

Welcome!

A Workshop to Help Chart the Future of Data Science at NIH

Philip E. Bourne Ph.D.
Associate Director for Data Science
philip.bourne@nih.gov

September 3, 2014 NIH Bethesda USA



Goals for Today

- Everybody's viewpoint gets heard and appreciated
- Actions are identified around:
 - Strategic directions
 - Policies
 - Funding initiatives – foundational & use cases
- Have you leave as advocates and supporters
- Identify who and how we can continue to engage with to carry out these actions
- Transparently report to the broader community what transpired here



A Few Logistics

- It is an open space style meeting which also means it is an open microphone - grab it at any time but do state name and affiliation
- Use the microphone for the benefit of those remote
- There are Google docs open for comment:
 - <https://docs.google.com/document/d/1qObNNoLa9Mrk3ULjoI70xxUelgWgcq0DSxwZLuownKA/edit>
 - <https://docs.google.com/document/d/1VduurNhd2-4wA0owxbTIfDT7wOybP0R9fv2EEjJYX34/edit>
- Hashtag is #ADDSUP
- A summary of the meeting will also be posted as an open document for comment



Agenda

- Provide the needed background and context (ADDS team)
- Open discussion about what is being done and what needs to be done (All)
- Identify topics for more in-depth discussion (All)
- Join subgroups to discuss these topics (All)
 - Moderators and scribes will be identified
- Subgroups report back (Moderators)
- Discussion of outcomes and next steps (All)

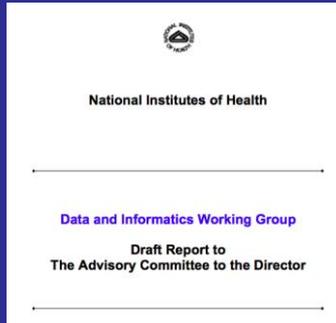


Who We Represent as Stakeholders

- Basic scientists
- Clinicians
- Data scientists/providers
- Funders
- Library and information scientists
- Pharma representatives
- Publishers (open and closed access)
- Cloud providers
- Healthcare industry



Some Background



6/12

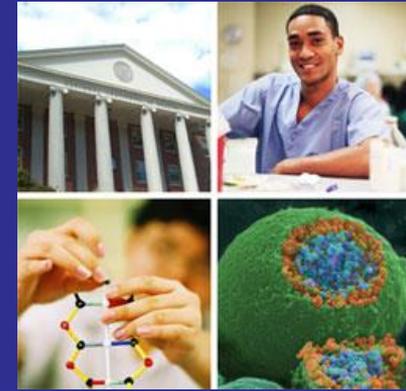
2/14

3/14

- Recommendations:
 - 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



Draft ADDS Team 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*



Current Examples of Collaborators in the Ecosystem

- NIH
 - 27 ICs
- Agencies
 - NSF
 - DOE
 - DARPA
 - NIST
- Government
 - OSTP
 - HHS HDI
 - ONC
 - CDC
 - FDA
- Private sector
 - Phrma
 - Google, Amazon, Microsoft
- Organizations
 - GA4GH
 - PCORI
 - RDA
 - FORCE11
 - CCC
 - CATS
 - FASEB
 - Biophysical Society
 - Sloan Foundation
 - Moore Foundation



Some Goals & *Examples* of the Digital Enterprise

- Sustainability – *The Commons & new business mods*
- 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:
 - \$27M in FY14
 - \$100M 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



Economic



Innovation

F



Collaboration

Deliverable

Exa



NIH

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



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



Possible Topics

- Training & Education
- Sustainability
- Community engagement
- Culture change
- Public private partnerships
- New data science
- Research lifecycle infrastructure
- Metrics and reward systems

