THE BD2K TRAINING COORDINATING CENTER (TCC): A RESOURCE FOR THE DATA SCIENCE COMMUNITY

John Darrell Van Horn, Ph.D.
USC Mark and Mary Stevens Neuroimaging and Informatics Institute
University of Southern California

November 29th, 2016
Presentation Outline

• Guiding principles
• BD2K Training Program Overview
• Role of the TCC
• Training Resource Indexing through ERuDIte
• User interactivity using BigDataU.org
• Data Science Seminar Series
• Data Science Innovation Labs
• RoAD Trip Science Rotations Program
• Big Data Biomedicine: The Movie
• Conclusions
FAIR with a “Silent E”?  

• FAIR – findable, accessible, interoperable, and re-useable (ELIXIR, FORCE11, BD2K, and others)  
• Things that are “fair” are balanced, equitable, open to all  
• FAIRE – Old English spelling for a celebration like a carnival; typically a village fete (UK); can also be referred to as a fair or a festival (US)  
• FAIRE, FAIR-E, FAIR(E), FAIRe or just FAIR (with an “invisible” E)  
• Adds to the collection of FAIR extensions and implications  
• Irrespective of how it is indicated, Education is a critical element of data, tools, methods, resources, etc which we wish to consider FAIR(E).
BD2K Training Programs

- Training programs across the BD2K enterprise represent a broad range of undergraduate, graduate, and post-doctoral programs, career path development, in-person workshops seminars, virtual events, video lectures, among other unique activities.
- While funded through a variety of NIH grant mechanisms, these BD2K training programs are, in fact, part of an integrated, collective whole.
- Through close interactions with these programs, the NIH and TCC seek to promote data science as a 21st Century response to the need for more scientists with the computational skills to take on our nation’s most serious biomedical research challenges.

U24  
R25  
K01  
T15/T32  
U54 Centers  
dR25
# List of BD2K Training Effort Awards (2015-2016)

## Training/Educational Development (R25)

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorr, David A.</td>
<td>Oregon Health &amp; Science University</td>
</tr>
<tr>
<td>Shojaie, Ali</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Pathak, Jyotishman</td>
<td>Mayo Clinic Rochester</td>
</tr>
<tr>
<td>Recht, Michael P.</td>
<td>New York University School of Medicine</td>
</tr>
<tr>
<td>Kovatch, Patricia</td>
<td>Icahn School of Medicine at Mount Sinai</td>
</tr>
<tr>
<td>Mukherjee, Bhraman</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Hoffmann, Alexander</td>
<td>University of California Los Angeles</td>
</tr>
<tr>
<td>Chuang, Jeffrey Hs-M</td>
<td>Jackson Laboratory</td>
</tr>
<tr>
<td>Fowlkes, Charless</td>
<td>University of California-Irvine</td>
</tr>
<tr>
<td>Shaw, Joseph R.</td>
<td>Mount Desert Island Biological Lab</td>
</tr>
<tr>
<td>Zhang, Min</td>
<td>Purdue University</td>
</tr>
<tr>
<td>Martin, Elaine R.</td>
<td>Univ of Massachusetts Med Sch Worcester</td>
</tr>
<tr>
<td>Haddad, Bassem R</td>
<td>Georgetown University</td>
</tr>
<tr>
<td>Surkie, Alisa</td>
<td>New York University School of Medicine</td>
</tr>
<tr>
<td>Lawson, Catherine L</td>
<td>Rutgers, The State Univ of N.J.</td>
</tr>
<tr>
<td>Seymour, Anne</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>Caffo, Brian Scott</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>Irizarry, Rafael Angel</td>
<td>Harvard School of Public Health</td>
</tr>
<tr>
<td>Pevzner, Pavel A</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>Hersh, William R</td>
<td>Oregon Health &amp; Science University</td>
</tr>
<tr>
<td>Amaro, Rommie E</td>
<td>University of California San Diego</td>
</tr>
<tr>
<td>Lee, Christopher</td>
<td>University of California Los Angeles</td>
</tr>
<tr>
<td>Bohland, Jason W</td>
<td>Boston University (Charles River Campus)</td>
</tr>
<tr>
<td>Elgin, Sarah C.R.</td>
<td>Washington University</td>
</tr>
</tbody>
</table>

