Update from the Associate Director for Data Science

BD2K & ADDS activities

Philip E. Bourne, PhD, FACMI
Associate Director for Data Science

BD2K Multi-Council Working Group
April 25, 2016
Outline

• Scientific Data Council: reorganization, mission update, working groups.
• Data science vs data management at NIH.
• Big Data to Knowledge: update on funding and strategy.
• Commons update.
• Upcoming events.
Scientific Data Council Update

• Reorganized with updated charge:
  – Continued oversight of BD2K.
  – 3 co-chairs: Phil Bourne (Data Science), Carrie Wolinetz (Policy), Jon Lorsch (IC Director).
  – Aligned major, trans-NIH data sharing committees with the SDC, for more coherent and streamlined governance.
  – Sustainability WG established.
Scientific Data Council (SDC) Governance

Scientific Data Council
(Co-chairs: Phil Bourne, Jon Lorsch, and Carrie Wolinetz)

Data Sharing Oversight Subcommittee
(Co-chairs: Belinda Seto and Jim Ostell)

Data Sustainability Subcommittee
(Co-chairs: Allan Darry and Susan Gregurick)

Genomic Data Sharing Working Group
(Co-chairs: Dina Paltoo and Laura Lyman Rodriguez)

Digital Scientific Data Policy and Implementation Working Group
(Co-chairs: Dina Paltoo and JP Kim)

Other Working Groups (e.g., Common Data Elements)

Genomic Data Submission and Management Taskforce
(Co-chairs: TBD)

Digital Scientific Data Technical Advisory Task Force
(Co-Chairs: TBD)
Data Science vs Data Management

• Understanding IC pain points.
• NLM Director to be named shortly – will begin discussions with them on the *when* and *how* of BD2K migration to NLM.
• Renewed NCBI oversight to further integrate with NIH data management needs.
• Anticipate high level data management strategic planning process in conjunction with the SDC.
BD2K Update

• **120** awards, **30** US states & **3** countries.
• **10** different mechanisms (U54, U01, U41, R01, U24, T32, T15, K01, R25, P50).
• Diversity of data science areas.
• Hitting all areas identified by the DIWG report.
• Starting to see outcomes of awards:
  – Early releases of tools/software (Spring/summer 2016) e.g., BioCADDIE.
  – **84** trainees (T15/32), **25** courses (R25).
120 BD2K Awards
30 US states
Creating a Collaborative Data Science Community

• Supporting communities e.g. GA4GH, RDA, FORCE11.
• Innovation Lab with NSF.
• International biomedical funder collaborations:
  – Wellcome Trust & HHMI - open science prize.
  – ELIXIR - Alignment of tools and research related to indexing and standards.
  – European Open Science Cloud - Shared vision of implementing the Commons Framework.
Open Science Prize

96 submissions received

Solvers from 45 countries spanning 5 continents

Over 100 entities represented

Each team has US and international members
Commons Update

• Groundswell of support from the community:
  – European Open Science Cloud adopting 3-tied commons framework.
  – Reference datasets deployed.
  – Workshop with the Common Fund in May.
• Commons credits first issue in May.
• Indexing by the BioCADDIE prototype.
• Standards Coordinating Center Portal in June.
• Some IC’s getting on board.
Looking Forward

• All Hands Meeting, Nov 29-Dec 1, 2016.
• The BD2K Center for Causal Discovery Datathon at Carnegie Mellon occurring June 17-18: https://t.co/LqEk83oN1v.
• Big Data in Translational Genomics released BRCA Exchange https://t.co/AxILNXCpbK.
• BD2K-related hackathons - ongoing.
Data Science at NIH

- https://datascience.nih.gov
- bd2k@nih.gov
- @NIH_BD2K
- #BD2K