

# The Future Home of Data

Johann-Christoph Freytag

Institut für Informatik

Humboldt-Universität zu Berlin

Germany

[freytag@dbis.informatik.hu-berlin.de](mailto:freytag@dbis.informatik.hu-berlin.de)

## 1. Overview

Over the last year the question of how and where to store data and how to access it has become pressing issue. Especially with the growing importance of E-Commerce and the widely used Internet access it is not clear any more if one approach satisfies all the needs of different communities accessing and processing data.

XML currently seems to be the most favored form for presenting, exchanging and possibly storing data. New technology “waves” such as Web Services, an “interaction model” between businesses and customers (B2C) or among business themselves (B2B) heavily rely on XML data because semantic information can be intertwined with the data. Many companies in the E-Commerce space assume XML to be the “universal data format” for the future when storing and accessing data.

Database management systems (DBMSs) have been the “modern” technology for the last 40 years (with different data models and different levels of functionality and sophistication) to store, to access and to manipulate large amounts of data. This technology comes with many properties embedded in subsystems of DBMSs that is essential for reliable, efficient information processing in the business world, such as transaction management and query processing. Over the years, DBMSs have changed their assumptions where data might reside for access and processing. The initial assumption that all data is stored centrally changed to a “distributed model” by assuming that data is spread over several locations. The latter model quickly extended into a federated one assuming

---

*Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment*

**Proceedings of the 28<sup>th</sup> VLDB Conference,  
Hong Kong, China, 2002**

that “data sources” might be autonomous and not always under the control of one (database) system at the same time acknowledging that data might come in different formats.

Despite the “new technology waves” one has to acknowledge that data is still stored as “flat files” without structural information or other relevant “meta data” that might be important for efficient access and correct processing.

Besides the form of the data its location, i.e. where to find data that might be relevant and important for a user to perform a particular task, has become an important issue. Having many devices such as mobile phones, Notebooks, PDAs and other mobile devices, those often need data from other sources, at the same time storing new/additional data that might be relevant for others to perform their tasks successfully. Some devices, such as mobile phone, exist in large numbers; they perform tasks different from a computing device, still their capabilities as storage processing devices (limited capabilities) make them an important data generation source and storage device that must be included into current and future processing environments.

Panel members from industry and academia have been asked to address the following issues in the statements:

- What are the different alternatives for storing and accessing data for the future?
- What are the characteristics for storage now and in the future?
- Is XML the universal answer to all future needs, does it take over the “world of data”?
- Should existing DBMSs be thrown away?
- Should existing DBMSs be changed to adapt to new requirements and challenges?
- Are XML-DBMS the answer to all needs?
- Is the role/functionality of DBMSs changing?

- How do we deal with the many (mobile) devices as devices for data storage and data processing?
- What are the “right” assumptions about distribution and heterogeneity of data?
- What is the industrial and/or academic approach to deal with this challenge?
- What are the trade-offs between different approaches?

## **2. Panelists**

- Mike Franklin, Univ. of Berkeley, CA, U.S.A.
- Guido Moerkotte, Univ. Mannheim, Germany
- Guy Lohman, IBM Almaden Research, U.S.A.
- Paul Larson, Microsoft Research, U.S.A.
- N.N.