# TABLE OF CONTENTS

1. SIGMOD Officers, Committees, and Awardees

3. Editor’s Notes

5. ACM TODS EIC Selection (M. Tamer Ozsu)

**Articles**

7. Database Research Opportunities in Computer Games (Walker White, Christoph Koch, Nitin Gupta, Johannes Gehrke, and Alan Demers)

15. Simple off the shelf abstractions for XML Schema (Wim Martens, Frank Neven, and Thomas Schwentick)

**Surveys** (Cesar Galindo-Legaria, editor)

23. Overview and Semantic Issues of Text Mining (Anna Stavrianou, Periklis Andritsos, and Nicolas Nicoloyannis)

**Distinguished Profiles in Data Management** (Marianne Winslett, editor)

35. Kyu-Young Whang Speaks Out

41. Boon Thau Loo Speaks Out

**Research Centers** (Ugur Cetintemel, editor)

47. Community Systems Research at Yahoo!

**Reports** (Brian Cooper, editor)

55. Report on the First International Workshop on Database Preservation (PresDB’07) (Vassilis Christophides and Peter Buneman)
SIGMOD Record

SIGMOD Record is a quarterly publication of the Special Interest Group on Management of Data (SIGMOD) of the Association for Computing Machinery (ACM). SIGMOD is dedicated to the study, development, and application of database and information technology. SIGMOD Record Web Edition is also freely available online at http://www.sigmod.org/record.

SIGMOD Record solicits contributions of articles, technical notes, reports, and proposals for special sections. Conference announcements and calls for papers are published if relevant to the interests of the group and, in most cases, are limited to one page. Submitted technical papers are reviewed for importance and correctness. Priority is given to papers that deal with current issues of interest to a broad audience.

Papers should be submitted electronically in PDF format through the SIGMOD REcord Electronic Submission System (RECESS) at http://db.cs.pitt.edu/recess and they should follow a format similar to that of the SIGMOD conference proceedings (but with a larger font): 10 point font, single-spaced, 2-column, 8.5" by 11" page size with 1" margins all around and no page numbers. They should also be formatted for letter size pages and must contain all fonts embedded. Submitted articles are limited to 6 pages unless prior agreement with the Editor.

By submitting your article for distribution in this Special Interest Group publication, you hereby grant to ACM the following non-exclusive, perpetual, worldwide rights: 1) to publish in print on condition of acceptance by the editor; 2) to digitize and post your article in the electronic version of this publication; 3) to include the article in the ACM Digital Library; and 4) to allow users to copy and distribute the article for noncommercial, educational or research purposes. However, as a contributing author, you retain copyright to your article and ACM will make every effort to refer requests for commercial use directly to you. Therefore, ACM is asking all newsletter authors to include their contact information in their submissions. Opinions expressed in articles and letters are those of the author(s) and do not necessarily express the opinions of the ACM or SIGMOD. Author(s) should be contacted for reprint authorization.

SIGMOD Record Editor:
Alexandros Labrinidis
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260-9161, USA
<labrinid AT cs.pitt.edu>

Associate Editors:
- Bagdalena Balazinska, University of Washington (Systems and Prototypes), <magda AT cs.washington.edu>
- Denilson Barbosa, University of Calgary (Web Edition), <denilson AT ucalgary.ca>
- Ugur Çetintemel (Research Centers), <ugur AT cs.brown.edu>
- Brian Cooper, Yahoo! (Reports), <cooperb AT yahoo-inc.com>
- Andrew Eisenberg, IBM Corporation (Standards), <andrew.eisenberg AT us.ibm.com>
- Cesar Galindo-Legaria, Microsoft Research (Research Surveys), <cesarg AT microsoft.com>
- Leonid Libkin, University of Edinburgh (Database Principles), <libkin AT inf.ed.ac.uk>
- Jim Melton, Oracle Corporation (Standards), <jim.melton AT acm.org>
- Len Seligman, The MITRE Corporation (Industry Perspectives), <seligman AT mitre.org>
- Marianne Winslett, University of Illinois (Distinguished Profiles in Data Management), <winslett AT cs.uiuc.edu>

SIGMOD Record (ISSN 0163-5808) is published quarterly by the Association for Computing Machinery, 2 Penn Plaza, Suite 701, New York, NY 10121-0701. Periodicals postage paid at New York, NY 10001, and at additional mailing offices.