## Training/Career Development (K01s)

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avants, Brian</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Callcut, Rachael A</td>
<td>University of California, San Francisco</td>
</tr>
<tr>
<td>Chen, Jonathan Hailin</td>
<td>Stanford University</td>
</tr>
<tr>
<td>Coffman, Donna Lynn</td>
<td>Pennsylvania State University</td>
</tr>
<tr>
<td>Farhat, Maha</td>
<td>Massachusetts General Hospital</td>
</tr>
<tr>
<td>Garmire, Lana X</td>
<td>University of Hawaii at Manoa</td>
</tr>
<tr>
<td>Gliske, Stephen V</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>Inakura, Haruka</td>
<td>Stanford University</td>
</tr>
<tr>
<td>Johnson, Michael</td>
<td>Johns Hopkins University</td>
</tr>
<tr>
<td>Hiroshi</td>
<td></td>
</tr>
<tr>
<td>Landau, Dan</td>
<td>Dana-Farber Cancer Institute</td>
</tr>
<tr>
<td>Lee, George</td>
<td>Case Western Reserve University</td>
</tr>
<tr>
<td>Nemati, Shamim</td>
<td>Emory University</td>
</tr>
<tr>
<td>Nguyen, Quynh</td>
<td>University of Utah</td>
</tr>
<tr>
<td>Nsoesie, Elaine O.</td>
<td>Children's Hospital Corporation</td>
</tr>
<tr>
<td>Paguirangan, Amy</td>
<td>Fred Hutchinson Cancer Research Center</td>
</tr>
<tr>
<td>Park, Soojin</td>
<td>Columbia University Health Sciences</td>
</tr>
<tr>
<td>Pearson, John</td>
<td>Duke University</td>
</tr>
<tr>
<td>Prokop, Jeremy W.</td>
<td>Medical College of Wisconsin</td>
</tr>
<tr>
<td>Schmitt, James E</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Van Panhuus, Willem</td>
<td>University of Pittsburgh at Pittsburgh</td>
</tr>
</tbody>
</table>

## Diversity (dR25)

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canner, Judith Elena</td>
<td>California State Univ, Monterey Bay</td>
</tr>
<tr>
<td>Qian, Lei</td>
<td>Fisk University</td>
</tr>
<tr>
<td>McEligot, Archana J</td>
<td>California State University Fullerton</td>
</tr>
<tr>
<td>Garcia-Arraras, Jose E</td>
<td>University of Puerto Rico Rio Piedras</td>
</tr>
</tbody>
</table>

## Institutional Training (T32/T15)

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canner, Judith Elena</td>
<td>California State Univ, Monterey Bay</td>
</tr>
<tr>
<td>Qian, Lei</td>
<td>Fisk University</td>
</tr>
<tr>
<td>McEligot, Archana J</td>
<td>California State University Fullerton</td>
</tr>
<tr>
<td>Garcia-Arraras, Jose E</td>
<td>University of Puerto Rico Rio Piedras</td>
</tr>
<tr>
<td>Altman, Russ</td>
<td>Stanford University</td>
</tr>
<tr>
<td>Amos, Christopher</td>
<td>Dartmouth College</td>
</tr>
<tr>
<td>Daniels, Michael</td>
<td>University of Texas, Austin</td>
</tr>
<tr>
<td>Malin, Bradley</td>
<td>Vanderbilt University</td>
</tr>
<tr>
<td>Newton, Michael</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>Papin, Jason</td>
<td>University of Virginia</td>
</tr>
<tr>
<td>Quackenbush, John</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Ritchie, Marylyn</td>
<td>Pennsylvania State University</td>
</tr>
<tr>
<td>Shya, Chi-Ren</td>
<td>University of Missouri</td>
</tr>
<tr>
<td>van der Laan, Mark</td>
<td>University of California, Berkeley</td>
</tr>
</tbody>
</table>

## Training Coordination Center (U24)

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Horn, John Darrell</td>
<td>University of Southern California</td>
</tr>
</tbody>
</table>
BD2K Training Coordinating Center (TCC)

NIH U24 Award

- Bigdatau.org
- Innovation Lab
- Calls for Applications
- Training Events
- BD2K Training News

Training Coordinating

- U54, R25, T32, K01s
- Working Groups
- Core Competencies
- Resource Discovery
- Diversity
- Career Paths

Indexing Training Materials

- "ERuDITe"
- "Knowledge map"
- Personalized Training
- Curriculum construction
- Training "Workflows"

Science Rotations

- Matching young biomedical researchers with senior quantitative scientists
- Fund two-week intensive residencies

Public Outreach

- Facebook Page
- Google Calendar
- Mailing Lists
- USC School of Cinematic Arts

Supplemental Projects

- Scoping Workshop
  - mHealth Mobile Health
  - NIH and NSF Program Officials
  - Craft the themes for the Innovation Lab

- Innovative Lab
  - 30 participants and 7 mentors
  - Background in biomed and statistics/computer science

- National Science Foundation
  - Innovation Lab Travel support for quantitative scientists
  - Mathematicians
  - Statisticians
  - Computer Science

- International Meeting
  - On training, held at USC, including
  - TCC
  - Elixir
  - H3Africa
  - NIH and NSF
  - Others

- Career Paths
  - University of Illinois, Urbana-Champaign
  - Bring university leadership together to discuss the shifting notion of data science in higher education
Educational Resource Discovery Index (ERuDItE)

• Facilitate the discovery, access, and citation of educational resources through the development of a living educational resource discovery index (ERuDItE).

