



# ICEIS 2007

## ***NINTH INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS***

# Proceedings

*Databases and Information Systems Integration*

FUNCHAL, PORTUGAL, June 12-16, 2007

SPONSORED BY

FCT - Fundação para a Ciência e Tecnologia

ORGANIZED BY



CO-ORGANIZED BY

University of Madeira

IN COOPERATION WITH

ACM/SIGMIS and AAAI

# ICEIS 2007

Proceedings of the  
Ninth International Conference on  
Enterprise Information Systems

Volume DISI

Funchal, Madeira, Portugal

June 12 – 16, 2007

Co-organized by  
**INSTICC – Institute for Systems and Technologies of Information, Control  
and Communication**  
and  
**Universidade da Madeira**

Sponsored by  
**FCT – Fundação para a Ciência e Tecnologia**

In Cooperation with  
**ACM SIGMIS – Special Interest Group on Management  
Information Systems**  
**AAAI – Association for the Advancement of Artificial Intelligence**

Copyright © 2007 INSTICC – Institute for Systems and Technologies of  
Information, Control and Communication  
All rights reserved

Edited by Jorge Cardoso, José Cordeiro and Joaquim Filipe

Printed in Portugal

ISBN: 978-972-8865-88-7

Depósito Legal: 258799/07

<http://www.iceis.org>

[secretariat@iceis.org](mailto:secretariat@iceis.org)

# BRIEF CONTENTS

---

INVITED SPEAKERS.....	IV
SPECIAL SESSION CHAIRS .....	V
ORGANIZING AND STEERING COMMITTEES .....	VI
SENIOR PROGRAM COMMITTEE.....	VII
PROGRAM COMMITTEE .....	VIII
AUXILIARY REVIEWERS .....	XII
SELECTED PAPERS BOOK .....	XIV
SPONSOR .....	XIV
FOREWORD.....	XV
CONTENTS.....	XVII

# INVITED SPEAKERS

---

**Amit Sheth**

Kno.e.sis Center, Wright State University

USA

**Wil van der Aalst**

Eindhoven University of Technology

The Netherlands

**Kurt Sandkuhl**

Jönköping University

Sweden

**Christoph Bussler**

BEA Systems, Inc.

