You are cordially invited
to view the

# Big Data to Knowledge (BD2K)

## *Workshop on Enhancing Training for Biomedical Big Data*

29 July, 10am-6pm
30 July, 8:30am-1:30pm
http://videocast.nih.gov/

**Workshop Co-chairs:** Karen Bandeen-Roche and Issac Kohane
**Contact:** BD2Kworkshops@mail.nih.gov

**Workshop Goals:**
1) Identify the knowledge, skills, and resources needed by biomedical research to organize, process, manage, and utilize large, complex data sets, and
2) Recommend and prioritize specific objectives for the NIH in training for Big Data.

This information will be used by NIH staff to develop short- and long-term training initiatives that prepare and empower the community to maximize the use of Big Data for research aimed at understanding human biology and improving human health.

**Background:**
Big Data to Knowledge (BD2K), a new NIH initiative, aims to enable scientists to effectively manage and utilize the large, complex data sets (Big Data) that are already being generated and whose number and value will increase in the future. The BD2K initiative is based on a set of recommendations on data and informatics from a working group to the Advisory Committee to the Director, NIH (see http://www.nih.gov/news/health/dec2012/od-07.htm). Enhancing training for biomedical Big Data is one of the four components of the BD2K initiative.

The NIH seeks to increase the ability of the scientific workforce to utilize biomedical Big Data. Big Data creates challenges to the data pipeline, from acquisition and processing of the data after their generation through analysis and visualization. Utilization and analysis of this data will require new knowledge and skills beyond those traditionally employed in biomedical research. Furthermore, such abilities will be required at all levels, from students through established faculty, in a diverse and sustainable workforce. The workshop will consider a refocus of traditional training programs toward being cross-disciplinary, and the development of focused, short-term training programs that are potentially technology-enabled, web-based, or otherwise widely accessible to investigators at all levels.

The workshop will (a) identify the knowledge and skills needed by individuals and collaborating teams to work productively with biomedical Big Data, and (b) discuss new resources and programs for educating and training both students and practicing scientists with the necessary knowledge and skills. The workshop will address the long- and short-term training needs of professionals and trainees with the purposes of increasing the number of: (1) informaticians and computational/quantitative scientists wishing to apply their skills and knowledge in the biomedical, behavioral, and clinical sciences and (2) biomedical, behavioral, and clinical scientists who have the requisite knowledge and skills to effectively access, organize, analyze, and integrate large and complex data sets.

**Monday, 29 July**

| | | |
|---|---|---|
| 10:00 | Welcome, Introductions, and Overview of BD2K | Mark Guyer |
| 10:30 | Purpose of the Workshop | Michelle Dunn |
| 10:45 | Summary of *Request for Information* Responses | Richard Baird |
| 11:15 | Intersection of BD2K with Director's Workforce and Diversity Initiatives | Sally Rockey |

11:45   Lunch – not provided

12:45   Discussion of the Goal and Vision for BD2K Training      Karen Bandeen-Roche, Zak Kohane

1:15   Data Challenges and Competencies Needed (10 min presentation + 10 min discussion)
- Electronic Health Records          Dan Masys
- Imaging          Ron Kikinis
- Genomics          Mike Boehnke
- Integration of Large or Small Datasets          Mark Musen

2:45   Discussion of BD2K Knowledge and Skills          Participants
- What are the necessary knowledge and skills that a Big Data team must include?
- How do the knowledge and skills needed vary according to the individual's:
  - primary relationship to Big Data?
    - needing to be conversant
    - applying routine methods and tools
    - leading novel applications
    - developing new methods and tools
  - primary training as basic, clinical, or quantitative scientists?
- How do we allow institutions adequate flexibility and still achieve the BD2K goals?

3:15   Break:  Refreshments will not be provided

3:30   BD2K Characteristics of Long-term Training and Career Award Programs:          Participants
Approach
- What type of person should long-term training aim to produce?
- How should individuals be cross-trained?
- How could the curriculum and other program components be modified or developed so that a cross-trained student would not have a longer time from matriculation to graduation?
- The generation of new methods and software are essential for biomedical Big Data.  Since computational and quantitative skills are broadly applicable, how should training programs encourage deployment or specialization of these skills in the biomedical field?
- What are the essential elements (e.g. courses, laboratory, clinical, or research rotations in industry, health care organizations, or government labs with big data) of a training program for a cross-trained student?

Environment
- What kind of an environment would be effective for BD2K-supported training?
- What would be a critical mass of students for a viable interdisciplinary program?
- What training program characteristics foster interaction between students trained in different disciplines, so that they learn from one another?

Policy
- Should NIH encourage common core elements in all BD2K-supported training programs?
- Should ALL training programs incorporate some elements of Big Data knowledge and skills into their curriculum?
- What should be the outcome of BD2K training programs and how should they be evaluated?

| | | |
|---|---|---|
| 5:45 | Brief Summary of Recommendations | Z. Kohane, K. Bandeen-Roche |
| 6:00 | Adjourn until Tuesday, 30 July 8:30am | |

**Tuesday, 30 July**

8:30    Distillation of Day 1                                                    K. Bandeen-Roche and Z. Kohane

9:00    Characteristics of BD2K Programs for Short-term training                 Participants
- Who should the target audience be—undergraduates, faculty at undergraduate institutions, graduate students, postdoctoral fellows, new and experienced investigators, clinicians?
- What can short-term training accomplish?  What concepts and skills can be conveyed via in this format? How would the success of such a program be evaluated?  What are the metrics of success?

10:00   Characteristics of BD2K Programs for Innovative Training Technology      Participants
- What innovative uses of technology could help 1) large numbers of students become familiar with basic core knowledge, or 2) established investigators acquire updated skills or an appreciation of new skills?
- How can online material be made interactive and adaptive to personalize delivery based on the learner's prior knowledge?
- How can NIH promote the development of training technologies specialized to biomedical Big Data?

11:00   Characteristics of BD2K Programs for Curriculum                          Participants
- Should NIH support curriculum development to encourage integrated, intersecting curricula?

11:30   Break (Working Lunch) -- Refreshments will not be provided

12:00   General Discussion                                                       Participants
- Are there particular challenges to keeping content updated?  How can sharing be encouraged?
- How should success of the programs be evaluated?  How can this activity be used to increase the number of students in research who are from underrepresented groups or less research-intensive institutions?
- What other training modalities should be considered (e.g. working groups, internships, etc.)?
- Of all the activities discussed, how would you prioritize them?
- Additional advice?

1:00    Summary of Workshop                                                      K. Bandeen-Roche, Z. Kohane

1:30    Adjourn