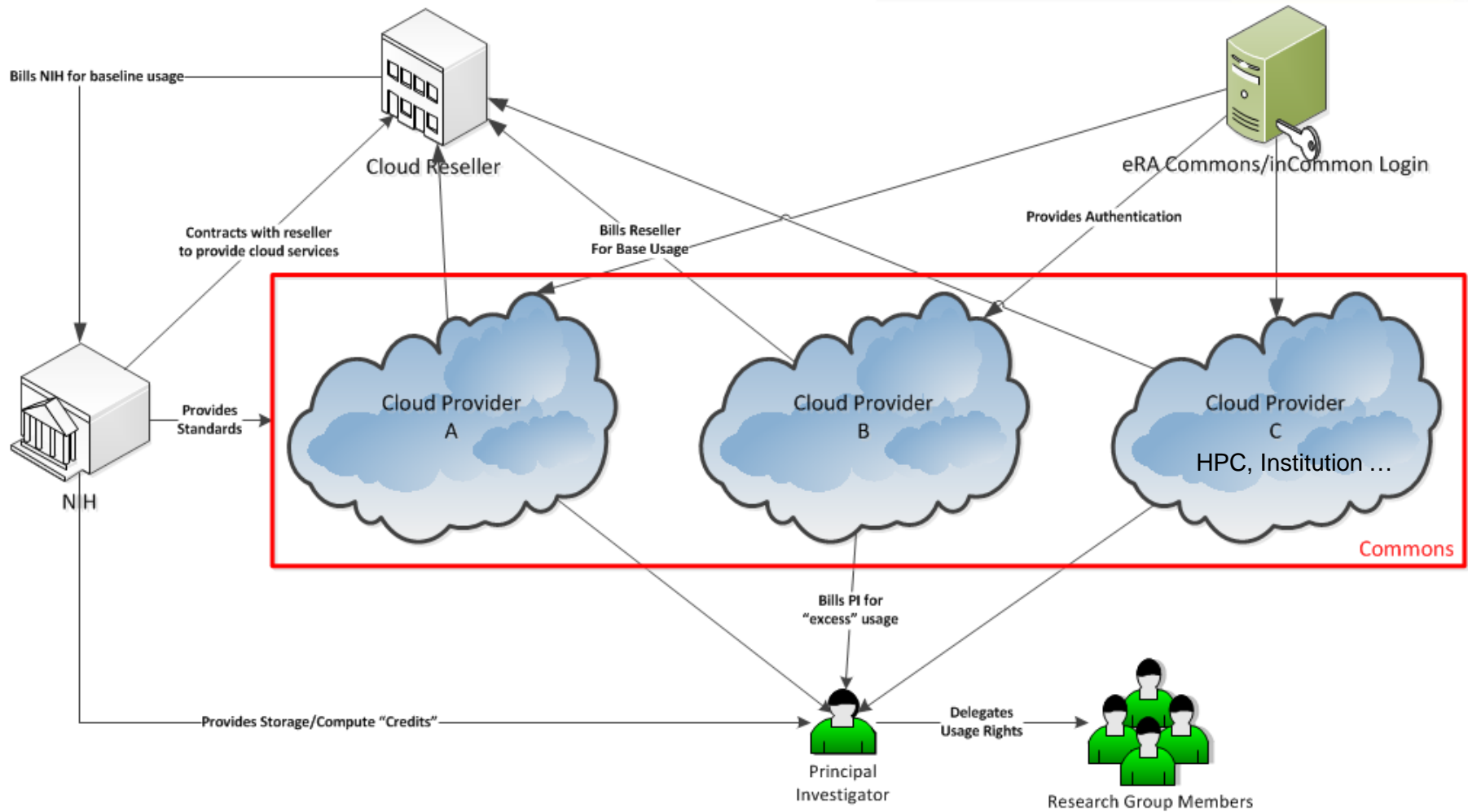
*Cloud, Research Objects,  
Business Models*

# What Does the Commons Enable?

- Dropbox like storage
- The opportunity to apply quality metrics
- Bring compute to the data
- A place to collaborate
- A place to discover

