

Training and Workforce Development Breakout Session
BD2K AHM Friday, November 13th 2015

Co-chairs: Jack Van Horn and Erica Rosemond
Room: Balcony A

Participants interested in [the training and education efforts of the BD2K initiative](#) were invited to an open discussion.

- Additional Discussion Items?
 - Ways to address diversity in needs & interests of data science consumers?
 - Are there standards for resource dissemination?
- Potential topics for discussion:
 - Defining the field: What are the core competencies for biomedical data science?
 - Depends on the domain (home dept criteria satisfied, plus additional requirements)
 - One person's view:
<http://informaticsprofessor.blogspot.com/2015/07/what-is-difference-if-any-between.html>
 - Moving target; need broad categories of competencies
 - Focus on literacy and communication across disciplines - for research and for team-based projects
 - Competencies for different levels of learners; different backgrounds of the students (clinicians - focused content for different backgrounds)
 - Applied questions - to real life biomedical problems; can we have metadata with the course that describe applied biomedical questions
 - Build a team around the biomedical question - skillsets of needed participants are identified via directed/specific questions
 - Learning objectives needed; what should the students know at the end of the training
 - Soft skills are equally as important - writing, scholarship
 - Meeting needs of a variety of investigators: What are the most pressing needs of the current workforce of data science consumers? Of data science developers?
 - Resource discovery: How can discovery and access of training resources be improved? How can needs assessments for content be kept up-to-date?
 - EdX and Coursera deposits - need a resource similar to this for BD2K (TCC is developing an educational discovery index)
 - FAIR-E - FAIR tag/metadata for an educational resource
 - Archiving content - sustainability questions of educational materials
 - Updating curriculum - problems with archiving these resources
 - Standard metadata schema - intended audiences and links to it; filter in rights to the materials;
 - Data federation - pull resources together
 - NCIP Hub - check it out
 - Priming the pump: What are best practices for building a diverse pipeline of undergraduates? How has training evolved over the last few years in response to Big Data? In what ways does it still need to evolve?
 - Encourage females and URM - success stories?

- Real world applications - follow-up with individual students - using targeted questions approach; challenging STEM courses- need to test conceptual learning real-time
 - Women in technology - high school or earlier - target; need female role models as teachers; life style issues etc; have to start very early in the pipeline
- Career paths: What are the challenges in retention, reward and post-recruitment mentoring for early career scientists? How are issues of academic/university advancement relevant to Big Data researchers? What are some grant writing strategies for Big Data biomedical research?
 - Alternatives to publishing programs/packages
 - MS degrees - look to the Informatics field
 - Changing the culture / metrics for publishing software versus the biological (etc) paper
 - Sensitize applicants to promoting themselves and their contributions when it comes to tenure promotion
 - Use other metrics over than of the h-index
 - How to evaluate team science - may be a way to provide a metric in the realm of educational materials
 - MS in data science - gone viral (Berkeley etc - 100% employment)
- Working together: How can we promote regional and national BD2K training events?
- What will these students/participants be doing AFTER the training is complete?
 - Health-care, pharma, academia, other industry (depending on the skills they acquire through the program)
 - Show/demonstrate how exciting biomedical science is and provide specific opportunities to continue in biomedical data science (e.g., through connections with health care companies) - retain students/participants in biomedical research and healthcare; we don't want to lose these people to non-scientific opportunities (return on investment)

Additional Comments Based on Reports Back to the Group (later that day):

- Courses from an introductory to an advanced level are needed
- Computational resources for courses - hands-on learning/applications; cloud computing
- Suggest using a cloud credit model for training and education resources. Could there be an amazon partnership that includes a discount? NCI cloud pilots will be available in the Spring of 2016
- There is a need for hackathons
- Please see the API Working Group Break-Out notes - this group would like to educational efforts with engineers and others outside of BD2K; would like to see recruitment of people with limited experience / education in this domain area
- SMART API effort - educate across the centers

- Suggest to have a "Building a Digital Ecosystem Working Group" - suggest that all tools could be uploaded to the Commons to be used by trainees - perhaps the Centers can help with this