

The ADDS Office

The Office of the Associate Director for Data Science (ADDS) sets a strategic vision for and leads data science activities and funding initiatives across the National Institutes for Health (NIH). Trans-NIH extramural funding initiatives in data science fall under the umbrella of the Big Data to Knowledge (BD2K) Initiative, a joint collaboration between the NIH Common Fund and each of the 27 Institutes and Centers. The ADDS Office has four main focus areas: training, sustainability, open science, and communications.

Data Science Activities @ NIH

Pi Day (3.14)

Celebration of the Intersection between the mathematical and biomedical sciences.



Frontiers in Data Science Lecture Series

Lectures and webinars featuring ideas at the forefront of data science designed to inspire innovation and exploration.

Data Science Distinguished Seminar Series

Lectures featuring high-profile biomedical scientists who exemplify the roles that computational and quantitative sciences play in today's most innovative biomedical research and practices.

BD2K Guide to the Fundamentals in Data Science Webinar Series

Weekly virtual talks presented by experts from across the country covering Big Data topics relevant to modern biomedical research.

Data Science Training, Education, and Workforce Development

The ADDS office supports a portfolio of dynamic training classes and workshops in data science, designed with the goal of establishing an effective and diverse biomedical data science workforce.

Contact Us:

BD2K

Jennie Larkin, Ph.D.
Jennie.Larkin@nih.gov

The Commons

Vivien Bonazzi, Ph.D.
Vivien.Bonazzi@nih.gov

Senior Science Advisor

Mark Guyer, Ph.D.
Mark.Guyer@nih.gov

Training

Michelle Dunn, Ph.D.
Michelle.Dunn@nih.gov

Sustainability

Allen Dearry, Ph.D.
Allen.Dearry@nih.gov

Open Innovation & Policy

Elizabeth Kittrie, M.M.
Elizabeth.Kittrie@nih.gov

Communications

Lisa Dunnebacke, M.S.
Lisa.Dunnebacke@nih.gov

For More Information:



Data Science at NIH website:
www.datascience.nih.gov

Input/Output Blog:

News and discussion forum for the biomedical data science community. Visit www.datascience.nih.gov/blog.



Join the BD2K Updates listserv:

Weekly news roundup from across the BD2K research community.

For additional information or to subscribe, email: BD2K@nih.gov.



@NIH_BD2K
#BigData
#DataScience
#NIH_BD2K



@NIHbd2k



NIH Big Data to Knowledge (BD2K)

NIH Big Data to Knowledge Initiative



I believe the future of research into health and well-being is going to be tied very much to our ability to sustain, trust, integrate, analyze/discover, disseminate, visualize, and comprehend digital data.

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

Big Data to Knowledge

BD2K is a trans-NIH initiative established to enable biomedical research as a digital research enterprise, to facilitate discovery and support new knowledge, and to maximize community engagement.

The BD2K initiative addresses four major aims that, in combination, are meant to enhance the utility of biomedical Big Data by:

- ▶ Facilitating broad use of biomedical digital assets by making them **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (**FAIR**).
- ▶ Conducting research and developing the methods, software, and tools needed to analyze biomedical Big Data.
- ▶ Enhancing training in the development and use of methods and tools necessary for biomedical Big Data science.
- ▶ Supporting a data ecosystem that accelerates discovery as part of a digital enterprise.

BD2K: Data Science Research

There is great potential for discovery and innovation as the quantity and accessibility of biomedical data continues to expand. However, this potential can never be realized without appropriate tools. Visit <https://datascience.nih.gov/bd2k/funded-programs>

Targeted Software Development

These awards fund software tools and methods development to tackle data management, transformation, and analysis challenges in areas of high need to the biomedical research community. Topic areas include: Applying Metadata, Data Compression and Reduction, Data Privacy, Data Provenance, Data Repurposing, Data Visualization, and Data Wrangling.

BD2K Centers

These are large-scale projects aiming to develop new approaches, methods, software tools, and related resources. The Centers also provide training to advance Big Data science in the context of biomedical research. The 13 BD2K Centers function with the

other BD2K grantees as a consortium and collaborate with one another for the purpose of furthering every aspect of the field of biomedical data science research. The Centers are located all across the United States and include 11 Centers of Excellence for Big Data Computing and two Centers that are collaborative projects with the NIH Common Fund's Library of Integrated Network-Based Cellular Signatures (LINCS) program.

BD2K: Developing the Workforce

BD2K Training and Diversity programs support the establishment of an effective and diverse biomedical data science workforce. Accomplishing this requires support for training of early career scientists and graduate students, as well as development and dissemination of short courses, training materials, and open educational resources in data science and data management. BD2K also supports specialized programs for K12 students and to enhance diversity among undergraduate students. To build interdisciplinary data science teams, BD2K and the National Science Foundation together support a series of Innovation Labs and the Quantitative Approaches to Biomedical Big Data (QuBBD) program. The BD2K Training Coordination Center is developing ERuDite, an Educational Resource Discovery Index, to enable learners to find resources about data science, whether funded through BD2K or another source.

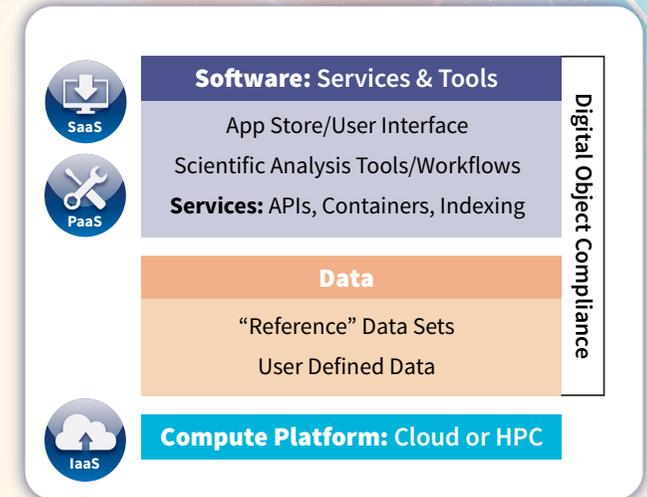


Developing the Data Commons

The Data Commons is a framework that fosters the development of a digital ecosystem—a plug-and-play model that allows multiple participants to connect to it, interact with each other, and create value.

Characteristics of the Data Commons

- ▶ Treats products of research (data, metadata, software, papers, workflows, standards, etc.) as digital objects.
- ▶ These digital objects exist in a shared virtual space where they can be: shared, managed, found, and re-used.
- ▶ Digital Objects can be part of the Data Commons if they are **FAIR**.



Implementing the Data Commons

- ▶ **Reference Data Sets:** Making high-value data sets and associated tools available in the cloud.
- ▶ **Commons Framework Pilot:** Facilitating connectivity, interoperability, and access to digital objects.
- ▶ **Cloud Credit Model:** Facilitating access to cloud resources for NIH-funded investigators.
- ▶ **Resource Indexing:** Developing and implementing solutions to help find and use digital objects made available in the Data Commons.

Fostering Open Science

Open Science and Open Data are critical links in maximizing the potential of biomedical data and increasing the rate of discovery. ADDS encourages researchers to follow the **FAIR** principles.

The Open Science Prize

The Open Science Prize is a collaboration between the NIH, Wellcome Trust, and Howard Hughes Medical Institute that challenges innovators to use Open Data in novel ways to improve human health. We received 96 entries from teams representing 45 countries across 6 continents, demonstrating the global interest in Open Science. Visit <https://www.openscienceprize.org/>

