

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

Dan Sullivan<sup>1</sup>, Joseph L. Gabbard Jr<sup>2</sup>, Maulik Shukla<sup>1</sup>, Bruno Sobral<sup>1</sup>

<sup>1</sup> Virginia Bioinformatics Institute  
Washington Street, MC 0477  
Virginia Tech, Blacksburg, Virginia 24061

<sup>2</sup> Center for Human-Computer Interaction  
2202 Kraft Drive  
Virginia Tech, Blacksburg, Virginia 24061

### **Abstract**

Systems biology and infectious disease (host-pathogen-environment) research is becoming increasingly dependent on integrating data from diverse and dynamic sources. Maintaining integrated resources over long periods of time presents distinct challenges. This paper describes experiences and lessons learned from integrating data in two five-year projects focused on pathosystems biology: the Pathosystems Resource Integration Center (PATRIC, <http://patric.vbi.vt.edu/>), with a goal of developing bioinformatics resources for the research and countermeasures development communities based on genomics data, and the Resource Center for Biodefense Proteomics Research ((RCBPR) (RBPRC), <http://www.proteomicsresource.org/>), with a goal of developing resources based on the experiment data such as microarray and proteomics data from diverse sources and technologies. Some challenges include: integrating genomic sequence and experiment data, data synchronization, data quality control, and usability engineering. We present examples of a variety of data integration problems drawn from our experiences with PATRIC and RBPRC, as well as open research questions related to long term sustainability, and describe the next steps to meeting these challenges.