

**NIH BD2K Data Catalog Workshop  
August 21-22, 2013  
NCI Shady Grove, Room TE406  
9609 Medical Center Drive  
Rockville, MD**

**Program**

**Workshop Goal:** To identify the least duplicative and burdensome, and most sustainable and scalable method to create and maintain a NIH Data Catalog.

**Day 1 (8:30 – 5:15)**

- 8:30 – 8:35 Introduction and Welcome: **Jennie Larkin, NHLBI and Ron Margolis, NIDDK**
- Social media for the workshop: **Vivien Bonazzi, NHGRI and Chris Wellington, NHGRI**
- Workshop focus: **Ron Margolis, NIDDK**
- 8:35 – 9:05 BD2K Program Goals: **Eric Green, Acting Associate Director for Data Sciences, NIH**
- 9:05 – 9:15 Framework for a Data Catalog: **Fran Berman, RPI**
- 9:15 – 10:00 **Plenary 1:** “Data Catalog implementation and lessons learned from NOAA’s experience” **Ken Casey, NOAA**
- 10:00 – 10:20 **BREAK**
- 10:20 – 11:30 **Panel 1: Community-specific Biomedical Data Catalogs**
- 3 presentations 10 minutes each, addressing questions (below), followed by a 40 minute panel discussion.
- Session Lead: George Alter, ICPSR**
- Neuroscience Information Framework: **Jeff Grethe, UCSD and NIF**
- Introduction and Invitation to MOPED and DELSA: **Eugene Kolker, Seattle Children’s Hospital**
- BioSiteMaps: **Charles Borromeo, University of Pittsburgh**



1. What community do you serve?
2. What services do you provide?
3. How do you help your community to discover data?
4. How do you encourage citation of data?
5. How do you measure your success?

**Questions for Panel Discussion:**

1. How could an NIH Data Catalog interoperate with your existing services?
2. How could an NIH Data Catalog accelerate innovation in your research community?
3. What are the main obstacles to the success of an NIH Data Catalog in your research community?

11:30 – 12:30

**LUNCH**

12:30 – 2:00

**Implementing the White House OSTP Mandate**

12:30 – 12:45

Introduction to the OSTP mandate: **Mike Stebbins, OSTP**

12:45 – 2:00

**Panel 2: NIH and Federal Strategies for Implementing OSTP Policies**

1 slide (5 minutes each), followed by discussion and questions.

**Session Lead: Mike Stebbins, OSTP**

Perspective from the National Library of Medicine: **Jerry Sheehan, NLM/NIH**

Incorporating Grant System Information into a NIH Data Catalog: **Rick Ikeda, OD/NIH**

Perspective from the NIH Center for Scientific Review: **Valerie Durrant, CSR/NIH**

Perspective from the NIH Office of Science Policy: **Sarah Carr, OD/NIH**

Perspective from Health and Human Services: **Greg Downing, HHS**

Perspective from National Oceanic and Atmospheric Administration: **Ken Casey, NOAA**

Perspective from National Institute of Standards and Technology: **John Henry Scott, NIST**



**Questions for Panel Discussion:**

1. What does the OSTP mandate mean to federal agencies? What does it mean to NIH?
2. What are the plans for NIH to partner with other agencies to develop data catalogs?
3. What impact will the NIH catalog have on what is requested and expected from the NIH research community?
4. What are internal IT and other constraints in developing the NIH data catalog?
5. How does NIH see the data catalog as responding to the public access memo? What if data is listed in the catalog but not available?

2:00 – 3:30

**Breakout Session 1: Community-wide Guidance for Implementation of a NIH Data Catalog**

The goal of the breakouts will be to discuss the best way(s) to implement a NIH data catalog.

*Each group will be tasked with addressing assigned questions, and may then address other questions as time permits.*

**Questions for Breakout groups to address:**

1. What would a really useful NIH data catalog look like to you?
  - a. Are there things that you could do with the catalog that you cannot easily do now?
  - b. How should the NIH data catalog combine with other data resources?
2. What should be in the NIH data catalog?
  - a. How should the contents of the NIH data catalog be structured?
  - b. How should the contents of the NIH data catalog be discovered?
3. What should the NIH data catalog link to?
4. How should the contents of the NIH data catalog be cited?
5. How can the data catalog be made accessible to the research community? The general public?
6. What are appropriate metrics of use and success for the data catalog?

3:30 – 3:45

**BREAK**





3:45 – 5:15                    **Report out from Breakout Groups**  
5 minutes report out, 10 minutes discussion for each of 5-6 groups

6:30                              **Dinner on own**

## **Day 2 (8:30 – 1:00)**

8:30 – 8:35                    Introduction to Day 2

8:35 – 9:00                    Recap of Day 1  
**Fran Berman, RPI**

9:00 – 9:45                    **Plenary 2:** “Understanding the data behind web pages” **Ramanathan Guha, Google**

9:45 – 11:00                   **Breakout Session 2: Potential Use Cases for a NIH Data Catalog – Roadblocks and Solutions**

The goal of the breakouts will be to discuss potential roadblocks and new use cases for a NIH data catalog.

### **Questions for Breakout groups to address:**

1. What roadblocks could stand in the way of the creation of a successful NIH data catalog?
2. What use cases can you imagine for the NIH data catalog 3, 5, and 10 years from now?
3. How might the data catalog be developed /evolved to potentially support future new uses?

11:00 – 11:15                   **Breakout Group Leads put together Slides and BREAK**

11:15 – 12:45                   **Report out from Breakout Groups**  
(5 minutes for reports, 5 minutes discussion for each group)

12:45 – 1:00                    Final Summation: **Fran Berman, RPI**

1:00                              Adjourn

1:00 – 4:00                    **Speakers and session leads draft report.**  
The summary will appear on the BD2K web site.