

## Knowledge-Driven Infrastructure: Towards Intelligent Science Assistants

Yolanda Gill, Director of Knowledge Technologies and Associate Division Director  
Information Sciences Institute  
University of Southern California  
Los Angeles, California, USA

Advances in computing have enabled a data-centered revolution in science that has been significant, and yet I believe it will be dwarfed by what is ahead. We have seen high-end computing and distributed computing push the envelope in the scale of the phenomena that can be studied, leading to many many success stories of this kind of data-intensive computing in many areas of science.

In this talk, I will argue that we are beginning to see a second revolution enabled by knowledge-driven infrastructure that brings knowledge to the forefront. In scientific practice, data is augmented by knowledge that is scattered in papers, lab notebooks, emails, presentations, etc. Scientists are beginning to see beyond these informal textual forms and creating more declarative knowledge representations that enable more efficient communication among researchers and improve reproducibility. This includes semantic characterizations of data and software, process representations as workflows, and ontologies and metadata standards. This trend towards knowledge-driven infrastructure is transforming the way science is done. But there is a more profound innovation in this knowledge-driven infrastructure. By making scientific knowledge explicit, it becomes machine readable. Knowledge-driven infrastructure enables the development of Intelligent Science Assistants that partake in scientific processes, relieving scientists from tasks that are routine and facilitating tracking of experimental explorations and model revisions. I will present our work to date on knowledge-driven infrastructure, including intelligent workflow systems, constraint-based software characterization, and social knowledge collection.

I will also introduce the nascent discipline of Discovery Informatics that is catalyzing relevant research in artificial intelligence, visualization, data analytics, and social computing with the goal of improving and innovating science processes to accelerate discoveries.