

# DATA SCIENCE DIGEST:

*A Newsletter from the ADDS*

Summer 2016

## In this issue:

- Call for Public Feedback for DataMed DDI Prototype
- From the Desk of the ADDS: Preprints
- Innovation Lab 2016: Mobile Health

## Quick Links

- [Data Science Community News](#)
- [Data Science Events](#)
- [Data Science Workshops](#)
- [Data Science Opportunities](#)

## About ADDS News:

Published by the Office of the Associate Director for Data Science. News, article ideas, calendar events, letters, and photographs are welcome. Submissions may be edited.

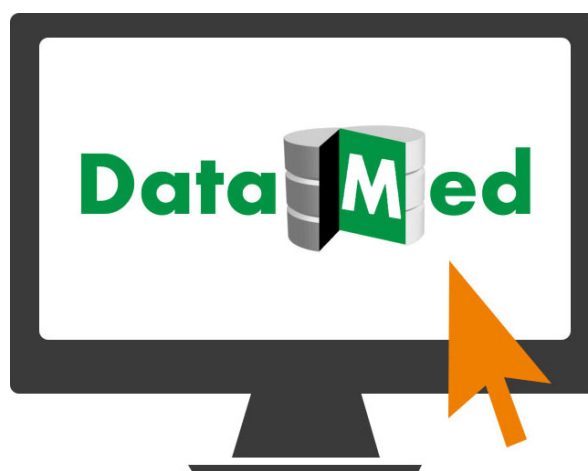
National Institutes of Health  
Building 1, 1 Center Drive  
Bethesda, MD 20892  
Telephone: 301-402-9819  
Email: [adds@nih.gov](mailto:adds@nih.gov)

Find us on 

 Follow @NIH\_BD2K

Subscribe to BD2K LISTSERV

## Call for Public Feedback for DataMed DDI Prototype



The National Institutes for Health Big Data to Knowledge (BD2K) initiative is pleased to announce an opportunity to evaluate and provide feedback on a prototype biomedical Data Discovery Index (DDI) called [DataMed](#). Developed through the biomedical and healthCARE Data Discovery Indexing Ecosystem project ([bioCADDIE](#)), the prototype allows users to find and access biomedical datasets from multiple sources based on key attributes. DataMed is one of several ongoing efforts to develop common practices for indexing biomedical Big Data. This opportunity to shape the prototype in a direction that will benefit all intended users is an exciting point in the evolution of the DataMed index and an essential ingredient is participation by a broad group of users.

DataMed is a work in progress and the DataMed development team welcomes your feedback here: <https://datamed.org/feedback.php> to increase their understanding of how the website is perceived and experienced by visitors like you. The information gathered will help the DataMed development team assess and respond to the community and understand user feedback and suggestions to help establish optimal approaches to Finding and Accessing biomedical data. User feedback will guide further DataMed development and BD2K insight into what works and what doesn't and contribute to further advances in biomedical research and patient care.

BioCADDIE is led by the University of California San Diego and brings together investigators from the University of Michigan, University of Texas-Houston, and Oxford University as well as a wider community of researchers, developers, and service providers via a number of Working Groups to embed DataMed into existing efforts and ensure a community-driven approach to its development. The ADDS Office would like to acknowledge the bioCADDIE team and the contributions of the NIH Core Team without which this public release would not have been possible: Dr. Ronald Margolis (NIDDK), Lead; Dr. Dawei Lin (NIAID) and Dr. Alison Yao (NIAID), co-Program Officers; Dr. Ian Fore (NCI), Scientific Officer; Dr. Claire Schulkey (NIAID) and Eric Choi (NIAID), Project Management.

The bioCADDIE Project held a Repositories Workshop “Data Indexing for DataMed” at the Estancia in La Jolla, CA on June 23, 2016. At this workshop, presenters explained the DataMed metadata model, named DatA Tag Suite ([DATS](#)), which also has a [schema.org](#)-annotated serialization. Participants were given hands-on experience mapping their metadata into the DATS model for indexing. Watch the [YouTube video](#).

