As part of the Internet era, most of today's data-management systems need to effectively cope with a networked world. The recent proliferation of sensor networks and Peer-to-Peer (P2P) applications for a variety of data collection and sharing tasks only emphasizes the above thesis, further blurring the traditional boundaries between networking and data management. The efficient operation of the Internet itself relies on data-monitoring applications that continuously gather operational data from numerous points in large ISP networks – clearly, effectively managing the underlying infrastructure and minimizing the monitoring burden on the production network is an important concern here. Compared to conventional distributed databases, this new era of networked data-management systems poses a host of new research challenges, owing to (a) the sheer scale and dynamic nature of distribution (e.g., P2P systems supporting thousands of volatile peers), (b) the large volume of massive, rapid-rate data streams (e.g., large ISPs collecting hundreds of GBs of monitoring data each day), and (c) specific data-source characteristics and limitations (e.g., failure-prone and resource-limited sensor devices). To effectively support a high-level query-processing interface over such systems, it is crucial to intelligently push computation *inside the network*, moving query-processing toward the origin of the data. This requirement represents a major departure from traditional databases, placing issues like routing efficiency, load balancing, and intelligent scheduling of communication at the core of effective query-processing strategies.

The collection of articles in this special issue is only a small sample of different research perspectives and efforts in the vibrant area of *in-network query processing*. The first article by Hellerstein et al. outlines an interesting research vision for a "Network Oracle", a global, real-time monitoring and information infrastructure for the Internet's core state. The authors offer convincing arguments on the importance, timeliness, and feasibility of such an infrastructure, and outline a high-level research agenda with several key challenges for our community. When dealing with distributed data collection and monitoring, communication efficiency is an important concern (e.g., in sensornets where communication is the major source of sensor battery drain). The next four articles discuss different approaches for improving communication efficiency through *approximation*. First, Olston and Widom give an overview of their recent work on approximate replication, an important technique for balancing distributed replica precision against communication efficiency. Second, Deligiannakis and Kotidis discuss data-reduction mechanisms (based on linear regression) and approximate-aggregation techniques for limiting communication overheads in sensornet environments. Third, Kollios et al. propose duplicate-insensitive data synopses for communication-efficient aggregation in sensornets that use broadcast or multi-path routing to improve robustness in the face of failures. Finally, Cormode and Garofalakis outline a general framework and algorithmic approach for continuous, approximate query tracking in distributed settings.

The article by Deshpande et al. outlines the design of their *Sensor Control System*, that brings together a number of interesting ideas from statistical models and control theory to enable efficient sensornet management. Shifting to the realm of P2P systems, the next two articles discuss challenges in effective query processing over large, widely-distributed, and dynamic collections of peers: Ganesan and Garcia-Molina propose novel algorithms for mapping data to peers in a manner that ensures efficient, dynamic load balancing and retains locality in the key space for efficient range and similarity queries (unlike conventional DHTs); Felber et al. propose a topology-aware DHT design for peer data management as well as novel indexing schemes for structured XML content and queries. Topology awareness is also the key theme in the article by Ahmad et al. that discusses the issues of query-operator placement and distribution for Internet-scale stream processors. Finally, Gnawali et al. discuss their experiences with implementing SQL functionality on top of a generic sensornet routing platform.

I truly hope that you will find this collection of articles as enjoyable and insightful as I did, and that this special issue will help stimulate new research ideas in this exciting area, at the intersection of networking and data management. My sincerest thanks to all the authors for their contributions.