USA

**Martin Curley**

Intel Corporation

USA

**K. Venkatesh Prasad**

Ford Motor

USA

**Larry Constantine**

University of Madeira, Portugal and

Constantine & Lockwood Ltd., USA

# SPECIAL SESSION CHAIRS

---

## SPECIAL SESSION ON BUSINESS INTELLIGENCE, KNOWLEDGE MANAGEMENT AND KNOWLEDGE MANAGEMENT SYSTEMS

**Aurora Vizcaíno**, Escuela Superior de Informática Universidad de Castilla-La Mancha, Spain

**Juan Pablo Soto**, Escuela Superior de Informática Universidad de Castilla-La Mancha, Spain

**Ezendu Ariwa**, London Metropolitan University, U.K.

## SPECIAL SESSION ON COMPUTER SUPPORTED COLLABORATIVE EDITING

**Claudia Ignat**, LORIA-INRIA Lorraine, France

**Pascal Molli**, LORIA- INRIA Lorraine, University Henri Poincaré, France

**Hala Skaf-Molli**, LORIA- INRIA Lorraine, University Henri Poincaré, France

## SPECIAL SESSION ON APPLICATIONS IN A REAL WORLD

**Wita Wojtkowski**, Boise State University, USA

## SPECIAL SESSION ON NEW INFORMATION SYSTEM AND APPROACHES FOR PRODUCT MAINTENANCE

**Stéphanie Minel**, LIPSI-ESTIA Biarritz, LAPS-Bordeaux University, France

**Jérémy Legardeur**, LIPSI-ESTIA Biarritz, LAPS-Bordeaux University, France

## SPECIAL SESSION ON COMPARATIVE EVALUATION OF SEMANTIC WEB SERVICE FRAMEWORKS

**Tiziana Margaria**, Institute for Computer Science, University Potsdam, Germany

**Ulrich Küster**, Institute for Computer Science, Friedrich-Schiller-University Jena, Germany

# ORGANIZING AND STEERING COMMITTEES

---

## CONFERENCE CHAIR

Joaquim Filipe, Polytechnic Institute of Setúbal / INSTICC, Portugal

## PROGRAM CO-CHAIRS

Jorge Cardoso, Universidade da Madeira, Madeira, Portugal

José Cordeiro, Polytechnic Institute of Setúbal / INSTICC, Portugal

## LOCAL ARRANGEMENTS

Mónica Saramago, INSTICC, Portugal

## PROCEEDINGS PRODUCTION

Andreia Costa, INSTICC, Portugal

Vítor Duarte, INSTICC, Portugal

Bruno Encarnação, INSTICC, Portugal

Luís Marques, INSTICC, Portugal

Vitor Pedrosa, INSTICC, Portugal

## CD-ROM PRODUCTION

Paulo Brito, INSTICC, Portugal

## WEBDESIGNER

Marina Carvalho, INSTICC, Portugal

## GRAPHICS PRODUCTION

Helder Coelhas, INSTICC, Portugal

## SECRETARIAT AND WEBMASTER

Vitor Pedrosa, INSTICC, Portugal

# SENIOR PROGRAM COMMITTEE

---

**Luís Amaral**, University of Minho, Portugal

**Peter Bøgh Andersen**, University of Aarhus,  
Denmark

**Senén Barro**, University of Santiago de Compostela,  
Spain

**Jean Bézivin**, ATLAS Group (INRIA & IRIN),  
University of Nantes, France

**Enrique Bonsón**, University of Huelva, Spain

**João Alvaro Carvalho**, University of Minho, Portugal

**Albert Cheng**, University of Houston, USA

**Bernard Coulette**, University of Toulouse 2, France

**Andrea De Lucia**, Università di Salerno, Italy

**Jan Dietz**, Delft University of Technology,  
The Netherlands

**Virginia Dignum**, Utrecht University,  
The Netherlands

**Schahram Dustdar**, Technical University of Vienna,  
Austria

**António Figueiredo**, University of Coimbra, Portugal

**Ulrich Frank**, ICB - University of Essen, Germany

**Nuno Guimarães**, University of Lisbon, Portugal

**Jatinder Gupta**, The University of Alabama in  
Huntsville, USA

**Erik Hollnagel**, École des Mines de Paris, France

**Dimitris Karagiannis**, University of Vienna, Austria

**Maurizio Lenzerini**, University of Rome La  
Sapienza, Italy

**Michel Leonard**, University of Geneva, Switzerland

**Kecheng Liu**, The University of Reading, UK

**Pericles Loucopoulos**, The University of Manchester,  
UK

**Paul Luker**, Higher Education Academy, UK

**Kalle Lyytinen**, Case Western Reserve University,  
USA

**Yannis Manolopoulos**, Aristotle University of  
Thessaloniki, Greece

**José Legatheaux Martins**, Faculty of Sciences and  
Technology, New University of Lisbon, Portugal

**Masao Johannes Matsumoto**, Kyushu Sangyo  
University, Japan

**Carmel McNaught**, The Chinese University of Hong  
Kong, China

**Luís Moniz Pereira**, Universidade Nova de Lisboa,  
Portugal

**George Papadopoulos**, University of Cyprus, Cyprus

**Marcin Paprzycki**, SWPS and IBS PAN, USA

**Alain Pirotte**, University of Louvain, Belgium

**Klaus Pohl**, University of Duisburg-Essen, Germany  
and Lero, The Irish software Engineering Centre,  
Ireland/Germany

**Matthias Rauterberg**, Technical University  
Eindhoven, The Netherlands

**Colette Rolland**, University of PARIS-1, France

**Narczyz Roztocki**, State University of New York  
(SUNY) at New Paltz, USA

**Abdel-Badeeh Salem**, Ain Shams University, Egypt

**Bernadette Sharp**, Staffordshire University, UK

**Timothy K. Shih**, Tamkang University, Taiwan

**Alexander Smirnov**, St. Petersburg Institute for  
Informatics and Automation of Russian Academy of  
Sciences - SPIRAS, Russia