The NIH Big Data to Knowledge (BD2K) initiative seeks to address the needs of the biomedical research community as it confronts the emerging challenges of data. The overall principles are the need to Find, Access, Interoperate, and Re-use (FAIR) data to catalyze greater advances in biomedical science and patient care. In order to make data findable and accessible, it must be indexed and made searchable. BioCADDIE is a NIH BD2K initiated community-based project. Its mission is to fulfill the mandate to establish a Data Discovery Index (DDI) to enhance the discoverability, citation, and access of biomedical data in accordance with the FAIR principles. For additional information, contact [Anupama Gururaj](#) or [bicaddie@ucsd.edu](mailto:bicaddie@ucsd.edu).

---

## From the Desk of the ADDS: Preprints

The Office of the Associate Director for Data Science (ADDS) has as part of its mission to accelerate the rate of scientific discovery through open science. It is not surprising, therefore, that we have taken significant interest in the growing momentum behind preprints in the life sciences. A preprint is a copy of a completed manuscript that is made available to the community prior to publication, typically in a peer reviewed journal. This implies that research in an unreviewed form is available to the community often many months before it would be otherwise. It is a feature of scholarly communication, which is well established in Physics and related fields through the [arXiv](#) preprint server. The pros and cons of preprints in the Life Sciences were discussed at the first meeting of [ASAPbio](#) held February 16-17, 2016.

Enthusiasm by the varied stakeholders at the meeting for pursuing preprints was high enough that a group of funders, including the NIH, got together at a [Funder's Meeting](#) on May 24, 2016 to see what collective action we might take. The outcome, in summary, is:

- Funders will develop, as far as possible, homogeneous policies for how we regard preprints as part of the scholarly record.
- ASAPbio will produce a proposal describing the governance, infrastructure, and standards to be used for a preprint service.
- Funders will disseminate these outcomes.

This article addresses the last bullet.

On a personal note, I am very excited that the community is considering preprints, since I see it as a first step in changing an antiquated system for disseminating knowledge, which while well suited to the analog era and slower pace of scientific advancement, is slowing us down in the fast-paced digital era. When lives are at stake, this is a welcome development.

Philip E. Bourne, Ph.D., FACMI

---

## Innovation Lab 2016: Mobile Health

A mathematician, a computer scientist, and a clinician sit down at a table ... this sounds like the opening of a joke, but it's not. It describes what happened at the 2016 Innovation Lab on Interdisciplinary Approaches to Biomedical Data Science Challenges of Wearable and/or Ambient Sensors held at the UCLA Lake Arrowhead Conference Center on June 15-19, 2016.



The Innovation Lab brought together 29 delegates from a variety of backgrounds interested in the theme of mobile health and sensors along with 6 mentors who are leading researchers in their respective fields. The goal is to develop new teams to address important biomedical problems through innovative data science approaches and methods. The delegates went through a series of activities designed to encourage creativity and innovation under the guidance of professional facilitators. First, they learned what knowledge was in the room and what the current state of the art is. Then, they went through an interactive process that drew out ideas for what the main data science challenges are and how they might be addressed. The week culminated in the delegates forming multidisciplinary teams and sketching out a research program.

One team is exploring building predictive models to forecast pain and integrating them into wearable devices for patient decision support. Another team is integrating mHealth data from wearables to explore how patient-controlled variables, such as sleep and physical activity, influence Crohn's disease and ulcerative colitis symptoms. The delegates were then guided through the development of their research programs by six highly regarded mentors: Todd Coleman (UCSD), John Harer (Duke), Kathy Kim (UC Davis), Xihong Lin (Harvard), Jack Van Horn (USC), and Yazhen Zhang (UW Madison). These mentors helped the teams develop and refine innovative solutions to biomedical problems that require data science.

This Innovation Lab was organized by the BD2K Training Coordination Center (TCC), along with facilitators from KnowInnovation, and was funded as part of the Quantitative Biomedical Big Data (QuBBD) initiative, a joint effort between the Division of Mathematical Sciences of the National Science Foundation and the NIH Big Data to Knowledge (BD2K) Initiative.

---