

# DATA SCIENCE SEMINAR

Bringing together thought-leading speakers and researchers across campus to discuss topics related to data analysis, visualization and applications to domain sciences. Everyone interested is welcome to attend.

Wednesday, Apr. 11

Location: Physics/Astronomy Auditorium, room A102

TIME | 3:30 - 4:20pm

<http://data.uw.edu/seminar>

SPR 2018



John Darrell Van Horn, M.Eng., Ph.D.

Associate Professor of Neurology, University of Southern California

## Making data science training resources FAIR

### ABSTRACT

In our rapidly evolving information era, methods for handling large quantities of data obtained in biomedical research have emerged as powerful tools for confronting critical research questions. These methods are having significant impacts in diverse domains ranging from genomics, to health informatics, to environmental research, and beyond. The NIH's Big Data to Knowledge (BD2K) Training Consortium, in particular, has worked to empower current and future generations of researchers with a comprehensive understanding of the data science ecosystem, giving them the ability to explore, prepare, analyze, visualize, and interpret Big Data. To this end, the BD2K Training Coordinating Center (TCC) was funded to facilitate in-person and online learning, and to open the concepts of data science to the widest possible audience. In this presentation, I will describe the activities of the BD2K TCC, particularly the construction of the Educational Resource Discovery Index (ERuDIte). ERuDIte identifies, collects, describes, and organizes over 10,000 data science training resources, including: online data science materials from BD2K awardees; open online courses; and videos from scientific lectures and tutorials. Given the richness of online training materials and the constant evolution of biomedical data science, computational methods applying information retrieval, natural language processing, and machine learning techniques are required. In effect, data science is being used to inform training in data science where the so-called FAIR principles apply equally to these resources as well as to the datatypes and methods they describe. As a result, the work of the TCC has aimed to democratize novel insights and discoveries brought forth via large-scale data science training. This presentation will be of interest to anyone seeking to personalize their own data science education, craft unique online training curricula, and/or share their own online training content.

### BIO

John Darrell Van Horn, M.Eng., Ph.D. is an Associate Professor of Neurology at the University of Southern California. He is also a Senior Lecturer at the University of Washington. He is currently the Director of the Center for Data Science and Informatics at the University of Washington. He is also the Director of the Center for Data Science and Informatics at the University of Washington. He is also the Director of the Center for Data Science and Informatics at the University of Washington.

The University of Washington is committed to providing access, equal opportunity and reasonable accommodation in its services, programs, activities, education and employment for individuals with disabilities. To request disability accommodation contact the Disability Services Office at least ten days in advance at: 206-543-6450/V, 206-543-6452/TTY, 206-685-7264 (FAX), or [dso@u.washington.edu](mailto:dso@u.washington.edu).