

The explosion in the volume of digital data and its wide availability is revolutionizing many domains. While science is becoming increasingly data intensive, in industry, data has become a commodity — a key means to attain efficiency and generate revenue. This explosion has also democratized data access. Large-scale data analysis and management have historically been carried out in silos with the assistance of highly-trained database experts. Now, masses of data enthusiasts, from scientists to journalists with no knowledge of databases (or the means to pay for experts), are faced with complex analysis tasks and the need to manage data. Even though there are robust and efficient databases management systems, these are hard to configure and use and thus, out-of-reach for data enthusiasts. The problem is compounded due to the fact that complex analyses often involve data that are diverse and require operations that go beyond what is supported by relational databases or warehouses. Consequently, users need to weave complex workflows that combine multiple tools — a task that is difficult even for experts. These new users and applications present new challenges to database research 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 that addresses some of these challenges; examples of how database research has been successfully applied to problems that involve large-scale unstructured and structured data in different domains; and articles that relate limitations in current database technology from the point of view of users of the technology. Halevy and McGregor discuss challenges data journalists face and how tools such as Google Fusion Tables, that combine data management and visualization, have given journalists more power to discover and tell stories. Howe and Halperin argue that, in spite of numerous success stories, database systems are underused in the long tail of science, where spreadsheets abound. They describe SQLShare, a web-based query-as-a-service system developed at the University of Washington, that departs from the conventional database design, instead emphasizing a simple Upload-Query-Share workflow and exposing a direct, full-SQL query interface over raw tabular data. They relate real use cases for their system and how it has improved data analysis and management in different scientific domains. Fekete and Silva examine the interplay between databases and visual analytics. They observe that visualization researchers often re-implement database functionality in their tools in a sub-optimal way and discuss a number of data management services and features that are needed for visual analytics. They also posit that research in the intersection of databases and visual analytics would enrich both fields. Li and Jagadish explore the interaction between human computer interaction and databases, and discuss both lessons the database community can learn as well as new research challenges that arise for database usability. Weikum et al. give an overview of scalable techniques for data and text analytics that have been successfully applied in computational linguistics. They leverage big data on the Web and other resources to enhance semantics-centric tasks dealing with natural language texts. Kvilekval et al. discuss the challenges involved in managing biological images and how Bisque, an image database platform they have developed, addresses the limitations of traditional databases and standalone analysis tools for managing and exploring image collections.

There are two recurring themes in these papers. First, the importance of cross-domain synergies. Not only can database research benefit other areas, but we can also learn from and derive new research questions based on the needs of other areas, both within computer science and across different scientific domains. Second, these papers also highlight the need to go beyond traditional database systems and to make database technology more accessible—both easier to use for end-users and easier to integrate with other systems. I hope this special issue will energize discussions around these themes.

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.