**Ronald Stamper**, Staffordshire University, UK and  
University of Twente, NL, UK/The Netherlands

**David Taniar**, Monash University, Australia

**Miguel Toro**, University of Seville, Spain

**Antonio Vallecillo**, Universidad de Málaga, Spain

**Michalis Vazirgiannis**, Athens University of  
Economics & Business, Greece

**François Vernadat**, European Commission,  
Luxembourg

**Ioannis Vlahavas**, Aristotle University of  
Thessaloniki, Greece

**Frank Wang**, Cambridge-Cranfield High  
Performance Computing Facilities, UK

**Merrill Warkentin**, Mississippi State University,  
USA

**Hans Weigand**, Tilburg University, The Netherlands

**Yoneo Yano**, The University of Tokushima, Japan

**Mohd Syazwan Abdullah**, Universiti Utara Malaysia,  
Malaysia



# PROGRAM COMMITTEE

---

**Rama Akkiraju**, IBM Research, USA

**Patrick Albers**, ESEO - Ecole Supérieure d'Electronique de l'Ouest, France

**Salah Al-Sharhan**, Gulf University for Science and Technology, Kuwait

**Andreas Andreou**, University of Cyprus, Cyprus

**Pedro Antunes**, Faculty of Sciences, University of Lisboa, Portugal

**Ezendu Ariwa**, London Metropolitan University, UK

**Juan Carlos Augusto**, University of Ulster at Jordanstown, UK

**Bart Baesens**, University of Southampton, UK

**Cecilia Baranauskas**, UNICAMP - Universidade Estadual de Campinas, Brazil

**Steve Barker**, King's College London University, UK

**Balbir Barn**, Thames Valley University, UK

**Daniela Barreiro Claro**, Universidade Federal da Bahia, Brazil

**Nick Bassiliades**, Aristotle University of Thessaloniki, Greece

**Remi Bastide**, LIIHS-IRIT, France

**Nadia Bellalem**, University NANCY 2, France

**Orlando Belo**, University of Minho, Portugal

**Hatem Ben Sta**, Tunisia University, Tunisia

**Sadok Ben Yahia**, Faculty of Sciences of Tunis, Tunisia

**Peter Bertok**, RMIT University, Australia

**Robert Biddle**, Carleton University, Canada

**Oliver Bittel**, HTWG Konstanz - University of Applied Sciences, Germany

**Luis Borges Gouveia**, University Fernando Pessoa, Portugal

**Djamel Bouchaffra**, Oakland University, Michigan, USA

**Danielle Boulanger**, University of Lyon, France

**Jean-Louis Boulanger**, University of Technologie of Compiègne, France

**José Ângelo Braga de Vasconcelos**, Universidade Fernando Pessoa, Portugal

**Sjaak Brinkkemper**, Utrecht University, The Netherlands

**Miguel Calejo**, Declarativa, Portugal

**Coral Calero**, University of Castilla-La Mancha, Spain

**Luis M. Camarinha-Matos**, New University of Lisbon / Uninova, Portugal

**Olivier Camp**, Ecole Supérieure d'Electronique de l'Ouest, France

**Fernando Carvalho**, Universidade Federal do Ceará, Brazil

**Malu Castellanos**, HP Laboratories, USA

**Jose Jesus Castro-Schez**, Universidad de Castilla-La Mancha, Spain

**Luca Cernuzzi**, Universidad Católica, Paraguay

**Maria Filomena Cerqueira de Castro Lopes**, Universidade Portucalense Infante D. Henrique, Portugal

**Laurent Chapelier**, Fortis Banque Luxembourg, France

**William Cheng-Chung Chu**, TungHai University, Taiwan

**Chrisment Claude**, IRIT/SIG, France

**Francesco Colace**, Università degli Studi di Salerno, Italy

**Cesar Collazos**, Universidad del Cauca, Colombia

**Jose Eduardo Corcoles**, LoUISE Research Group Castilla-La Mancha University, Spain

**Antonio Corral**, University of Almeria, Spain

**Julie Cowie**, University of Stirling, UK

**Sharon Cox**, University of Central England, UK

**Alfredo Cuzzocrea**, University of Calabria, Italy

**Mohamed Dahchour**, Institut National des Postes et Télécommunications (INPT), Morocco

**Sergio de Cesare**, Brunel University, UK

**Nuno de Magalhães Ribeiro**, Centro de Estudos e Recursos Multimediáticos (CEREM), UFP, Portugal

**Suash Deb**, National Institute of Science & Technology, India

**Vincenzo Deufemia**, University of Salerno, Italy

## PROGRAM COMMITTEE (CONT.)

---

**Rajiv Dharaskar**, Nagpur University, MIET, CSI,  
India

**José Javier Dolado**, University of the Basque  
Country, Spain

**Alan Eardley**, Staffordshire University, UK

**Hans-Dieter Ehrich**, Technische Universitaet  
Braunschweig, Germany

**David Emery**, Staffordshire University, UK

**Jean-Max Estay**, Université Catholique de l'Ouest,  
France

**Yaniv Eytani**, University of Illinois at  
Urbana-Champaign, USA

**Antonio Fariña**, University of A Coruña, Spain

**Antonio Fernández-Caballero**, Universidad de  
Castilla-La Mancha, Spain

**Eduardo Fernández-Medina**, University of  
Castilla-La Mancha, Spain

**Edilson Ferneda**, Universidade Católica de Brasília,  
Brazil

**Paulo Ferreira**, INESC-ID/IST - Technical University  
of Lisbon, Portugal

**Filomena Ferrucci**, University of Salerno, Italy

**Andre Flory**, INSA, France

**Donal Flynn**, University of Manchester, UK

**Ana Fred**, IT - IST - Technical University of Lisbon,  
Portugal

