



Hot Topics in Big Data: What You Need to Know Now!

A One-Day Workshop Co-sponsored by CENDI and NFAIS and Hosted by FLICC
Library of Congress, Mumford Room, 101 Independence Avenue, SE, Washington, DC 20540

December 11, 2012

9:00 am - 4:30 pm

PRELIMINARY AGENDA

9:00 am - 9:15 am: Welcome and Opening Remarks

9:15 am - 10:00 am: Big Data: An Overview of Key Concepts and Issues
George Strawn, NITRD (confirmed)

This session will provide a compelling overview of the current status or landscape of “Big Data.” Issues that will be covered are: The growth of scholarly and scientific data – why is this an issue today and where is it going; structured versus unstructured data; the volume and diversity of data depending on subject discipline; storage and use capacity, etc.; the need for policy and planning; staff skill sets (data management and curation); new work/analytic tools required to maximize data value; and linking, repositories, etc.

10:00 am - 10:15 am: Break and Networking Opportunity

10:15 am - 11:45 am: Institutional Players and Their Roles and the Challenges: Scientific Data Centers and Infrastructures
Carol Meyer, Earth Science Information Partnership (confirmed)
Myron Gutmann, NSF Director for Social Behavior and Economic Sciences (confirmed)
Third speaker to be determined

The speakers on this session are from organizations that represent a large group of libraries and information centers to provide a broad perspective of the current trends. Each of the organizations will be asked to answer the following questions:

- What has changed in your information landscape in the last 2-3 years?
- Has your organization been impacted by the growth of data?
- If “yes,” in what way and how are you adapting (issues such as staff skills, workflow issues, data management, systems, risk management, policies, etc. – be specific with regards to the hot topics/issues addressed at a higher level during the opening keynote)
- What are your biggest challenges?
- What changes do you see in the next five years?

11:45 am – 12:30 pm: Lunch *(will be provided)*

Hot Topics in Big Data: What You Need to Know Now! ~ Agenda Continued

12:30 pm – 2:00 pm **Data Citation, Data Linking and Supplemental Materials**
Giri Palanisamy, Oak Ridge National Laboratory (confirmed)
David Martinsen, American Chemical Society (confirmed)
Hillary Spencer, Nature Publishing (confirmed)

This session will look at the issues related to data citation, data linking, and supplemental materials covering issues related to publishers, librarians, users, etc.

2:00 pm - 2:15 pm: **Break and Networking Opportunity**

2:15 pm- 3:45 pm: **Big Data and Emerging Skill Set Requirements**
Michael Chui, McKinsey Study (confirmed)
David E. Schindel, Executive Secretary, Consortium for the Barcode of Life, Smithsonian Institution (confirmed)

This session will focus session on manpower issues at the macro level and relationship to skill sets at the macro level. It will touch on the McKinsey report as presented at the Bid Data Podcast in March 2012, which addressed overall numbers of professionals needed and related findings. Also included in this session is a look at specific skills at the micro level from the education and training perspective.

3:45 pm - 4:30 pm: **Data Analytics and Data Mining: The Approaching Tidal Wave**
Dennis Gannon, Director, Cloud Research Strategy, Microsoft Research Connections (confirmed)

The closing keynote talk will be an overview of the field of data analytics --definitions, description, what all the fuss is about, the successful applications and the opportunities and challenges. It will include what Microsoft is currently doing but will contain some examples about new initiatives and why they are developing them.