

## **Workshop on Community-Based Data and Metadata Standards Development: *Best practices to support healthy development and maximize impact***

**Location: John Edward Porter Neurosciences Research Center GE620/630**

***February 25-26, 2015***

***Chairs: Melissa Haendel, Ph.D. and Christopher Chute, M.D., Dr. P.H.***

---

### **MEETING GOALS**

- Identify specific gaps in the development of data and metadata standards with greatest potential impact for NIH relevant research
- Recognize common technical, social and financial pain points in data and metadata standards development life cycle
- Explore possible NIH assistance mechanisms that could address pain points and accelerate and improve the quality of community-based data standards development
- Identify best practices for managing community based data standards development and evaluating near and longer term success

### **AGENDA**

---

#### **Day 1**

**8:30—8:45 a.m.**

***Welcome and Introductions***

**Philip E. Bourne, Ph.D.**

Associate Director for Data Science, NIH

*Overview of the larger BD2K initiative, meeting goals, and expected products*

**8:45—9:25 a.m.**

***Keynote***

**Peter Fox, Ph.D.**

President, Earth Science Information Partners Federation (ESIP)

*The Earth Science Information Partners Federation (ESIP) -- an open, networked community and its impact on innovation and discovery in earth sciences.*

**Session 1**

**9:25—10:40 a.m. Community-driven processes for developing data and metadata standards in the basic sciences and the potential impact(s) for advancing NIH-relevant research.**

*A glimpse of community standards development across the basic science spectrum.*

**Moderator/Chair:** **Melissa Haendel, Ph.D.**  
Oregon Health and Science University  
**Chris Mungall, Ph.D.**  
Lawrence Berkeley Laboratory  
**Chris Stoeckert, Ph.D.**  
University of Pennsylvania  
**Gary D. Bader, Ph.D.**  
University of Toronto  
**Ryan Brinkman, Ph.D.**  
University of British Columbia

**Panel Discussion**

- What are the common social, technical, and financial issues?
- In what similar or different ways have the issues been addressed?

**10:40—10:50 a.m. Break**

**Session 2**

**10:50—11:50 a.m. Challenges of the user community  
Panel Discussion**

**Moderator/Chair:** **Alexa McCray, Ph.D.**  
Harvard Medical School  
**Jessica Tenenbaum, Ph.D.**  
Duke Translational Medicine Institute  
**Olivier Bodenreider, M.D., Ph.D.**  
National Library of Medicine  
**Anita de Waard, Ph.D.**  
Elsevier  
**Lynn Etheredge, Ph.D.**  
Rapid Learning Project

The panel will address the following questions:

- What is a ‘community’ for the purposes of standards development, and how do communities drive data standards efforts?
- What are the different kinds of stakeholders, what are their roles in standards development, and how does this depend on the specific standards under development?
- How do you identify and reward change agents who can lead and sustain standards efforts?

**11:50—12:50 p.m. Working Lunch – generate list of questions for later breakout groups**

**12:50—1:30 p.m.**

*Keynote*

**Doug Fridsma, M.D., Ph.D.**

President and CEO, American Medical Informatics Association

*Accelerating development of national data standards and the potential impact of standards work for human health research.*

**Session 3**

**1:30—2:45 p.m.**

**Community-driven processes for developing data and metadata standards in clinical sciences and the potential impact(s) for advancing NIH-relevant research.**

*A glimpse of community standards development across the clinical spectrum.*

**Moderator/Chair:**

**Chris Chute, M.D., Dr. P.H.**

The Johns Hopkins University

**Clement J. McDonald, M.D.**

National Library of Medicine

**Jessica Tenenbaum, Ph.D.**

Duke Translational Medicine Institute

**Doug Fridsma, M.D., Ph.D.**

American Medical Informatics Association

**Tom Oniki, Ph.D.**

Homer Warner Center for Informatics Research

**Panel Discussion**

- What are the common social, technical, and financial issues?
- In what similar or different ways have the issues been addressed?

**2:45—3:00 p.m.**

**Break**

**Session 4**

**3:00—4:00 p.m.**

**Complexity and Data Diversity**

**Panel Discussion**

**Moderator/Chair:** **James Taylor, Ph.D.**  
The Johns Hopkins University  
**Philippe Rocca-Serra, Ph.D.**  
University of Oxford  
**Michel Dumontier, Ph.D.**  
Stanford Center for Biomedical Informatics Research  
**Charles Bailey, M.D., Ph.D.**  
University of Pennsylvania  
**Olivier Bodenreider, M.D., Ph.D.**  
National Library of Medicine

The panel discussion will address the following questions:

- Why do we need data standards? How will we know one is useful *a priori*?
- How do you know that you need a standard? Determining requirements.
- How do you know a standard is working well? How can you evaluate them?
- How to determine what standard will work best for any given application?

