

# Proceedings

## 18<sup>th</sup> International Conference on Scientific and Statistical Database Management

### SSDBM 2006

3 – 5 July, 2006  
Vienna, Austria



Los Alamitos, California

Washington • Tokyo

---

All rights reserved.

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

*The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.*

IEEE Computer Society Order Number P2590

ISBN 0-7695-2590-3

ISBN 978-0-7695-2590-7

ISSN Number 1551-6393

*Additional copies may be ordered from:*

IEEE Computer Society  
Customer Service Center  
10662 Los Vaqueros Circle  
P.O. Box 3014  
Los Alamitos, CA 90720-1314  
Tel: + 1 800 272 6657  
Fax: + 1 714 821 4641  
<http://computer.org/cspress>  
[csbooks@computer.org](mailto:csbooks@computer.org)

IEEE Service Center  
445 Hoes Lane  
P.O. Box 1331  
Piscataway, NJ 08855-1331  
Tel: + 1 732 981 0060  
Fax: + 1 732 981 9667  
[http://shop.ieee.org/store/  
customer-service@ieee.org](http://shop.ieee.org/store/customer-service@ieee.org)

IEEE Computer Society  
Asia/Pacific Office  
Watanabe Bldg., 1-4-2  
Minami-Aoyama  
Minato-ku, Tokyo 107-0062  
JAPAN  
Tel: + 81 3 3408 3118  
Fax: + 81 3 3408 3553  
[tokyo.ofc@computer.org](mailto:tokyo.ofc@computer.org)

*Individual paper REPRINTS may be ordered at: <[reprints@computer.org](mailto:reprints@computer.org)>*

Editorial production by Bob Werner  
Cover art production by Joe Daigle/Studio Productions  
Printed in the United States of America



IEEE Computer Society  
**Conference Publishing Services**  
<http://www.computer.org/proceedings/>

# Table of Contents

## 18<sup>th</sup> International Conference on Scientific and Statistical Database Management SSDBM 2006

<b>Foreword</b> .....	ix
<b>Committees</b> .....	x
<b>Reviewers</b> .....	xi

### **Session 1: Scientific Database Management I**

Containment of Partially Specified Tree-Pattern Queries.....	3
<i>Dimitri Theodoratos, Theodore Dalamagas, Pawel Placek, Stefanos Souldatos, and Timos Sellis</i>	
InterJoin: Exploiting Indexes and Materialized Views in XPath Evaluation.....	13
<i>Derek Phillips, Ning Zhang, Ihab Ilyas, and Tamer Özsu</i>	
The Holodex: Integrating Summarization with the Index .....	23
<i>Hung-chih Yang, Stott Parker, and Ruey-Lung Hsiao</i>	
Minimizing I/O Costs of Multi-Dimensional Queries with Bitmap Indices .....	33
<i>Doron Rotem, Kurt Stockinger, and Kesheng Wu</i>	

### **Keynote**

Data Management in Production Grids — Challenges and Techniques .....	45
<i>Erwin Laure</i>	

### **Session 2: Experiment Data Management I**

Data-Preservation in Scientific Workflow Middleware.....	49
<i>David Liu, Ghaleb Abdulla, Michael Franklin, Jim Garlick, and Marcus Miller</i>	
Efficient Data Harvesting for Tracing Phenomena in Sensor Networks .....	59
<i>Adesola Omotayo, Moustafa Hammad, and Ken Barker</i>	

## Session 3: Experiment Data Management II

A Framework for Querying Pedigree Data .....	71
<i>B. Elliott, S.F. Akgul, Z.M. Ozsoyoglu, and E. Manilich</i>	
Task-Oriented Integrated Use of Biological Web Data Sources .....	81
<i>Mustafa Kirac, Ali Cakmak, and Gultekin Ozsoyoglu</i>	
Scalability Management in Sensor-Network PhenomenaBases .....	91
<i>M.H. Ali, Walid Aref, and Ibrahim Kamel</i>	
The GOBASE: An Information Management System for Gene Ontology.....	101
<i>Ruey-Lung Hsiao and Stott Parker</i>	
A Demonstration of the SciFlo Grid Workflow Engine .....	105
<i>Brian Wilson, Gerald Manipon, Dominic Mazzoni, Benyang Tang, and Tom Yunck</i>	

## Session 4: Data Mining and Visualization

Mining Frequent 3D Sequential Patterns.....	109
<i>Zhenqiang Tan and Anthony Tung</i>	
Mining Hierarchies of Correlation Clusters.....	119
<i>Elke Aichert, Christian Böhm, Peer Kröger, and Arthur Zimek</i>	
PhyloMiner: A Tool for Evolutionary Data Analysis.....	129
<i>Sen Zhang, Katherine Herbert, Jason Wang, William Piel, and David Stockwell</i>	
On Data and Visualization Models for Signaling Pathways.....	133
<i>Nattakarn Ratprasartporn, Ali Cakmak, and Gultekin Ozsoyoglu</i>	
SQL.CT: Providing Data Management for Visual Exploration of CT Datasets .....	143
<i>Shirley Cohen and David Guzman</i>	

