

Systems Biology Research Symposium

Oral Presentation Session

Grand Ballroom
Tuesday, June 5th
7:00-8:30pm

Data Integration for Dynamic and Sustainable Systems Biology Resources: Challenges and Lessons

Dan Sullivan, Joe Gabbard, Maulik Shukla, and Bruno Sobral

Virginia Bioinformatics Institute, Virginia Tech, Washington Street, MC 0477, Blacksburg, VA 24061

Presenter's email address: dsulliva@vbi.vt.edu

Systems biology and host-pathogen research is increasingly dependent on integrating data from diverse sources. Maintaining integrated resources over long periods of time presents distinct challenges. This paper describes experiences and lessons learned from integrating data in two 5-year projects focused on pathosystems biology: the Pathosystems Resource Integration Center (PATRIC, <http://patric.vbi.vt.edu/>), with a goal of developing bioinformatics resources to the research community based on genomics data, and the Biodefense Proteomics Resource Center (BPRC), <http://www.proteomicsresource.org>), with goal of developing resources based on the experiment data such as microarrays and proteomics data from diverse sources. Some challenges include: integrating genomic sequence and experiment data, data synchronization, data quality control, graceful degradation, and exposing metadata. We present examples of a variety of data integration problems drawn from our experiences with PATRIC and BPRC as well as open research questions related to long term sustainability and describe next steps to meeting these challenges. PATRIC is funded through NIAID contract number HHSN266200400035C

Key words: pathosystems biology, data integration, systems architecture