POSTMASTER: Send address changes to SIGMOD Record, ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701.
SIGMOD Officers, Committees, and Awardees

Chair
Raghu Ramakrishnan
Yahoo! Research
2821 Mission College
Santa Clara, CA 95054
USA
<First8CharsOfLastName AT yahoo-inc.com>

Vice-Chair
Yannis Ioannidis
University of Athens
Department of Informatics & Telecom
Panepistimiopolis, Informatics Buildings
157 84 Ilissia, Athens
HELLAS
<yannis AT di.uoa.gr>

Secretary/Treasurer
Mary Fernández
ATT Labs - Research
180 Park Ave., Bldg 103, E277
Florham Park, NJ 07932-0971
USA
<mff AT research.att.com>

SIGMOD Executive Committee:

Advisory Board: Tamer Özu (Chair), University of Waterloo, <tozu AT cs.uwaterloo.ca>, Rakesh Agrawal, Phil Bernstein, Peter Buneman, David DeWitt, Hector Garcia-Molina, Jim Gray, Masaru Kitsuregawa, Jiawei Han, Alberto Laender, Krithi Ramamritham, Hans-Jörg Schek, Rick Snodgrass, and Gerhard Weikum.

Information Director:
Jeffrey Xu Yu, The Chinese University of Hong Kong, <yu AT se.cuhk.edu.hk>

Associate Information Directors:

SIGMOD Record Editor:
Alexandros Labrinidis, University of Pittsburgh, <labrinid AT cs.pitt.edu>

SIGMOD Record Associate Editors:
Magdalena Balazinska, Denilson Barbosa, Ugur Çetintemel, Brian Cooper, Andrew Eisenberg, Cesar Galindo-Legaria, Leonid Libkin, Jim Melton, Len Seligman, and Marianne Winslett.

SIGMOD Disc Editor:
Joachim Hammer, Microsoft Research, <Joachim.Hammer AT microsoft.com>

SIGMOD Anthology Editor:
Curtis Dyreson, Washington State University, <cdyreson AT eecs.wsu.edu>

SIGMOD Conference Coordinators:
Jianwen Su, UC Santa Barbara, <su AT cs.ucsb.edu>,
Lisa Singh, Georgetown University, <singh AT cs.georgetown.edu>

PODS Executive: Phokion Kolaitis (Chair), IBM Almaden, <kolaitis AT almaden.ibm.com>,
Foto Afrati, Catriel Beeri, Georg Gottlob, Leonid Libkin, and Jan Van Den Bussche.

Sister Society Liaisons:
Raghu Ramakrishnan (SIGKDD), Yannis Ioannidis (EDBT Endowment).

Awards Committee: Serge Abiteboul (Chair), INRIA, <serge.abiteboul AT inria.fr>,
Mike Carey, David Maier, Moshe Y. Vardi, and Gerhard Weikum.
SIGMOD Officers, Committees, and Awardees (continued)

**SIGMOD Edgar F. Codd Innovations Award**

*For innovative and highly significant contributions of enduring value to the development, understanding, or use of database systems and databases.* Until 2003, this award was known as the “SIGMOD Innovations Award.” In 2004, SIGMOD, with the unanimous approval of ACM Council, decided to rename the award to honor Dr. E.F. (Ted) Codd (1923 - 2003) who invented the relational data model and was responsible for the significant development of the database field as a scientific discipline. Recipients of the award are the following:

Jennifer Widom (2007)

**SIGMOD Contributions Award**

*For significant contributions to the field of database systems through research funding, education, and professional services.* Recipients of the award are the following:

Hans-Jörg Schek (2007)

**SIGMOD Doctoral Dissertation Award**

The annual ACM SIGMOD Doctoral Dissertation Award, inaugurated in 2006, recognizes excellent research by doctoral candidates in the database field.

- **2006 Winner:** Gerome Miklau, University of Washington
  
  *Runners-up:* Marcelo Arenas, University of Toronto; Yanlei Diao, University of California at Berkeley.

- **2007 Winner:** Boon Thau Loo, University of California at Berkeley
  
  *Honorable Mentions:* Xifeng Yan, University of Illinois at Urbana-Champaign; Martin Theobald, Saarland University

A complete listing of all SIGMOD Awards is available at: [http://www.sigmod.org/awards/](http://www.sigmod.org/awards/)

[Last updated on October 12, 2007]
Editor’s Notes

Welcome to the September 2007 issue of SIGMOD Record. I am happy to report that we are catching up with backlog and our publication schedule should return to normal with the December 2007 issue.

We start this issue with an important message from M. Tamer Ozsu, who gives a behind-the-scenes account of the TODS Editor-in-Chief selection process. Next, we have an interesting article by Johannes Gehrke and his colleagues at Cornell University who are identifying opportunities towards a database research agenda for Computer Games. With CS enrollment at American universities in decline, this makes for a very interesting read.

The next article is a contribution to the Database Principles column (edited by Leonid Libkin) on the expressiveness and complexity of XML Schema (by Wim Martens, Frank Neven, Thomas Schwentick), presented in an easy and accessible way.

