



**NIH BD2K Joint Kick-Off Meeting
Data Discovery Index Coordination Consortium (DDICC)
and BD2K Center Consortium**

November 2-4, 2014

Goals: For the DDICC and BD2K Center investigators to discuss the goals of their consortia and how to collaborate with each other, other BD2K projects, and the NIH Commons.

Meeting Overview

Day 1, Sunday, November 2, 6:30 pm - 9:00 pm

Hilton Washington DC/Rockville Hotel and Executive Meeting Center
1750 Rockville Pike, Rockville MD 20852

- 6:30p-7:00p Dinner, Twinbrook Ballroom
- 7:00p-9:00p Joint DDICC and Center Consortium Session, Twinbrook Ballroom

Day 2, Monday, November 3, 8:00 am - 7:30 pm

NIAID Conference Center, 5601 Fishers Lane, Rockville MD 20852
and dinner at the Hilton Washington DC/Rockville Hotel and Executive Meeting Center,
1750 Rockville Pike, Rockville MD 20852

- 8:00a-9:00a Registration and Meeting Check-In
- 9:00a-9:45a Joint DDICC and Center Consortium Session, Room 1D13
- 9:45a-10:00a Break
- 10:00a-12:30p Parallel Sessions (Part 1)
- 12:30p-1:30p Lunch, Cafeteria
- 1:30p-2:00p Joint DDICC and Center Consortium Session, Room 1D13
- 2:00p-5:45p Parallel Sessions (Part 2)
- 6:00p-7:30p Dinner, Hilton Rockville

Day 3, Tuesday, November 4, 7:00 am - 2:30 pm

NIAID Conference Center, 5601 Fishers Lane, Rockville MD 20852

- 7:00a-8:00a Meeting Check-in
- 8:00a-9:20a Joint DDICC and Center Consortium Session, Room 1D13
- 9:20a-9:30a Break
- 9:30a-2:30p Parallel Sessions (Part 3)

Day 1, Sunday, November 2

6:30p-9:00p **Joint DDICC and Center Consortium Session**
Hilton Rockville, Twinbrook Ballroom

6:30p-7:00p [Dinner](#)

7:00p-9:00p [Discussion of Data Science at NIH](#)
Phil Bourne, NIH/OD

This session will set the stage for the meeting by describing the role of BD2K in the larger data science ecosystem at NIH and beyond as well as defining the goals and expectation for the BD2K program. Attendees will be asked to address the issues raised in their subsequent presentations and discussions.



Day 2, Monday, November 3

- 8:00a-9:00a Registration and Meeting Check-In**
- 9:00a-9:45a Joint DDICC and Center Consortium Session
NIAID Conference Center Room 1D13
Moderator: Ron Margolis, NIH/NIDDK**
- 9:00a-9:15a [Welcome and Goals, Brief Overview of BD2K](#)
Phil Bourne, NIH/OD
- 9:15a-9:25a [Introduction and Objectives](#)
Mark Guyer, NIH/NHGRI
--The types of activities that consortia do
--The history that led to these consortia (HGP, NCBCs, DIWG report)
- 9:25a-9:35a [DDICC Overview](#)
Lucila Ohno-Machado, University of California, San Diego
- 9:35a-9:45a [BD2K Center Consortium Overview](#)
Mark Guyer, NIH/NHGRI
- 9:45a-10:00a Break
- 10:00a-12:30p Parallel Sessions (Part 1)**

The Data Discovery Index (Part 1)	BD2K Center Consortium (Part 1)
NIAID Conference Center Room 1D06	NIAID Conference Center Room 1D13
Moderator: Dawei Lin, NIH/NIAID	Moderator: Jennie Larkin, NIH/OD
10:00a-10:15a Development of the Data Discovery Index Concept Ron Margolis, NIH/NIDDK	10:00a-12:30p Overview of Each BD2K Center <i>Scientific goals, the major challenges to achieving these goals, their contributions to the Consortium and the digital ecosystem that can be used by the other Centers and more broadly, what they hope that other Centers will contribute to make them successful, and their training initiatives.</i>
10:15a-10:25a Implementation of the bioCADDIE Plan Lucila Ohno-Machado, University of California, San Diego	
10:25a-11:15a Task Forces and Pilots	
	10:00a-10:10a <i>Big Data for Discovery Science</i> Arthur Toga, University of Southern California



