As the sciences are becoming more and more data rich, most scientists are facing the challenges of dealing with ever increasing volumes of data in their research. It is generally agreed that tools are needed for modeling and organizing large amounts of scientific data for efficient access and easy querying that will work for domain scientists. While the database and data mining tools/techniques seem to be a perfect match for managing and analyzing scientific data, building such tools to help scientists requires working collaborations between domain scientists and computer scientists, more specifically database and data mining researchers, and involves several challenges. This issue of Data Engineering Bulletin explores the potential and challenges for database research to help other sciences in this data-rich era, and presents a spectrum of database techniques and approaches for scientific applications.

In the first paper, Jim Gray and Alex Szalay discuss the potential and the need for "Bridging the Gap between Databases and Science" in general. They also discuss what is needed for making such a collaborative research be successful based on their experiences in building the World-Wide Telescope and the Sky Server Project. As more powerful telescopes are being built, more data is being generated, and the volume of data to be processed in Astronomy is reaching challenging proportions. The next paper also describes a collaborative research by database and astronomy researchers. Calimlin, Cordes, Demers, Deneva, Gehrke, Kifer, Riedewald, and Shanmugasundaram describe their initial approach and the challenges in building a data management and data mining system for astronomy data for the Arecibo Telescope.

Data volume and workload requirements make scientific databases as good candidates for automated physical design techniques. Papadomanolakis and Alimaki discuss database partitioning techniques by utilizing the workload information, and present an automatic schema design tool for partitioning database tables. They evaluate this approach experimentally using the astronomy data from the Sky Server project. Howe and Maier advocate in situ processing of large scientific data repositories as a means of harnessing database techniques. They argue that using a data model interface by superimposing a specialized data model over native data repositories can provide the benefits of using a database management system without the substantial cost of massive bulk loading and reloading that come with the database deployment.

In many scientific database applications, similarity search is either commonly used by itself, or as a building block (e.g., clustering, phylogeny construction). Kriegel, Pfeifle and Schonauer discuss similarity search, and the utilization of a set-based similarity measure in two diverse applications in science and engineering domains. Data provenance (or data lineage) is another very important problem in scientific databases. Supporting data provenance in bio-scientific database applications is becoming more important with the explosion of curated databases in molecular biology and genomics. However, there is relatively few previous work on tracing the lineage of data in scientific databases. Wang-Chiew Tan, in her paper, gives motivation and an overview of research in this important subject, and presents a prototype for tracing the provenance and flow of relational data.

Graphs are very useful in modeling data in scientific problems and applications. The last two papers of the issue are on graph databases. M. Ozsoyoglu, Nadeau, G. Ozsoyoglu, and Tasan present an integrated set of tools for querying, browsing, and visualizing metabolic pathways in multiple abstraction levels, where pathways are conceptually modeled as a graph structured database. Finally, Stephens, Rung and Lopez briefly describe support for graph data in a commercial database management system.

I hope you will enjoy this issue, and find it useful and interesting. My special thanks to all the authors for their contributions, and Wanhong Xu and Murat Tasan for their assistance in the assembly of this special issue of the Bulletin.