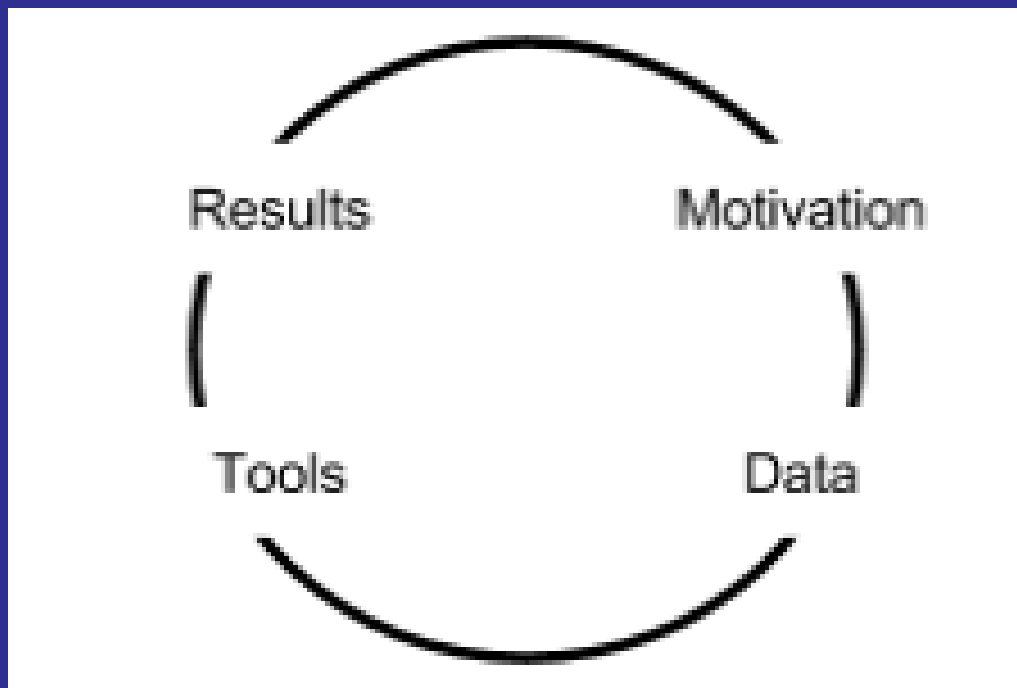


# One Possible Commons Business Model





# Pilots Around A Virtuous Cycle Expect a Funding Call



# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## Programmatic Theme

Sustainability\*

Education\*

Innovation\*

Process

Collaboration

## Deliverable

Commons

Training Center

BD2K

Modified Review

Communication

## Example Features

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists

\* Hires made



# The Biomedical Research Digital Enterprise

# Training (Michelle Dunn)



- Training Goals:
  - Develop a sufficient cadre of researchers skilled in the science of Big Data
  - Elevate general competencies in data usage and analysis across the biomedical research workforce
  - Combat the Google bus
- How:
  - Traditional training grants
  - Work with IC's on a needs assessment
  - Standards for course descriptions with EU
  - Work with institutions on raising awareness
  - Virtual/physical training center(s)?



# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## Programmatic Theme

Sustainability\*

Education\*

Innovation\*

Process

Collaboration

## Deliverable

Commons

Training Center

BD2K

Modified Review

Communication

## Example Features

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists

\* Hires made



# The Biomedical Research Digital Enterprise

# BD2K Innovation FY 14 (Jennie Larkin and Mark Guyer)



- Data Discovery Index Coordination Consortium (10/14)
- Targeted Software Development (under review)
- 11 Investigator-initiated Centers of Excellence for Big Data (10/14)
- BD2K-LINCS-Perturbation Data Coordination and Integration Center (11-12/14)



# BD2K Innovation FY 15 (Jennie Larkin and Mark Guyer)



- Governance model to foster the ecosystem
- Workshops identified, others will be considered
  - Sustainability
  - Standards
  - ELSI for research use of clinical data
  - Private sector engagement for research use of clinical data
  - Using EHRs for outcomes research
  - Gaming community contribution to biomedical research
- Software index?
- Standards framework?
- Other?



# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## Programmatic Theme

Sustainability\*

Education\*

Innovation\*

Process

Collaboration

## Deliverable

Commons

Training Center

BD2K

Modified Review

Communication

## Example Features

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists

\* Hires made



# The Biomedical Research Digital Enterprise

# Process Current Efforts



- Clinical data harmonization
- Data citation
- Machine readable data sharing plans
- New review models, audiences etc.
  - Open review
  - Micro funding
  - Standing data committees to explore best practices
  - Crowd sourcing





# Associate Director for Data Science

Scientific Data Council

External Advisory Board

## Programmatic Theme

Sustainability\*

Education\*

Innovation\*

Process

Collaboration

## Deliverable

Commons

Training Center

BD2K

Modified Review

Communication

## Example Features

- Cloud – Data & Compute
- Search
- Security
- Reproducibility Standards
- App Store

- Coordinate
- Hands-on
- Syllabus
- MOOCs

- Community
- Centers
- Training Grants
- Catalogs
- Standards
- Analysis

- Data Resource Support
- Metrics
- Best Practices
- Evaluation
- Portfolio Analysis

- IC's
- Researchers
- Federal Agencies
- International Partners
- Computer Scientists

\* Hires made



# The Biomedical Research Digital Enterprise

# Collaboration Current Efforts

- Joint public – private partnership workshop with NOAA?
- 2 joint workshops with NSF + Dear Colleague letter
- OSTP – Open Data 2.0
- HIRO's big data meeting
- ELIXIR working groups





# NIH...

*philip.bourne@nih.gov*

# Turning Discovery Into Health

