

# **Warehousing The World: Challenges From New Types of Data**

Torben Bach Pedersen

Department of Computer Science, Aalborg University, Denmark  
tbp@cs.aau.dk  
<http://www.cs.aau.dk/~tbp>

## **Keynote Abstract**

Data warehouses (DWs) have become widely used and successful in many enterprises, by allowing the storage and analysis of large amounts of structured business data. DWs are based on a multidimensional data model, where important business events, e.g., sales, are modeled as facts, characterized by a number of hierarchical dimensions, e.g., time and products, with associated numerical measures, e.g., sales price. The multidimensional model is unique in providing a framework that is both intuitive and efficient, allowing data to be viewed and analyzed at the desired level of detail with excellent performance. Traditional data warehouses have worked very well for traditional, so-called structured data, but recently enterprises have become aware that DWs are in fact only solving a small part of their real integration and analysis needs.

Already today, many different types of data are found in most enterprises, including structured, relational data, multidimensional data in DWs, text data in documents, emails, and web pages, and semi-structured/XML data such as electronic catalogs. Based on current developments within mobile, pervasive and ubiquitous computing, most enterprises will also have to manage large quantities of geo-related data, as well as data from a large amount of sensors. Finally, many analytical models of data have been developed through data mining.

The main problem with current technologies is that all these different types of data/models cannot be integrated and analyzed in a coherent fashion. Instead, applications must develop separate ad-hoc solutions for integration and analysis, typically for each pair of data types, e.g., relational and text. This obviously is both expensive and error-prone. Additionally, privacy protection is often given low priority. This situation inspires the vision of developing a breakthrough set of technologies that extend the benefits of DWs to a much wider range of data, making it feasible to literally "warehouse the world". To do this, five unique challenges must be addressed:

1. Warehousing data about the physical world
2. Integrating structured, semi-structured, and unstructured data in DWs
3. Integrating the past, the present, and the future
4. Warehousing imperfect data