**Session 5**

**4:00—4:30 p.m.      Synthesis Breakouts**

*Three breakout sessions will identify key points and messages from the earlier discussions and develop and refine breakout topics for Day 2.*

**Facilitators:** **Philippe Rocca-Serra, Ph.D.**  
University of Oxford  
**Alexa McCray, Ph.D.**  
Harvard Medical School  
**Weida Tong, Ph.D.**  
U.S. Food and Drug Administration

**Rapporteurs:** **Lesley Skalla, Ph.D.**  
MDB, Inc.  
**Sheri D. Schully, Ph.D.**  
National Cancer Institute  
**Sherri De Coronado**  
National Cancer Institute

**4:30—5:00 p.m.      Summary of the Day**

## Day 2

**8:30—8:35 a.m. Introduction to Day 2 Activities**

### Session 6

**8:35—9:35 a.m. Tools and Rules: What are the technical hurdles that must be overcome?  
Panel Discussion**

**Moderator/Chair:** **Eric Neumann, Ph.D.**  
Foundation Medicine  
**James Overton, Ph.D.**  
Knocean  
**Tom Oniki, Ph.D.**  
Homer Warner Center for Informatics Research  
**Anand Basu, M.B.B.S., M.D.**  
Mayo Clinic  
**Mike Feolo, Ph.D.**  
National Center for Biotechnology Information  
**Ryan Brinkman, Ph.D.**  
University of British Columbia

The panel will address the following:

- What are the tools used in standards development, and where do they fail?
- Who is developing these tools? How are they being evaluated?
- Commonalities, differences, and solutions gleaned from the earlier presentations.
- Identification of those with the biggest adverse impact; discuss possible solutions.
- Creating a toolbox of resources for those engaged in data standards development (e.g., best practices, advisory support).
- Creating a community portal for collaborative development and extensions of standards (e.g., GitHub).

**9:35—9:45 a.m. Break**

### Session 7

**9:45—10:45 a.m. Breakout Groups:** Gaps in data and metadata standards development processes. Areas of greatest need and impact for data and metadata standards development.

**Facilitators:** **Patrice (Polly) Tremoulet, Ph.D.**  
Drexel University  
**Todd Carpenter, M.S.**  
National Information Standards Organization

**Ryan Brinkman, Ph.D.**  
University of British Columbia

**Rapporteurs:** **Veerasamy (Ravi) Ravichandran, Ph.D.**  
National Institute of General Medical Sciences

**Richard Conroy, Ph.D.**  
National Institute of Biomedical Imaging and Bioengineering

**Lesley Skalla, Ph.D.**  
MDB, Inc.

Questions to be addressed include:

- What constitutes due diligence for assessing the existing data standards landscape?
- What factors should be considered in prioritizing needs for development of data and metadata standards?
- How do you determine the potential or eventual impact of a data standard?
- What specific use cases represent urgent and unmet needs for data and metadata standards?

**10:30—10:45 a.m. Report-out**

**10:45—10:50 a.m. Break**

**10:50 – 11:50 a.m. Breakout Groups:** Ensuring community engagement and models for maintenance and sustainability. Best practices for engaging communities throughout the standards development lifecycle and models for their maintenance.

**Facilitators:** **Eva Huala, Ph.D.**  
Phoenix Bioinformatics

**Warren Kibbe, Ph.D.**  
National Cancer Institute

**Michel Dumontier, Ph.D.**  
Stanford Center for Biomedical Informatics Research

**Rapporteurs:** **Yaffa Rubinstein, Ph.D.**  
National Center for Advancing Translational Sciences

**Rosemary McKaig, Ph.D.**  
National Institute of Allergy and Infectious Diseases

**Lesley Skalla, Ph.D.**  
MDB, Inc.

Questions to be addressed include:

- What are best practices for identifying relevant stakeholders?
- What are the incentives for participation of different stakeholders?

- How do governance models support (or discourage) community engagement?
- How does engagement translate into maintenance and sustainability?

**11:35—11:50 a.m. Report-out**

**11:50—12:20 p.m. Whole Group Discussion:** How to support data standards development? Issues to be addressed will be informed by earlier panel and breakout discussions.

- Accelerating the development and adoption of high quality data and metadata standards for NIH relevant research.
- Leading cultural change to recognize and reward individual and collective contributions to data standards development work.

**12:20—12:30 p.m. Summary and Next Steps**

**12:30 p.m. Main Meeting Adjournment**

DRAFT