Stony Brook University

Biomedical Informatics Grand Rounds

Jonas Almeida, PhD;

Chief Data Scientist, *Division of Cancer Epidemiology & Genetics, National Cancer Institute*

Data, Analyze Thyself!

Wednesday, February 19, 2020 3 pm—4 pm
Health Science Center L3-LH6

Abstract:

The intersection of FAIR principles with data-intensive infrastructure for cancer epidemiology studies presents a number of challenges to the configuration of Data Commons. Ultimately, data analysis will have to rely of code migrating to where the data is hosted. In that end, we'll discuss lessons learned from large scale cohort studies at NCI's Division of Cancer Epidemiology and Genetics (DCEG), and review the role of consumer-facing Cloud+Web initiatives such as those enabled by NIH STRIDES program. The discussion includes live experimentation with prototype applications developed both at NCI and at Stony Brook.

Bio:

Dr. Jonas Almeida leads a multidisciplinary research program that combines systems biology, computational statistics, and software engineering for biomedical applications. The primary focus of his research is to accelerate the investigation of epidemiologic and genetic causes of cancer by developing innovative methods to advance the computational research infrastructure for conducting precision prevention studies of cancer. Dr. Almeida seeks to identify and deliver consumer-facing architectures for precision medicine and prevention that employ cloud computing, web applications, and machine learning. He explores this interrelated computational ecosystem by developing portable software solutions that can migrate between data sources – from consumer genomics to wearable sensing devices, and between different contexts of application from patients to caregivers. As Chief Data Scientist for the NCI Division of Cancer Epidemiology and Genetics (DCEG), Dr. Almeida has the dual responsibilities of 1) leading efforts for the integrated creation, management, and analysis of data-intensive knowledge bases, establishing cost-effective scalable researcher-facing analytic infrastructure, and defining new infrastructure to extract, manage, and analyze data in a scalable way to support epidemiological research within the Division; and 2) conducting independent research to advance real-time analytics of cancer "Big Data". He provides support for the data and technology infrastructure used by the new prospective multi-center cohort study that will serve as an important Division-wide resource. Dr. Almeida received his Ph.D. in Biological Engineering from the University Nova of Lisbon, Portugal in 1995. After a postdoctoral fellowship in computational statistics and machine learning at the University of Tennessee and Oak Ridge National Laboratory, he became an Assistant Professor in Chemistry at the University of Lisbon in 1996, followed by an appointment as an Associate Professor of Biostatistics at the Medical University of South Carolina in 2001 and as a tenured Professor of Bioinformatics in the Division of Applied Mathematics of the University of Texas MD Anderson Cancer Center in 2006. In 2008, MD Anderson awarded Dr. Almeida the endowed Abell-Hanger Distinguished Professorship in Bioinformatics. In 2011, he was recruited to be the inaugural Director of a new Division of Informatics and tenured Professor in the Department of Pathology of the University of Alabama at Birmingham. In 2015, he joined Stony Brook University as a tenured Professor and Chief Technology Officer. Dr. Almeida was appointed Chief Data Scientist of DCEG in 2019.

CME Credit Available

Questions? Please call the Biomedical Informatics Department at 631-638-2590.