

## Introduction to the Minitrack on Data Warehousing

Barbara Wixom  
McIntire School of Commerce  
The University of Virginia  
Charlottesville, VA 22903  
[bwixom@mindspring.com](mailto:bwixom@mindspring.com)

Paul Gray  
.School of Information Science  
Claremont Graduate University  
Claremont, CA 91711  
[paul.gray@cgu.edu](mailto:paul.gray@cgu.edu)

Hugh J. Watson  
Terry College of Business  
The University of Georgia  
Athens, GA 30602-6256  
[hwatson@arches.uga.edu](mailto:hwatson@arches.uga.edu)

Data warehousing continues to grow in importance. Surveys consistently rank it as one of the most important strategic initiatives in companies today. Customer relationship management, performance management, e-commerce, and connecting companies along the supply chain all rely on data warehousing. Data warehousing has moved from being an innovation in a few leading-edge companies, to being a requirement for doing business in a large number of firms.

This minitrack was created as a vehicle for researchers around the world to share their ideas and work. Over the years, the number of submissions and quality of the papers has continued to increase. The papers in this year's minitrack show the quality research that is being done.

The first three papers of this minitrack investigate organizational and managerial issues regarding data warehousing. The second group of papers examines technical facets of data warehousing, with a focus on improving the performance and effectiveness of data warehousing through better data integration, data modeling, and data virtualization.

In "Understanding How Metadata and Explanations Can Better Support Data Warehousing and Related Decision Support Systems: An Exploratory Case Study," Hess and Wells examine the topic of metadata used to support the decision making process. With data warehousing, users are asking questions that they previously could not ask, and they are thinking in new ways. Thus, they rely on metadata to understand the data access tools and the data itself that enable their decisions.

Metadata extends much farther past simple meanings and relationships that have been available to users in the operational world. To make decisions, users need to know information such as when the data were last refreshed, what sources provided the data, and how data were combined or manipulated. However, the effects of metadata and explanations on that support have not been documented. Hess and Wells undertake an exploratory case study at two organizations in different industries. The DSS-decision performance model suggested by Todd and Benbasat was employed as a framework, and constructs from Gregor and Benbasat's meta-analysis of explanation use were used to guide the investigation of the explanations provided by the data warehouse applications.

McGregor, Bryan, Curry, and Tracy also use the case study method in their paper "The e-Baby Data Warehouse: A Case Study." The primary objective of this study was to create a framework for clinical and psychological data capture, storage structures, and Internet data access for data warehousing in the healthcare industry. The organization at the heart of the case is the Neonatal Intensive Care Unit (NICU) At Nepean Hospital in Perth Australia. The case illustrates how a data warehouse infrastructure was put in place to support the immediate and demanding needs of a NICU. A valuable contribution of this paper is the description of the process that was used to put the infrastructure in place.

A third case study in this minitrack describes the data warehouse in place at Boeing Employee Credit Union

(BECU). O'Hara and Brohman focus on the managerial issues of a data warehouse that supports various customer relationship management applications in "Boeing Employees' Credit Union: Leveraging the Data Warehousing Investment." BECU has leveraged their \$2 million data warehouse investment by providing three distinct types of service: (1) information and decision support to BECU employees, (2) personalized information to members, and (3) information and decision support to other credit unions. This case study details these three types of service, the impact of each on BECU's performance, along with future directions and the lessons learned.

As applications for data warehousing, such as CRM, continue to grow, the number of data sources that an organization has to manage increases exponentially. Integrating this data is a growing problem and one that is addressed by Miller, Yu, and Nilakantra in "Integration of Relational Databases and Record-Based Legacy Systems for Populating Data Warehouses." Their research investigates the situation where data sources are restricted to relational databases and record-based legacy systems owned by the same organization. Specifically, the authors provide a mechanism for developing a relational model for a set of data sources, provide a method for generating correct queries over the model, and create an architecture for executing the queries based on the mobile agent paradigm. A prototype of the system has been designed and implemented.

A second issue of equal importance to data integration is creating an effective data model for a data warehouse. In "Data Modeling for the Data Warehouse: A Case Study," Sampson and Atkins explore how an understanding of the semantic content of a data model may impact on the effectiveness of the implemented data warehouse. A conceptual framework based on semiotics was devised to explain how a data model may be interpreted through the meaning levels of understanding, connotation and generation, and also how a data model is created from an existing meaning structure by intention, generation and action.

The final paper in the minitrack presents the principles of virtualization and several vendor options that incorporate its principles. Virtualization is the mechanism that allows the next generation online storage to be seamlessly integrated. It also allows cost sensitive installations to make use of alternative disk types to be used effectively without requiring complex installation management. Milligan and Selkirk in "Online Storage Virtualization: The Key to Managing the Data Explosion" describe virtualization as an important factor for the success of online data storage systems today and in the future.