## Session 5: Scientific Database Management II

HDF5-FastQuery: Accelerating Complex Queries on HDF Datasets Using Fast Bitmap Indices.....	149
<i>Luke Gosink, John Shalf, Kurt Stockinger, Kesheng Wu, and Wes Bethel</i>	
Materialized Views in the Presence of Reporting Functions.....	159
<i>Dirk Habich, Wolfgang Lehner, and Michael Just</i>	
Probabilistic Ranking Queries on Gaussians .....	169
<i>Christian Böhm, Alexey Pryakhin, and Matthias Schubert</i>	
Efficient Storage Allocation of Large-Scale Extendible Multi-dimensional Scientific Datasets .....	179
<i>Ekow Otoo and Doron Rotem</i>	
Efficient Query Processing in Arbitrary Subspaces Using Vector Approximations.....	184
<i>Hans-Peter Kriegel, Peer Kröger, Matthias Schubert, and Ziyue Zhu</i>	

## Keynote

Reuseable Tools for Metadata Management — Pie in the Sky?.....	191
<i>Sergey Melnik</i>	

## Session 6: Data Provenance

How to Cite Curated Databases and How to make them Citable .....	195
<i>Peter Buneman</i>	
Provenance and Reliability: Managing Metadata for Statistical Models.....	204
<i>Andrew Westlake</i>	

## Session 7: Ontologies and Metadata Management

Semantic Map of Services for Structural Bioinformatics .....	217
<i>Pierre Tufféry, Zoé Lacroix, and Hervé Ménager</i>	
BioFacets: Faceted Classification for Biological Information.....	225
<i>M. Mahoui, Z. Miled, A. Godse, H. Kulkarni, and N. Li</i>	
Edits — Data Cleansing at the Data Entry to Assert Semantic Consistency of Metric Data.....	235
<i>Hans-J. Lenz, Veit Köppen, and Roland Müller</i>	

## Session 8: Data Streams and Time Series

Efficient Methods on Predictions for Similarity Search over Stream Time Series.....	241
<i>Xiang Lian and Lei Chen</i>	
Time Series Analysis Using the Concept of Adaptable Threshold Similarity .....	251
<i>Johannes Assfalg, Hans-Peter Kriegel, Peer Kröger, Peter Kunath, Alexey Pryakhin, and Matthias Renz</i>	
Exploring Data Streams with Nonparametric Estimators .....	261
<i>Christopher Heinz and Bernhard Seeger</i>	
Indexing Time-Evolving Data with Variable Lifetimes .....	265
<i>Lukasz Golab, Piyush Prahladka, and M. Tamer Özsu</i>	
Sampling Trajectory Streams with Spatiotemporal Criteria .....	275
<i>Michalis Potamias, Kostas Patroumpas, and Timos Sellis</i>	
An Extensible Infrastructure for Processing Distributed Geospatial Data Streams.....	285
<i>Carlos Rueda, Michael Gertz, Bertram Ludäscher, and Bernd Hamann</i>	

## Session 9: Conceptual Models and OLAP

The Composite OLAP-Object Data Model: Removing an Unnecessary Barrier.....	291
<i>Elaheh Pourabbas and Arie Shoshani</i>	
Accuracy Control in Compressed Multidimensional Data Cubes for Quality of Answer-Based OLAP Tools.....	301
<i>Alfredo Cuzzocrea</i>	
Optimistic Coarse-Grained Cache Semantics for Data Marts.....	311
<i>Maik Thiele, Albrecht and Wolfgang Lehner</i>	
A Disc-Based Approach to Data Summarization and Privacy Preservation.....	321
<i>Rong Ge, Martin Ester, Wen Jin, and Zengjian Hu</i>	

## Session 10: Scientific Database Management III

On-line Aggregation and Filtering of Pattern-Based Queries.....	333
<i>Cédric du Mouza, Phillipe Rigaux, and Michel Scholl</i>	
ACache: Using Caching to Improve the Performance of Multiple Sequence Alignments.....	343
<i>Xun Tu, Kajal Claypool, and Cindy Chen</i>	
Progressive Spatial Join.....	353
<i>Wee Hyong Tok, Stéphane Bressan, and Mong Li Lee</i>	

## Session 11: Algorithms

A Dynamic Mobility Histogram Construction Method Based on Markov Chains.....	359
<i>Yoshiharu Ishikawa, Yoji Machida, and Hiroyuki Kitagawa</i>	
Dynamic Travel Time Maps — Enabling Efficient Navigation.....	369
<i>Dieter Pfoser, Nectaria Tryfona, and Agnès Voisard</i>	
Addressing the Need for Map-Matching Speed: Localizing Global Curve-Matching Algorithms.....	379
<i>Carola Wenk, Randall Salas, and Dieter Pfoser</i>	
Inferences on Kernel Density Estimates by Solving Nonlinear Systems.....	389
<i>Stefania Bellavia, Stefano Lodi, and Benedetta Morini</i>	

<b>Author Index</b> .....	398
---------------------------	-----