Letter from the Special Issue Editor

The explosion in the volume of digital data and its wide availability is revolutionizing many scientific domains. At the same time, scientists faced with this data deluge must overcome many challenges to manage and explore these data. Complex processes are needed to acquire, process, and analyze the data. Even through there are robust and efficient databases systems, they fail to meet many of the requirements of emerging scientific applications which involve diverse data and require operations that go beyond what is currently supported. These new users and applications present new research problems in data management as well as a great opportunity for our community to have practical impact.

In this issue, we have collected a set of articles that highlight new directions for database research; relate limitations in current data management technology; and provide examples of how database research has been successfully applied to scientific problems in different domains, including neuroscience, astronomy, ecology. Motivated by needs in simulation sciences, Heinis et al. present a series of techniques they have developed to enable the construction and analysis of bigger and more detailed spatial models. They discuss the application of these techniques to real neuroscience datasets and show that they obtain a considerable performance improvement over the state of the art. In the context of a collaboration between astronomers and database experts, Vandeplas et al. are working on a platform for processing next-generation telescope image collections. They aim to process large volumes of sky images and allow questions to be efficiently answered over these data. Their paper describes the architecture of AscotDB, the system they have built, as well as techniques they have developed to address performance and usability issues. Stonebraker et al. presents an overview of the scientific database research at M.I.T, and summarize their work on making SciDB elastic, providing skew-aware join strategies, and producing scalable visualizations of scientific data. Alex Szalay considers the problem of analyzing very large simulation data, which are becoming increasingly harder to access, analyze and visualize. To allow broader usage of these data, he posits that analyses and visualizations must move to where the data resides and discusses the challenges in creating such interactive laboratories. Talbert et al. discuss challenges faced by ecologists in the field of species distribution modeling. Because of the scale of the data and the number of different models that are available, analyses are complex and require computationally-intensive sensitivity analysis accounting for various sources of uncertainty. While there exists technology to support these analyses, they are out of reach for many scientists who do not have a computer science background. To address this problem, Talbert et al. propose solutions that make use of scientific workflow systems. The last paper in this special issue examines the problem of reproducibility in science. While reproducibility is essential in this era of data-intensive science, the practice has not been widely adopted. One reason that is often cited is the fact that creating reproducible experiments is hard and time consuming. Chirigati et al. posit that this due in part to the lack of appropriate tools that support the tasks required for reproducibility. Besides characterizing the key tasks involved in the lifecycle of reproducible experiments, they propose a computational reproducibility benchmark. The benchmark aims to provide a means to categorize existing tools and better understand the features they support and how well they are supported.

These papers underscore importance of cross-domain synergies. They provide concrete examples of how database research has benefitted different scientific domains and how new research questions can be derived based on the needs of other areas. They also give evidence that data management is an essential component of science, and that our community has many challenging and significant problems to tackle.

I would like to thank all of the authors who agreed to share their work and experiences, as well as Dave Lomet who has provided invaluable guidance during the process of putting this issue together.

<div align="right">

Juliana Freire
New York University
New York, New York

</div>