<p>Data Shop</p> <p>10:25a-10:35a <i>Pilot project 1.1: Linking Publications and Underlying Data Sets Using Natural Language Processing</i> Hua Xu, University of Texas, Houston</p>	<p>10:10a-10:20a <i>Center for Expanded Data Annotation and Retrieval (CEDAR)</i> Mark Musen, Stanford University</p>
<p>10:35a-10:45a <i>Pilot project 1.2: Rating Relationships in Imaging Reports: Machine Versus Crowdsourcing Approaches</i> Ricky Taira, University of California, Los Angeles</p>	<p>10:20a-10:30a <i>Center for Big Data in Translational Genomics</i> David Haussler, UC Santa Cruz</p>
<p>Data Showcase</p> <p>10:45a-10:55a <i>Pilot project 2.1: Data Recommendation using Machine Learning and Crowdsourcing</i> Xiaoqian Jiang, University of California, San Diego</p>	<p>10:30a-10:45a Break</p> <p>10:45a-10:55a <i>A Community Effort to Translate Protein Data to Knowledge: An Integrated Platform</i> Peipei Ping, UCLA</p>
<p>10:55a-11:05a <i>Pilot project 2.2: Intelligent Search Expansion (iSee) and Visualization of Datasets</i> Todd Johnson, University of Texas, Houston</p>	<p>10:55a-11:05a <i>KnowEng, a Scalable Knowledge Engine for Large-Scale Genomic Data</i> Jiawei Han, University of Illinois Urbana-Champaign</p>
<p>Data Market</p> <p>11:05a-11:15a <i>Pilot project 3.2: Citation and Data Access Metrics for a Molecular Data Bank and Related Resources</i> Peter Rose, University of California, San Diego Chun-Nan Hsu, University of California, San Diego</p>	<p>11:05a-11:15a <i>ENIGMA Center for Worldwide Medicine, Imaging, and Genomics</i> Paul Thompson, University of Southern California</p>
<p>11:15a-12:30p Discussion: How to Achieve and Maintain Integration and Synergy Across the Consortium?</p>	<p>11:15a-11:25a <i>Patient-Centered Information Commons</i> Isaac Kohane, Harvard University Medical School</p> <p>11:25a-11:35a <i>Center for Predictive Computational Phenotyping</i> Mark Craven, University of Wisconsin - Madison</p> <p>11:35a-11:50a Break</p> <p>11:50a-12:00 <i>Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data</i> Greg Cooper, University of Pittsburgh</p> <p>12:00 -12:10p <i>Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K)</i> Santosh Kumar, University of Memphis</p> <p>12:10p-12:20p <i>National Center for Mobility Data Integration to Insight (The Mobilize Center)</i> Scott Delp, Stanford University</p>



	12:20p-12:30p <i>LINCS-BD2K Perturbation Data Coordination and Integration Center</i> Avi Ma'ayan, Icahn School of Medicine at Mount Sinai
--	---

12:30p-1:30p Lunch, Cafeteria

**1:30p-2:00pa Joint DDICC and Center Consortium Session
NIAID Conference Center Room 1D13**

[DDICC and BD2K Center Consortium Shared Activities](#)

Ron Margolis, NIH/NIDDK
 --Points of coordination
 --Shared policies

2:00p-5:45p Parallel Sessions (Part 2)

The Data Discovery Index (Part 2)	BD2K Center Consortium (Part 2)
NIAID Conference Center Room 1D06	NIAID Conference Center Room 1D13
Moderator: Ian Fore, NIH/NCI	Moderator: Phil Bourne, NIH/OD
2:00p-3:30p 8 Supplement Projects: Relevance to the DDI and Plan of Action	Discussion of Collaborative Possibilities and Contributions to the Ecosystem
2:00p-2:10p <i>Data Discovery Index for Dynamic Biomedical Data</i> Rai Winslow, Johns Hopkins University	2:00p-2:30p The Commons, the Cloud, and Interoperability
2:10p-2:20p <i>Using Biomedical Knowledge to Identify Plausible Signals for Pharmacovigilance</i> Trevor Cohen, University of Texas, Houston	2:00p-2:05p <i>An initial Commons Framework: Sharing, Collaboration, and Interoperability</i> Vivien Bonazzi, NIH/OD George Komatsoulis, NIH/NCBI
2:20p-2:30p <i>Discovering and Applying Knowledge in Clinical Databases</i> George Hripcsak, Columbia University Medical Center	2:05p-2:15p <i>An Example of Interoperability from the Global Alliance for Genomics and Health</i> David Haussler, UC Santa Cruz Angel Pizarro, Amazon Web Services (AWS) Jonathan Bingham, Google Bob Davidson, Microsoft
2:30p-2:40p Break	2:15p-2:30p Discussion
2:40p-2:50p <i>Generation of a Centralized and Integrated Resource for Exposure Data</i> Carolyn Mattingly, NC State University	