**Lixin Fu**, University of Carolina, USA

**Jose A. Gallud**, University of Castilla-La Mancha,  
Spain

**Juan Garbajosa**, Universidad Politécnica de Madrid  
(UPM) - Technical University of Madrid, Spain

**Aristogiannis Garmpis**, Technological Educational  
Institution of Messolonghi, Greece

**Leonardo Garrido**, Monterrey Institute of  
Technology / Center for Intelligent Systems, Mexico

**Marcela Genero**, University of Castilla-La Mancha,  
Spain

**Joseph Giampapa**, Carnegie Mellon University, USA

**Paolo Giorgini**, University of Trento, Italy

**Raúl Giráldez**, Pablo de Olavide University of  
Seville, Spain

**Pascual González**, Universidad de Castilla-La  
Mancha, Spain

**Gustavo Gonzalez-Sanchez**, University of Girona,  
Spain

**Robert Goodwin**, Flinders University of South  
Australia, Australia

**Jaap Gordijn**, Vrije Universiteit Amsterdam,  
The Netherlands

**Silvia Gordillo**, LIFIA-UNLP, Argentina

**Feliz Gouveia**, University Fernando Pessoa / CEREM,  
Portugal

**Virginie Govaere**, INRS, France

**Sven Groppe**, University of Innsbruck, Austria

**Rune Gustavsson**, Blekinge Institute of Technology,  
Sweden

**Sissel Guttormsen Schär**, University of Bern,  
Switzerland

**Beda Christoph Hammerschmidt**, Oracle  
Corporation, USA

**Thorsten Hampel**, University of Paderborn, Germany

**Christian Heinlein**, University of Ulm, Germany

**Ajantha Herath**, Richard Stockton College of New  
Jersey, USA

**Suvineetha Herath**, Richard Stockton College of New  
Jersey, USA

**Francisco Herrera**, University of Granada, Spain

**Colin Higgins**, University of Nottingham, UK

**Peter Higgins**, Swinburne University of Technology,  
Australia

**Jun Hong**, Queen's University Belfast, UK

**Nguyen Hong Quang**, Institut de la Francophonie  
pour l'Informatique (IFI), Viet Nam

**Jiankun Hu**, RMIT University, Australia

**Kaiyin Huang**, South China Normal University,  
China

**Patrick C. K. Hung**, University of Ontario Institute of  
Technology (UOIT), Canada

**Joshua Ignatius**, Intelligent Insights International,  
Malaysia

**Hamid Jahankhani**, University of East London, UK

## PROGRAM COMMITTEE (CONT.)

---

**Arturo Jaime**, Universidad del Pais Vasco, Spain

**Ivan Jelinek**, Czech Technical University in Prague, Czech Republic

**Luis Jiménez Linares**, UCLM, Spain

**Paul Johannesson**, Royal Institute of Technology, Sweden

**Luis Joyanes**, Universidad Pontificia de Salamanca, Spain

**Nikos Karacapilidis**, University of Patras, Greece

**Nikitas Karanikolas**, Technological Educational Institute of Athens (TEI-A), Greece

**Stamatis Karnouskos**, SAP Research, Germany

**Hiroyuki Kawano**, Nanzan University, Japan

**Nicolas Kemper Valverde**, Universidad Nacional Autónoma de México, Mexico

**A. Rahman Khan**, Intel Corporation, USA

**John Krogstie**, IDI, NTNU and SINTEF, Norway

**Stan Kurkovsky**, Central Connecticut State University, USA

**Joaquín Lasheras**, University of Murcia, Spain

**James P. Lawler**, Pace University, USA

**Chul-Hwan Lee**, University of Pittsburgh, USA

**Alain Leger**, France Telecom R&D, France

**Carlos León de Mora**, University of Sevilla, Spain

**Joerg Leukel**, University of Hohenheim, Germany

**Hareton Leung**, Hong Kong Polytechnic University, China

**Therese Libourel**, LIRMM, France

**John Lim**, National University of Singapore, Singapore

**Selma Limam Mansar**, Zayed University, UAE

**ZongKai Lin**, Chinese Academy of Sciences, China

**Matti Linna**, University of Vaasa, Finland

**Jan Ljungberg**, Gothenburg University, Sweden

**Stephane Loiseau**, LERIA, France

**João Correia Lopes**, University of Porto, Portugal

**Víctor López-Jaquero**, LoUISE Group, University of Castilla-La Mancha, Spain

**María Dolores Lozano**, University of Castilla-La Mancha, Spain

**Miguel R. Luaces**, Universidade da Coruña, Spain

**Christopher Lueg**, University of Tasmania, Australia

**Mark Lycett**, Brunel University, UK

**Edmundo Madeira**, UNICAMP - Universidade Estadual de Campinas, Brazil

**Laurent Magnin**, University of Montreal, Canada

**Sam Makki**, University of Toledo, USA

**Mirko Malekovic**, FOI - Zagreb University, Croatia

**Nuno Mamede**, IST / L2F of INESC-ID Lisboa, Portugal

**João Bosco Manguiera Sobral**, UFSC / CPGCC - Computer Science Pos-Graduate, Brazil

**Esperanza Marcos**, Universidad Rey Juan Carlos, Spain

**Farhi Marir**, London Metropolitan University, UK

**Maria João Marques Martins**, IST - Technical University of Lisbon, Portugal

**Herve Martin**, LSR-IMAG, France

**Miguel Angel Martinez**, University of Murcia, Spain

**Johannes Mayer**, University of Ulm, Germany

**Andreas Meier**, University of Fribourg, Switzerland

**Emilia Mendes**, The University of Auckland, New Zealand

**Engelbert Mephu Nguifo**, Université d'Artois - IUT de Lens, CRIL-CNRS, France

**John Miller**, University of Georgia, USA

**Subhas Misra**, Carleton University, Canada

**Sudip Misra**, Cornell University, USA

**Michele Missikoff**, IASI-CNR, Italy

**Ghodrat Moghadampour**, Vaasa University of Applied Sciences, Finland

**Pascal Molli**, LORIA, Université Henri Poincaré, Nancy 1, France

**Francisco Montero**, LoUISE Research Group, University of Castilla-La Mancha, Spain

**Paula Morais**, Universidade Portucalense, Portugal

**Fernando Moreira**, Universidade Portucalense, Portugal

## PROGRAM COMMITTEE (CONT.)

---

