



E-Science

David De Roure, *University of Southampton*

Yolanda Gil, *Information Sciences Institute*

James A. Hendler, *University of Maryland*

This special issue brings you an overview of trends in multidisciplinary and large-scale science that are drawing in AI techniques to address the complexity and diversity of future scientific environments.

The science domain is no stranger to AI research. Many scientific questions have interested AI researchers and resulted in valuable contributions to science as well as advances in AI itself, from early knowledge systems¹ to machine learning² to natural language.³ In putting together this issue, we aim to convey new directions and opportunities that e-science is raising and that will challenge our field in the coming years.

The Expert Opinion department piece by David De Roure and James Hendler, "E-Science: The Grid and the Semantic Web," articulates how grid computing and Semantic Web technology must be brought together to enable future scientific endeavors. Because of their need for high-performance computing resources, many scientists are drawn to grid computing as the infrastructure to support data management and analysis across sites and organizations. Grids provide basic facilities for robust, efficient file management, transfer, and sharing, and they

support distributed computation by managing the execution of complex job workflows. Not yet a decade old, grid computing started to build from the bottom up and is already confronting the need for a more knowledge-rich infrastructure to address the challenges that multidisciplinary large-scale science raises. Although coming from a different direction, the Semantic Web vision also was motivated by the need to support scientific collaboration. By enabling transparent document sharing, metadata annotations, and semantic integration, it addresses multidis-

ciplinary distributed science research at the end-user level. The article articulates how grid computing and Semantic Web technology are complementary and outlines the challenges and immediate benefits of their integration.

Chris Wroe and his colleagues describe already ongoing work in this direction in "Automating Experiments Using Semantic Data on a Bioinformatics Grid." The article describes how they use Semantic Web languages to integrate services and data on grid computing. Their work supports interactive bioinformatics experimentation by pulling information from distributed heterogeneous services and integrating it through expressive semantic representations of their content and the operations they support. They introduce the notion of *scientific workflows*. These workflows consist of individual processing steps whose results are consumed by others and that overall generate the desired data products of scientific analysis.

In "Artificial Intelligence and Grids: Workflow Planning and Beyond," Yolanda Gil and her colleagues describe AI techniques that automatically generate complex, detailed workflows that can execute on a grid. Workflow generation is a complex decision space that is well suited to planning techniques, heuristics, and constraint reasoning to reason about each processing step's requirements and the data and resources available. Although manual creation of these workflows has been possible in the past, it becomes unmanageable as the tasks' size and complexity continue to increase. Gil and her colleagues also argue that knowledge-rich descriptions of the grid execution environment will enable more robust workflow execution and intelligent replanning and resource management.

Images are a crucial ingredient of scientific analysis. To understand physical systems and their structures, scientists need to access, process, integrate, and visualize images that provide alternative perspectives or are collected from different instruments. Grid computing and digital libraries provide basic mechanisms to store, index, manage, and access very large distributed collections of heterogeneous image data. In coming years, we should see a significant surge in the use of AI for representation and reasoning about these repositories' contents, for mining and discovery of novel relations, and for advanced management techniques. Jane Hunter, John Drennan, and Suzanne Little's article "Real-

izing the Hydrogen Economy through Semantic Web Technologies" exemplifies the need for image analysis and visualization techniques to support scientific and engineering advances. It describes a project that uses Semantic Web technology to develop novel hydrogen-based energy sources.

In "Semantics and Knowledge Grids: Building the Next-Generation Grid," Mario Cannataro and Domenico Talia look at future prospects of using knowledge to support the overall scientific task. Knowledge discovery and data mining, ontology-driven organization of grid-related knowledge, and intelligent data exploration and visualization will be part of future advanced applications of grid-computing infrastructure for science.

In "Semantic Services for Grid-Based, Large-Scale Science," William Johnston looks into the requirements envisioned by science research and illustrates factors contributing to the complexity of future large-scale scientific endeavors. Drawing from studies in climate modeling and high-energy physics, he emphasizes the challenges presented by the size of the data to be analyzed, the size and complexity of the data process-

ing and simulations involved, the growing diversity of required resources (physical instruments, computing resources, storage, and people), and scarcity of resources in a very dynamic distributed setting.

Scientists are choosing a high road paved with exciting and ever-challenging questions. AI research has many exciting opportunities to contribute to fundamental advances in all scientific disciplines. ■

References

1. R.K. Lindsay et al., *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project*, McGraw-Hill, 1980.
2. U.M. Fayyad et al., "Automated Analysis and Exploration of Image Databases: Results, Progress, and Challenges," *J. Intelligent Information Systems*, vol. 4, no. 1, Jan. 1995, pp. 7–25.
3. E. Rivas and S.R. Eddy, "The Language of RNA: A Formal Grammar That Includes Pseudoknots," *Bioinformatics*, vol. 16, no. 4, Apr. 2000, pp. 334–340.

The Authors



David De Roure is a professor of computer science at the University of Southampton, where he is head of Grid and Pervasive Computing in the School of Electronics and Computer Science and codirector of the Southampton Regional e-Science Center. His interests include large-scale distributed systems and the relationship between semantic, pervasive, and grid computing. He received his PhD in distributed systems from the University of Southampton. Contact him at the School of Electronics and Computer Science, Univ. of Southampton, Southampton, SO17 1BJ, UK; dder@ecs.soton.ac.uk.



Yolanda Gil is an associate division director at the University of Southern California's Information Sciences Institute and a research associate professor in the university's Computer Science Department. She's the principal investigator of several projects and conducts research in interactive knowledge capture, intelligent user interfaces, planning, and knowledge-intensive applications in defense and sciences. She received her PhD in computer science from Carnegie Mellon University. Contact her at USC/ISI, 4676 Admiralty Way, Marina del Rey, CA 90292; gil@isi.edu.



James A. Hendler is a professor of computer science at the University of Maryland and the Director of Semantic Web and Agent Technologies at the Maryland Information and Network Dynamics Laboratory. He received his PhD in artificial intelligence from Brown University. He's a fellow of the AAAI, the former Chief Scientist for Information Systems at DARPA, and the cochair of the Web Ontology Working Group for the World Wide Web Consortium. Contact him at the Dept. of Computer Science, Univ. of Maryland, College Park, MD 20742; hendler@cs.umd.edu; www.cs.umd.edu/~hendler.