2:50p-3:00p	<i>Natural Language Processing for Clinical and Translational Research</i> Hongfang Liu, Mayo Clinic	2:30p-3:00p	Identification of Collaborative Topics
3:00p-3:10p	<i>Journal-database Linkage: A Model for Effective Leverage of Biomedical Datasets</i> Lauren Becnel, Baylor College of Medicine Neil McKenna, Baylor College of Medicine	3:00p-3:15p	Break
3:10p-3:20p	<i>A Web-based Database and Toolkit for Effective Neuroimaging</i> Tor Wager, University of Colorado, Boulder	3:15p-4:15p	Break-out Groups Break-out Group 1, Room 1D13 Lisa Brooks, NIH/NHGRI Break-out Group 2, Room LD20A Susan Gregurick, NIH/NIGMS Break-out Group 3, Room LD20B Grace Peng, NIH/NIBIB
3:20p-3:30p	<i>Access, Discoverability and Citation for Repertoire Sequencing (REP-SEQ) Data</i> Steven Kleinstejn, Yale University	4:15p-5:30p	Report Back and Discussion
3:30p-4:00p	Discussion: Integration of Supplement Projects with the DDICC		
4:00p-4:15p	Break		
4:15p-5:45p	Related BD2K Efforts at the NIH and Beyond		
4:15p-4:30p	<i>National Library of Medicine</i> Betsy Humphreys, NIH/NLM		
4:30p-4:45p	<i>National Database for Autism Research</i> Greg Farber, NIH/NIMH		
4:45p-5:00p	<i>National Institute of Standards and Technology</i> Wo Chang, NIST		
5:00p-5:15p	<i>BD2K Software Discovery Index</i> Vivien Bonazzi, NIH/OD		
5:15p-5:45p	Discussion: What are the Possible Points of Intersection with the DDICC?		

6:00p-7:30p Dinner, Hilton Rockville



Day 3, Tuesday, November 4

- 7:00a-8:00a Meeting Check-In**
- 8:00a-9:20a Joint DDICC and Center Consortium Session
NIAID Conference Center Room 1D13**
- 8:00a-8:20a [Diversity and BD2K Activities](#)
Hannah Valantine, NIH/OD
- 8:20a-9:20a [Opportunities for Joint Activities](#)
Phil Bourne, NIH/OD
Jennie Larkin, NIH/OD
Vivien Bonazzi, NIH/OD
 --Standards Development
 -- Software Development
 -- Shared Infrastructure
 --Other?
- 9:20a-9:30a Break**
- 9:30a-2:30p Parallel Sessions (Part 3)**

The Data Discovery Index (Part 3)	BD2K Center Consortium (Part 3)
NIAID Conference Center Room 1D06	NIAID Conference Center Room 1D13
Moderator: Alison Yao, NIH/NIAID	Moderator: Phil Bourne, NIH/OD
9:30a-10:30a Community Engagement Maryann Martone, University of California, San Diego Lucila Ohno-Machado, University of California, San Diego --Workshops --Web portal --Other venues for outreach	9:30a-10:30a Break-out Groups: Identification of Collaborative Activities Break-out Group 4, Room 1D13 Maria Giovanni, NIH/NIAID Break-out Group 5, Room LD10 Vinay Pai, NIH/NIBIB Break-out Group 6, Room LD40 Mary Rodgers, NIH/NIBIB
10:30a-11:00a Discussion: Integration of the DDICC, the Supplements and Future Pilots	10:30a-10:50a Break
11:00a-11:30a Challenges and Opportunities Lucila Ohno-Machado, University of California, San Diego	10:50a-12:00 Report Back and Discussion
	12:00-1:00p Lunch (cafeteria)

<p>11:30a-12:30p DDICC Steering Committee Meeting Closed Session, Room LD40 Lucila Ohno-Machado, University of California, San Diego</p> <ul style="list-style-type: none"><i>--Governance plan</i><ul style="list-style-type: none"><i>--Role of the NIH</i><i>--Roles of the investigators</i><i>--Integration of BD2K DDI Supplement projects</i><i>--Ideas for future supplement projects/use cases</i><i>--Schedule of meetings</i> <p>12:30a-12:45p Summary and Wrap-up Ron Margolis, NIH/NIDDK</p>	<p>1:00p-1:45p Discussion: Center Consortium Governance Model</p> <p>1:45p-2:15p How Each Center Will Collaborate with the Other Centers and Address the Goals of the NIH Data Science Program <i>2 minutes each Center</i></p> <p>2:15p-2:30p Final Discussion</p>
---	---