• ERuDItE is a framework which may be enriched in multiple ways
  • The learning objectives and content can be organized into a framework by experts
  • "learned" through mining and clustering of the metadata
  • The TCC seeks to develop methods and technology to tag indexed educational resources with these learning objectives as new resources are added, to help researchers find training materials of interest to them.

• Personalize the discovery of biomedical data science educational resources.

• Leverage social media, usage statistics, etc to enhance what people view and take advantage of
ERuDItte Knowledge Maps (Version 1.0)

- Identify/Organize Training Meta
- Indexing Courses in ERuDItte
- Compute Similarities
- Invoke Machine Learning
- Extract Training Concepts
- Render ERuDItte Mappings
- Apply User Navigation
- Enable Personalized Training
Big Data U Website

bigdatau.org

- About TCC and the TCC Team
- TCC Interactions
- BD2K Training Grants
- Calendar of all BD2K Training Events
- BD2K Data Science Seminar Series
- TCC News
- Data Science Innovation Lab
- RoAD-Trip Program
- About ERuDIte
- Explore ERuDIte
- ERuDIte Dashboard
ERuDIte User Dashboard

- Users can draw from ERuDIte to populate topic-specific learning plans on their own personal dashboard
- Resources can be arranged in any way the user wishes
  - Drag-n-drop
  - Auto-arranged
- Icons with each resource “square” indicate
  - The type of resource
  - Average “Star Rating”
  - Whether the user has completed that resource
  - etc
- User-defined resources can be easily added, stored locally, or “in the cloud”
- Learning plans and user-defined resources can be easily shared
- Based on a user-profile and/or learning plans, resource recommendations can be made
- Dashboard is persistent and available via any location
- Mobile friendly as much as possible

See Poster and Demo by Sumiko Abe and Jeana Kamdar
The BD2K Guide to the Fundamentals of Data Science Series

Every Friday beginning September 9, 2016
12pm - 1pm Eastern/ 9am - 10am Pacific

http://www.bigdatau.org/data-science-seminars
Data Science Innovation Lab 2016

• Mentored, week-long residential program, June 14-19, 2016

• Mobile Health (mHealth) as Big Data

• Lake Arrowhead Conference Center

• http://www.bigdatau.org/innovationlab
Data Science Innovation Lab 2017

• Mentored, week-long residential program, June 18-22, 2017
• The Microbiome
• Wylie Inn and Conference Center
  • Northeast of Boston, MA
• http://www.bigdatau.org/innovationlab
Data Science “RoAD Trip” Program

• Create new partnerships
  • Junior biomedical scientists
  • Senior data scientists

• Novel data challenges necessitating data science methods and resources

• Matching plus peer-review selection process

• Junior fellows “take to the road” to spend ~2 weeks at the laboratories of the senior mentors

• Process, model, analyze, and visualize “big data”

• Develop next steps
  • Papers, conference presentations, grant proposals, etc
Big Data Biomedicine: The Movie

SPECIAL PREMIER TODAY!

Tuesday November 29th, 2016
5:30pm and 6pm in Salon E
Conclusions

• We believe that the TCC is quickly solidifying into a robust compendium of educational resources from around the BD2K and beyond
• We are confident that ERuDIt e will scale effectively as more resources are automatically included (e.g. courses, video lectures, documents, slide sets)
• Data science being applied has been remarkable and will continue to inform the ongoing organization of ERuDIt e (e.g. topics, keywords, clustering, etc)
• The attractive Big Data U website user experience we are developing will personalize how resources can be included, arranged, shared, etc via user-specified dashboards
• A variety of “maps” and interactive search capabilities will enable users to “find” the resources which matter most to them and their training
• Our additional activities promote biomedical data science engagement as a participatory and community-oriented enterprise
• Our TCC efforts can be fully expected to put the “E” in FAIR
Thanks go to...
Contact the BD2K TCC

USC Mark and Mary Stevens Neuroimaging and Informatics Institute
University of Southern California
2025 Zonal Avenue, SHN
Los Angeles, CA 90033

URL: http://bigdatau.ini.usc.edu,
http://www.bigdatau.org
Phone: 323-44-BRAIN
Email: jvanhorn@usc.edu