Executive Summary of the 2012 NSF Workshop on Discovery Informatics

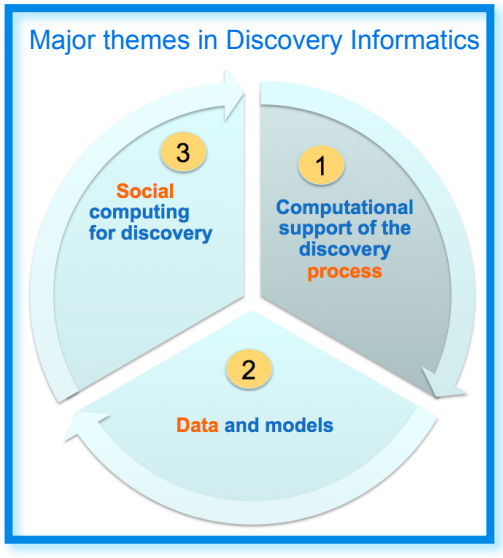
Advances in computing are transforming nearly all areas of science and engineering. In turn, the pursuit of new discoveries has resulted in innovations across all areas of computing. We are now facing the limits of our ability to gain insight from the volume, variety, and velocity of available data, posing fundamental challenges that can only be addressed through symbiotic advances in computing. Our ability to understand and gain insight from data of unprecedented complexity could be greatly increased with appropriate intelligent assistance and automation.

The Workshop on Discovery Informatics was convened to articulate the research challenges concerned with the management of knowledge and of the complex processes involved in scientific discovery. Workshop participants identified an expansive range of fundamental research challenges for information and intelligent systems brought into focus around three themes:

1. *New computational approaches are needed to manage the complexity of discovery processes that surpass human cognitive abilities.* This complexity often hampers scientists' knowledge and ability to analyze the large amounts of data at their disposal. We have reached a point where cognitive limitations are constraining scientific progress. We need to make scientific processes easily inspectable and reproducible. Innovations are needed to augment human abilities to analyze complex data through sophisticated processes, and enable understanding and insight.

2. *New computational approaches are needed to increase the connections between knowledge and data and exploit them to facilitate scientists' understanding of complex phenomena.* Data leads to new scientific knowledge, but the connection between knowledge and data is often not explicitly preserved in existing computational frameworks. As more complex data becomes available with increasing volume, variety, and velocity, the exploration of models becomes unmanageable, hurting our ability to do science effectively. We must develop general mechanisms for automated data-driven model refinement, data collection guided by models, and model-driven data analysis.

3. *New computational approaches are needed to flexibly combine diverse human abilities to tackle science problems that may not be otherwise considered possible.* New opportunities for discovery lie in the amalgamation of human expertise and effort. Although collaborations among scientists are common we have limited ability to facilitate unplanned, cross-disciplinary collaborations. In addition, we need better mechanisms to bring to bear human creativity to complement brute force computation, and open up science to valuable problem solving from massive amounts of volunteer contributors.



Existing relevant research efforts are scattered across disciplines and lack the critical mass needed to make a significant impact on these challenging aspects of science. Advances in these areas will transform the practice of science in two ways: 1) improving *existing* discovery processes that are unmanageable and suffer from human cognitive limitations, and 2) developing *new* discovery processes that increase our ability to understand challenging scientific phenomena. Further, outcomes in these areas are not domain specific, and can be leveraged across different science and engineering disciplines, having multiplicative returns, avoiding the inefficient, redundant development of computing innovations that would otherwise be repeated in specific disciplines (e.g., bio-, geo-, eco-informatics).

Discovery Informatics focuses on computing advances aimed at identifying scientific discovery processes that require knowledge assimilation and reasoning, and applying principles of intelligent computing and information systems in order to understand, automate, improve, and innovate any aspects of those processes. A new initiative in Discovery Informatics would enable and catalyze the transformational innovations needed to have a broad impact on the improvement and innovation of scientific discovery processes.

Discovery Informatics would require advancing basic research in many areas of computing, including: information extraction and text understanding to process publications and lab notebooks; synthesis of models from first principles, hypotheses, or data analysis; dynamic and adaptive design of data analysis methods; design, execution, and steering of experiments; selective data collection; data and model visualization; theory and model revision; collaborative activities that improve data understanding and synthesis; intelligent interfaces for scientists; design of new processes for scientific discovery; and computational mechanisms to represent and communicate scientific knowledge to colleagues, researchers in other disciplines, students, and the public.

Discovery Informatics will accelerate 21st century science and will have outcomes vital to the nation in numerous ways. National security is in severe need of better technologies for data analysis, noticing the unusual, and discovering patterns. Personal health and preventive medicine depend on our ability to enable people to contribute to the scientific enterprise in meaningful ways, by contributing data, analysis, personal histories, and sensor data. Our future relies on a better understanding of environmental and sustainability factors that is well beyond our current abilities. Our national competitiveness will be significantly boosted by a significant push in our nation's capabilities as a knowledge economy that would result from a renewed strength in Discovery Informatics. Discovery Informatics will advance the frontiers of computing, particularly in emerging areas of information and intelligent systems, while enabling new discoveries and innovations in all areas of science and engineering.

Participants stressed the need to act immediately. There is no doubt that our ability to generate and share data has surpassed our ability to analyze it. There is no doubt that there is data available or ready to be collected that could lead to many great discoveries of societal importance. We should strive to be in a position where not only can we harness the vast amounts of data at our disposal, but we are also able to pose increasingly complex questions that current methods do not even allow us to begin to imagine.

From "Final Report on the 2012 NSF Workshop on Discovery Informatics." Yolanda Gil and Haym Hirsh (Eds). National Science Foundation Workshop Report, August 2012. Available from http://www.discoveryinformaticsinitiative.org/diw2012. The workshop was sponsored by the Division of Information and Intelligent Systems of the Directorate for Computer and Information Sciences at the National Science Foundation under grant number IIS-1151951.