**Gianluca Moro**, DEIS, University of Bologna, Italy

**Haralambos Mouratidis**, University of East London, UK

**Nathalie Nathalie**, Universidad de Málaga, Spain

**Ana Neves**, knowman - Consultadoria em Gestão, Lda, Portugal

**Jose Angel Olivas**, University of Castilla-La Mancha, Spain

**Luis Olsina Santos**, Universidad Nacional de La Pampa, Argentina

**Peter Oriogun**, London Metropolitan University, UK

**José R. Paramá**, University of A Coruña, Spain

**João Pascoal Faria**, FEUP - Faculty of Engineering of University of Porto, Portugal

**Vicente Pelechano**, Universidad Politécnica de Valencia, Spain

**Maria Carmen Penadés Gramaje**, Technical University of Valencia, Spain

**Gabriel Pereira Lopes**, Universidade Nova de Lisboa, Portugal

**Laurent Péridy**, IMA-UCO, France

**Dana Petcu**, Western University of Timisoara, Romania

**Steeff Peters**, Vrije Universiteit Amsterdam, The Netherlands

**Paolo Petta**, Medical University of Vienna, Austrian Research Institute for Artificial Intelligence, Austria

**José Pires**, Escola Superior de Tecnologia e Gestão / IPB, Portugal

**Geert Poels**, Ghent University, Belgium

**Abdul Razak Rahmat**, University Utara Malaysia, Malaysia

**Jolita Ralyte**, University of Geneva, Switzerland

**Pedro Ramos**, ISCTE, Portugal

**Hajo A. Reijers**, Eindhoven University of Technology, The Netherlands

**Ulrich Reimer**, University of Applied Sciences St. Gallen, Switzerland

**Marinette Revenu**, Greyc Ensicaen, France

**Yacine Rezzoui**, University of Salford, UK

**Simon Richir**, Presence & Innovation Lab. ENSAM Laval, France

**Maria Rigou**, University of Patras, Greece

**Roland Ritsch**, University of Applied Sciences St. Gallen, Switzerland

**David Rivreau**, Université Catholique de l'Ouest, France

**Daniel Rodriguez**, University of Alcalá, Spain

**Pilar Rodriguez**, Universidade Autónoma de Madrid, Spain

**Jimena Rodriguez Arrieta**, University of the Basque Country, Spain

**Oscar M. Rodriguez-Elias**, Universidad Autónoma de Baja California (UABC), Mexico

**Jose Raul Romero**, University of Cordoba, Spain

**Agostinho Rosa**, IST - Technical University of Lisbon, Portugal

**Gustavo Rossi**, LIFIA-UNLP, Argentina

**Francisco Ruiz**, University of Castilla-La Mancha, Spain

**Roberto Ruiz**, Pablo de Olavide University, Spain

**Ángeles S. Places**, University of A Coruña, Spain

**Manuel Santos**, Universidade do Minho, Portugal

**Daniel Schang**, ESEO, France

**Mareike Schoop**, University of Hohenheim, Germany

**Isabel Seruca**, Universidade Portucalense, Portugal

**Hanifa Shah**, Staffordshire University, UK

**Jianhua Shao**, Cardiff University, UK

**Alberto Silva**, INESC/IST - Technical University of Lisbon, Portugal

**Maria João Silva Costa Ferreira**, Universidade Portucalense, Portugal

**Janice Sipior**, Villanova University, USA

**Spiros Sirmakessis**, Technological Educational Institution of Messolongi, Greece

**Hala Skaf-Molli**, INRIA Lorraine - University Henri Poincaré, France

**Chantal Soule-Dupuy**, University of Toulouse 1 - IRIT, France

**Chris Stary**, University of Linz, Austria

## PROGRAM COMMITTEE (CONT.)

---

**Janis Stirna**, Jönköping University, Sweden

**Vijayan Sugumaran**, Oakland University, USA

**Lily Sun**, The University of Reading, UK

**Ramayah T.**, Universiti Sains Malaysia, Malaysia

**Sotirios Terzis**, University of Strathclyde, UK

**Philippe Thiran**, University of Namur, Belgium

**Claudine Toffolon**, Université du Maine - LIUM, France

**Robert Tolksdorf**, Freie Universität Berlin, Germany

**Ambrosio Toval**, University of Murcia, Spain

**Grigorios Tsoumakas**, Aristotle University of Thessaloniki, Greece

**Theodoros Tzouramanis**, University of the Aegean, Greece

**Gulden Uchyigit**, Imperial College, UK

**Athina Vakali**, Aristotle University of Thessaloniki, Greece

**Michael Vassilakopoulos**, Technological Educational Institute of Thessaloniki, Greece

**Christine Verdier**, University Joseph Fourier Grenoble, France

**Maria-Amparo Vila**, University of Granada, Spain

**Tuong Vinh Ho**, Institut de la Francophonie pour l'Informatique (IFI), Viet Nam

**Aurora Vizcaino**, Escuela Superior de Informática, Spain

**Bing Wang**, University of Hull, UK

