Filling the Workforce Gap in Data Science and Data Analytics

Catherine Blake  
Graduate School of Information and Library Science (GSLIS)  
Center for Informatics Research in Science and Scholarship (CIRSS)  
University of Illinois  
clblake@illinois.edu

Jeffrey M. Stanton  
School of Information Studies  
Syracuse University  
jmstanto@syr.edu

AnnaLee Saxenian  
School of Information and Department of City and Regional Planning  
UC Berkeley  
anno@ischool.berkeley.edu

Abstract

The goal of this workshop is to provide a forum for iSchool faculty who are developing programs in data analytics, eScience, eResearch, big data, and cyberinfrastructure to discuss best practices with respect to preparing students to fill the workforce need for managers and analysts to analyze big data and make decisions based on their findings.

Keywords: Data Science, Big Data, Socio-technical data analytics, Information and Data Science

Changes in how information is created, disseminated and re-used provide new opportunities to intervene within the information lifecycle to ensure that the information created today will be available for centuries to come. Jim Gray, the first advocate of data intensive science, urges us to “support the whole research cycle – from data capture and data curation to data analysis and data visualization.” (Hey, Tansley, & Tolle, 2009). The “data deluge” (Hey & Trefethen, 2003) is now a fundamental characteristic of e-science and “big science,” especially in disciplines such as in cancer (e.g., National Center for Biotechnology Information), astronomy (e.g., the Sloan Sky Survey), and atmospheric science (e.g., coupled climate models). The transition towards data-intensive science has created a critical shortage of both a knowledgeable workforce and best practices to address the challenges of data management, curation, and analysis (Kim & Stanton, 2012; Kim, Addom, & Stanton, 2011; Palmer et al, 2007; Palmer, Renear, & Cragin, 2008) and information synthesis (Blake & Pratt, 2006a, 2006b).

Programs such as “DataNet” have been created to integrate “library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise to provide reliable digital preservation, access, integration, and analysis capabilities for science and/or engineering data over a decades-long timeline” (NSF, 2007). The Institute of Museum and Library Services (IMLS) has also invested heavily in educational initiatives that train data curation leaders around the country. Now that data curation training programs are well underway, it is time to focus our attention on workforce gaps in the latter activities in the information lifecycle such as in the McKinsey Global Institute report published earlier this year, which stated that “A shortage of the analytical and managerial talent necessary to make the most of big data is a significant and pressing challenge …” (Manyika et al., 2011, p3). The report goes on to say that “The United States alone faces a shortage of 140,000 to 190,000 people with deep analytics skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings” (Manyika et al., 2011, pg 104).

The goal of this workshop is to provide a forum for iSchool faculty who are developing programs in data analytics, eScience, eResearch, big data, and cyberinfrastructure to discuss best practices with respect to preparing students to fill the workforce need for managers and analysts to analyze big data and make decisions based on their findings. Questions of interest include (but are not limited to):

- What differentiates programs in this area (e.g. a program in big data vs. one in cyberinfrastructure? data analytics? eScience?  
- To what extent can big data be incorporated into the iSchool classroom experience?  
- What opportunities might exist to share data and teaching models between institutions?
• For some students analytics is a very different way of thinking about information. What teaching strategies have been successful in bringing students up the learning curve without sacrificing intellectual rigor?
• To what extent can faculty infuse their analytics research into the classroom setting?
• iSchool programs in data science and data analytics tend emphasize the context in which data is collected and situate analytics within the broader work-flows in science, business and the community. To what extent can this balance between social and technical aspects of big data be maintained?
• What makes iSchools uniquely prepared to address research and educational requirements of data science, data analytics, cyberinfrastructure and eResearch?

The workshop organizers provide three different iSchool flavors of data science and analytics at different stages of development. Specifically, Dr. Blake will discuss the MS and PhD. Program Socio-technical Data Analytics at Illinois, Dr. Stanton will discuss the CAS Data Science Program at Syracuse and Dr. Saxenian will discuss the MS program in Information and Data Science Program at UC Berkeley. Each organizer will present their respective program and address one or more of the questions above and thus demonstrate that each iSchool brings a different perspective to the big data workforce need.

With that said the iSchools emphasis on people, information and technology makes them well prepared to address both the research and educational gaps in data science and data analytics.

This workshop will be of interest for students who are interested in learning more about how training needs in big data are being addressed in iSchools, and for faculty who are developing new programs. We will ask participants to submit a two page position statement describing the research and educational issues and one or more of the questions outlined above before the workshop. After the initial organizer presentations, workshop participants will be broken into small groups to address the questions above and in particular, identify gaps between training needs and existing programs. Group leaders will then report back to the group.

References


