

# **DISCOVERY INFORMATICS IN BIOLOGICAL AND BIOMEDICAL SCIENCES: RESEARCH CHALLENGES AND OPPORTUNITIES**

VASANT HONAVAR

*College of Information Sciences and Technology, Pennsylvania State University  
University Park, PA 16802, USA  
Email: [vhonavar@ist.psu.edu](mailto:vhonavar@ist.psu.edu)*

New discoveries in biological, biomedical and health sciences are increasingly being driven by our ability to acquire, share, integrate and analyze, and construct and simulate predictive models of biological systems. While much attention has focused on automating routine aspects of management and analysis of "big data", realizing the full potential of "big data" to accelerate discovery calls for automating many other aspects of the scientific process that have so far largely resisted automation: identifying gaps in the current state of knowledge; generating and prioritizing questions; designing studies; designing, prioritizing, planning, and executing experiments; interpreting results; forming hypotheses; drawing conclusions; replicating studies; validating claims; documenting studies; communicating results; reviewing results; and integrating results into the larger body of knowledge in a discipline. Against this background, the PSB workshop on Discovery Informatics in Biological and Biomedical Sciences explores the opportunities and challenges of automating discovery or assisting humans in discovery through advances (i) Understanding, formalization, and information processing accounts of, the entire scientific process; (ii) Design, development, and evaluation of the computational artifacts (representations, processes) that embody such understanding; and (iii) Application of the resulting artifacts and systems to advance science (by augmenting individual or collective human efforts, or by fully automating science).

The workshop, which is especially timely in the context of the NIH Big Data to Knowledge (BD2K) initiative, brings together a group of scientists with complementary expertise to explore research challenges and opportunities in the informatics of discovery in biomedical sciences including, but not limited to:

- Representation and modeling languages with precise formal semantics, for describing, sharing, and communicating scientific models, theories, and hypotheses in biomedical sciences.
- Novel approaches to interactive visualization and exploration of complex biomedical data.
- Sophisticated approaches to construction of comprehensible and communicable predictive models and discovery of causal mechanisms from disparate types of observational and experimental data, literature, images, spatial, temporal, richly structured e.g., network data in biomedical sciences.
- Effective approaches for acquiring and making effective use of background assumptions, hypotheses, knowledge, beliefs and conjectures, arguments, domain expertise, and process descriptions in biomedical sciences.
- Algorithms and tools for automating various aspects of discovery.

Several individuals have contributed to the workshop: Tim Clark, Harvard University; Michel Dumontier, Stanford University; Yolanda Gil, University of Southern California; Lawrence Hunter, University of Colorado at Denver; Marylyn Ritchie, Pennsylvania State University; Andrey Rzhetsky, University of Chicago; Alan Ruttenberg, University at Buffalo; Neil Smalheiser, University of Illinois at Chicago; Nigam Shah, Stanford University.