**Hans Weghorn**, University of Cooperative Education, Stuttgart, Germany

**Gerhard Weiss**, SCCH, Austria

**Graham Winstanley**, University of Brighton, UK

**Claus Witfelt**, ITU, Denmark

**Wita Wojtkowski**, Boise State University, USA

**Robert Wrembel**, Poznan University of Technology, Poland

**Baowen Xu**, Southeast University, China

**Haiping Xu**, University of Massachusetts Dartmouth, USA

**Hongji Yang**, De Montfort University, UK

**Jasmine Yeap**, Intelligent Insights International, Malaysia

**Kokou Yetongnon**, University of Bourgogne, France

**Jun Zhang**, SUN Yat-sen University, China

**Liping Zhao**, The University of Manchester, UK

**Shuigeng Zhou**, Fudan University, China

**Ester Zumpano**, University of Calabria, Italy

## AUXILIARY REVIEWERS

---

**Cesar Javier Acuña**, Universidad Rey Juan Carlos, Spain

**Pedro P. Alarcón**, UPM, Madrid, Spain

**Grigoris Antoniou**, University of Crete, Greece

**Jose M<sup>a</sup> Cavero Barca**, Universidad Rey Juan Carlos, Spain

**Manuel Bollaín**, UPM, Madrid, Spain

**M. Boughanem**, IRIT/SIG, France

**João Paulo Caldeira**, ISR/Politechnic Institute of Setubal, Portugal

**Angélica Caro**, University of Bio-Bio, Chile

**M<sup>a</sup> Valeria de Castro**, Universidad Rey Juan Carlos, Spain

**Isabel Nepomuceno Chamorro**, University of Seville, Spain

**Evandro de Barros Costa**, UFAL, Brazil

**Guillermo Covella**, GIDIS\_Web / Universidad Nacional de La Pampa, Argentina

**David Benavides Cuevas**, Universidad de Sevilla, Spain

**Yuhui Deng**, Cranfield University, UK

**Vincenzo Deufemia**, University of Salerno, Italy

## AUXILIARY REVIEWERS (CONT.)

---

**Norberto Diaz Diaz**, University of Seville, Spain

**Angelina Espinoza**, UPM, Madrid, Spain

**Jian Feng**, QilinSoft (China) Ltd, China

**Carlos Fernandes**, IRS/Technical University of Lisbon, Portugal

**Óscar Pedreira Fernández**, University of A Coruña, Spain

**Rita Francese**, University of Salerno, Italy

**Vittorio Fuccella**, University of Salerno, Italy

**Arantza Irastorza Goñi**, Universidad del Pais Vasco, Spain

**Carmine Gravino**, University of Salerno, Italy

**Na Helian**, London Metropolitan University, UK

**Mohamed Said El Hmam**, LGI2A / University of Artois, France

**Kaidong Huang**, Century Securities, China

**Geert Jonker**, Utrecht University, The Netherlands

**Ioannis Katakis**, Aristotle University of Thessaloniki, Greece

**Dimitrios Katsaros**, University of Thessaly, Greece

**Vasiliki Koutsonikola**, Aristotle University, Greece

**Marcelo Ladeira**, UnB, Brazil

**Elie Abiu Lahoud**, University of Borgogne, France

**Oriana Licchelli**, ESEO, France

**Mondher Maddouri**, URPAH-INSAT, Tunis, Tunisia

**Francisco Javier Lucas Martinez**, University of Murcia, Spain

**Sergio di Martino**, University of Salerno, Italy

**José P. Molina Massó**, University of Castilla-La Mancha, Spain

**Juan Manuel Vara Mesa**, Universidad Rey Juan Carlos, Spain

**Gabriele Monti**, University of Bologna, DEIS Cesena, Italy

**M<sup>a</sup> Ángeles Moraga**, University of Castilla-La Mancha, Spain

**Diego Seco Naveiras**, University of A Coruña, Spain

**Antonio De Nicola**, IASI-CNR, Italy

**Joaquin Nicolas**, University of Murcia, Spain

**Germana Menezes da Nóbrega**, UCB, Brazil

**George Pallis**, Aristotle University, Greece

**Efi Papatheocharous**, University of Cyprus, Cyprus

**Ignazio Passero**, University of Salerno, Italy

**Hércules Antonio do Prado**, Embrapa, Brazil

**Ajith Ranabahu**, University of Georgia, USA

**Michele Risi**, University of Salerno, Italy

**Ilias Sakellariou**, City College, Thessaloniki, Greece

**Ivo José Garcia dos Santos**, UNICAMP - University of Campinas, Brazil

**K. Sauvagnat**, IRIT/SIG, France

**Giuseppe Scanniello**, University of Basilicata, Italy

**Joaquin Peña Siles**, Universidad de Sevilla, Spain

**Mehdi Snene**, University of Geneva, Switzerland

**Konstantinos Stamos**, Aristotle University, Greece

**Francesco Taglino**, IASI-CNR, Italy

**Guilaine Talens**, University of Lyon, France

**Rui Tavares**, ISR/University of Evora, Portugal

**Olivier Teste**, IRIT/SIG, France

**Jean-Marc Thevenin**, University of Toulouse 1 and IRIT, France

**Norbert Tsopze**, CRIL / University of Artois, France

**Sining Wu**, Cranfield University, UK

**Agustín Yagüe**, UPM, Madrid, Spain

## SELECTED PAPERS BOOK

---

A number of selected papers presented at ICEIS 2007 will be published by Springer, in a book entitled Enterprise Information Systems IX. This selection will be done by the conference Chair and program co-chairs, among the papers actually presented at the conference, based on a rigorous review by the ICEIS 2007 program committee members.

## SPONSOR

---

The logo for FCT (Fundação para a Ciência e a Tecnologia) consists of the letters 'FCT' in a large, bold, dark green sans-serif font.

Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

# FOREWORD

---

This volume contains the proceedings of the Ninth International Conference on Enterprise Information Systems (ICEIS 2007) organized by the Institute for Systems and Technologies of Information Control and Communication (INSTICC) and the University of Madeira, in collaboration with ACM/SIGMIS and AAI. Furthermore, the conference was sponsored by the Portuguese Foundation for Science and Technology (FCT).

ICEIS has become a major point of contact between research scientists, engineers and practitioners in the area of business applications of information systems. This year, five simultaneous tracks were held, covering different aspects related to enterprise computing, including: “Databases and Information Systems Integration”, “Artificial Intelligence and Decision Support Systems”, “Information Systems Analysis and Specification”, “Software Agents and Internet Computing” and “Human-Computer Interaction”. All tracks describe research work that is often oriented towards real world applications and highlight the benefits of Information Systems and Technology for industry and services, thus making a bridge between the Academia and the Enterprise worlds.

Following the success of 2006, ICEIS 2007 also had a number of satellite workshops, related to the field of the conference. This year we collaborated in the organization of the following ten international workshops: 7<sup>th</sup> International Workshop on Pattern Recognition in Information Systems; 1<sup>st</sup> International Joint Workshop on Wireless Ubiquitous Computing; 5<sup>th</sup> International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems; 5<sup>th</sup> International Workshop on Security In Information Systems; 4<sup>th</sup> International Workshop on Natural Language Processing and Cognitive Science; 4<sup>th</sup> International Workshop on Computer Supported Activity Coordination; 3<sup>rd</sup> International Workshop on Model-Driven Enterprise Information Systems; 1<sup>st</sup> International Joint Workshop on Technologies for Collaborative Business Processes and Management of Enterprise Information Systems; 1<sup>st</sup> International Workshop on RFID Technology - Concepts, Applications, Challenges and 1<sup>st</sup> International Workshop on Human Resource Information Systems.

This year, ICEIS 2007 received 644 paper submissions from more than 40 countries in all continents. 72 papers were published and presented as full papers, i.e. completed work (8 pages/30' oral presentation), 198 papers reflecting work-in-progress or position papers were accepted for short presentation, and another 131 contributions were scheduled for poster presentation.

These numbers, leading to a “full-paper” acceptance ratio below 12%, and a total acceptance ratio below 65%, show the intention of preserving a high quality forum for the next editions of this conference. Additionally, as usual in the ICEIS conference series, a number of invited talks, presented by internationally recognized specialists in different areas, have positively contributed to reinforce the overall quality of the Conference and to provide a deeper understanding of the Enterprise Information Systems field.

A book of Selected Papers will be published, following the conference, by Springer in the newly created series “Lecture Notes in Business Information Processing” (LNBIP). This series brings the



successful LNCS approach to areas such as business information systems, e-business, B2B integration, Enterprise applications and industrial software development.

The program for this conference required the dedicated effort of many people. Firstly, we must thank the authors, whose research and development efforts are recorded here. Secondly, we thank the members of the program committee and the additional reviewers for their diligence and expert reviewing. Thirdly, we thank the invited speakers for their invaluable contribution and for taking the time to synthesise and prepare their talks. Fourthly, we thank the workshop chairs and the special session chairs whose collaboration with ICEIS was much appreciated. Finally, special thanks to all the members of the local organising committee, especially Jorge Cardoso, whose collaboration was fundamental for the success of this conference.

This year, the organization will distribute two awards to papers presented at the conference: the best paper award and the best student paper award, mainly based on the classifications provided by the Program Committee members.

We wish you all an exciting conference and an unforgettable stay in the lovely island of Madeira. We hope to meet you again next year for the 10<sup>th</sup> ICEIS, to be held in Barcelona - Spain, details of which are available at <http://www.iceis.org>.