We continue with an article in the Surveys Column (edited by Cesar Galindo-Legaria), on Text Mining (by Stavrianou, Andritsos, and Nicoloyannis). The previous survey article was published in the June 2005 issue of SIGMOD Record; I am very happy to see the column revitalized again and feature timely contributions on exciting topics.

In this issue we have for the first time two interviews in the Distinguished Profiles in Data Management Column (formerly known as the Distinguished DB Profiles Column) by Marianne Winslett. The first interview, of Kyu-Young Whang (from KAIST), follows the “classic” style for entries in the column, by featuring distinguished senior members of the database community. Read Kyu-Young Whang’s interview to find out (among many other things) about Academia and Startups in Korea, and what KISS stands for.

The second interview is part of an effort to highlight junior database researchers and features this year’s ACM SIGMOD Dissertation Award winner, Boon Thau Loo (PhD from UC Berkeley, currently at UPenn). Read Boon Thau Loo’s interview to find out why Datalog is cool again and what it feels like to finish your first semester as an assistant professor.

We continue with an article in the Research Centers Column (edited by Ugur Cetintemel), about Community Systems Research at Yahoo! (by the members of the Community Systems Group). The article highlights some of the technologies developed by the group, with cool names such as PNUTS, Pig, AppForge, Purple SOX, etc.

Finally, the issue concludes with an event report (edited by Brian Cooper) on the First International Workshop on Database Preservation (PresDB07) which was held in March 2007 in Edinburgh, Scotland.

Alexandros Labrinidis
October 2007
ACM TODS EIC Selection

As you should have heard by now (there was a DBWORLD announcement), there is a new Editor-in-Chief (EIC) of ACM Transactions on Database Systems (TODS). Meral Ozsoyoglu of Case Western Reserve University has replaced Rick Snodgrass as TODS EIC as of September 17, 2007. In this short note, I want to give the SIGMOD community information about the process that was followed.

First let me provide some background. ACM headquarters is organized along a number of directorates; the ones relevant for SIGMOD and TODS are those of SIGs, and publications. For each of these, there are policy boards that establish the policies that headquarters staff follow. SIGMOD organization is on the SIG side and SIGMOD Chair is a member of SIG Governing Board, ACM TODS is on the publications side, as are all ACM publications. Therefore, although SIGMOD is an important stakeholder and for the purposes of distributing ACM Digital Library income is considered the “owner” of TODS, the appointment of TODS EIC is the responsibility of the ACM Publications Board.

The EIC selection process starts by the Publications Board selecting one of its members to oversee the process. The Publications Board member then forms a EIC Nominating Committee and obtains Publications Board approval for the committee. The Nominating Committee then gathers a list of possible candidates for the position, produces a short list, asks for vision statements and other supporting material from the short-listed candidates and makes a recommendation to the Publications Board. The appointment is finalized when the Publications Board approves the recommendation.

In this year’s process, I was the Publications Board member charged with setting up the Nominating Committee and coordinating its work. I also chaired the Nominating Committee that additionally included Phil Bernstein (Microsoft Research), Mary Fernandez (AT&T Labs – Research), Phokion Kolaitis (IBM Almaden Research Center), Krithi Ramamritham (IIT Bombay), and Gerhard Weikum (Max-Planc Institute for Informatics). We issued a public call for nominations through DBWORLD, and consulted a phone interview with Rick Snodgrass as out-going EIC, and an in-person interview during SIGMOD/PODS 2007 with Raghu Ramakrishnan as SIGMOD Chair. Thus, the relevant stakeholders were consulted.

As a result of these, the Nominating Committee gathered an initial list of ten nominees. After discussions, the Nominating Committee reduced this list to a short list of four. These colleagues were contacted to check their availability and all but one let their names stand. These three nominees were asked to supply statements responding to a set of questions that the Nominating Committee posed as well as their vision for TODS. We were very pleased with the depth and thoughtfulness of these statements; it was clear that all of the nominees had spent considerable time thinking about where TODS is and where they would like it to go. Interestingly, there were overlaps between the statements.

Making the final choice from among three very able colleagues was naturally difficult, but the Nominating Committee decided to recommend Meral Ozsoyoglu as the next EIC of TODS. The Publications Board approved this recommendation with enthusiasm and appointed Meral to a four year (renewable) term.
The Nominating Committee feels that the entire process worked very well. We appreciated the engagement of the community in supplying us with a number of possible candidates, and the willingness of both Rick and Raghu to provide us with suggestions and more candidates.

The Nominating Committee is excited by the leadership that Meral will bring to ACM TODS and we are certain that TODS will maintain its position as a preeminent publication venue under her leadership. We would also like to thank Rick Snodgrass for his remarkable 15 years of exemplary service to ACM TODS (the last eight years as EIC). Under his leadership, the journal has been a dynamic publication venue and the source of a number of experiments and initiatives.

M. Tamer Özsu
October, 2007