Joaquim Filipe

I.P.Setúbal/ INSTICC, Portugal

José Cordeiro

I.P.Setúbal/INSTICC, Portugal

Jorge Cardoso

Universidade da Madeira, Portugal

# CONTENTS

---

## INVITED SPEAKERS

### KEYNOTE LECTURES

- THE 4 X 4 SEMANTIC MODEL - Exploiting Data, Functional, Non-functional and Execution Semantics  
Across Business Process, Workflow, Partner Services and Middleware Services Tiers IS-5  
*Amit P. Sheth and Karthik Gomadam*
- TRENDS IN BUSINESS PROCESS ANALYSIS - From Verification to Process Mining IS-13  
*Wil M. P. van der Aalst*
- INFORMATION LOGISTICS IN NETWORKED ORGANISATIONS - Issues, Concepts and  
Applications IS-23  
*Kurt Sandkubal*
- SERVICE-ORIENTED ARCHITECTURE - One Size Fits Nobody IS-31  
*Christoph Busler*
- INTRODUCING AN IT CAPABILITY MATURITY FRAMEWORK IS-35  
*Martin Curley*
- DRIVING AHEAD: JOINT ENTERPRISE-EMBEDDED COMPUTING IN SMART CLOUDS,  
SMART DUST AND INTELLIGENT AUTOMOBILES IS-49  
*K. Venkatesh Prasad*
- ENTERPRISE INFORMATION SYSTEMS FOR USE: FROM BUSINESS PROCESSES TO HUMAN  
ACTIVITY IS-51  
*Larry Constantine*

## DATABASES AND INFORMATION SYSTEMS INTEGRATION

### FULL PAPERS

- STATISTICS API: DBMS-INDEPENDENT ACCESS AND MANAGEMENT OF DBMS STATISTICS  
IN HETEROGENEOUS ENVIRONMENTS 5  
*Tobias Kraft and Bernhard Mitschang*
- USING AN INDEX OF PRECOMPUTED JOINS IN ORDER TO SPEED UP SPARQL PROCESSING 13  
*Sven Groppe, Jinghua Groppe and Volker Linnemann*
- ONE-TO-MANY DATA TRANSFORMATION OPERATIONS - Optimization and Execution on an  
RDBMS 21  
*Pauo Carreira, Helena Galhardas, João Pereira and Andrzej Wichert*
- MONITORING WEB DATA SOURCES USING TEMPORAL PROPERTIES AS AN EXTERNAL  
RESOURCES OF A DATA WAREHOUSE 28  
*Francisco Araque, Alberto Salguero and Cecilia Delgado*
- TRANSFORMATION OF LEGACY BUSINESS SOFTWARE INTO CLIENT-SERVER  
ARCHITECTURES 36  
*Thomas Rauber and Gudula Rünger*

A NEW ALGORITHM FOR TWIG PATTERN MATCHING <i>Yangjun Chen</i>	44
XML SCHEMA STRUCTURAL EQUIVALENCE <i>Angela C. Duta, Ken Barker and Reda Albaji</i>	52
MODELING DIMENSIONS IN THE XDW MODEL - A LVM-Driven Approach <i>R. Rajagan, Elizabeth Chang and Tharam S. Dillon</i>	60
INFORMATION SYSTEMS INTEGRATION DURING MERGERS - Integration Modes Typology and Integration Paths <i>Gérald Brunetto</i>	71
A VIRTUALIZATION APPROACH FOR REUSING MIDDLEWARE ADAPTERS <i>Ralf Wagner and Bernhard Mitschang</i>	78
XML INDEX COMPRESSION BY DTD SUBTRACTION <i>Stefan Böttcher, Rita Steinmetz and Niklas Klein</i>	86
DYNAMIC COMMIT TREE MANAGEMENT FOR SERVICE ORIENTED ARCHITECTURES <i>Stefan Böttcher and Sebastian Obermeier</i>	95
DETERMINING THE COSTS OF ERP IMPLEMENTATION <i>Rob J. Kusters, Fred J. Heemstra and Arjan Jonker</i>	102
<b>SHORT PAPERS</b>	
STAH-TREE: Hybrid Index for Spatio Temporal Aggregation <i>Marcin Goramski and Michał Faruga</i>	113
PREFERENCE RULES IN DATABASE QUERYING <i>Sergio Greco, Cristian Molinaro and Francesco Parisi</i>	119
SIMPLIFIED QUERY CONSTRUCTION - Queries Made as Easy as Possible <i>Brad Arshinoff, Damon Ratcliffe, Martin Saetre, Reda Albaji and Tansel Özyer</i>	125
DISTRIBUTED APPROACH OF CONTINUOUS QUERIES WITH KNN JOIN PROCESSING IN SPATIAL DATA WAREHOUSE <i>Marcin Goramski and Wojciech Gębczyk</i>	131
ACTIVITY WAREHOUSE: DATA MANAGEMENT FOR BUSINESS ACTIVITY MONITORING <i>Oscar Mangisengi, Mario Pichler, Dagmar Auer, Dirk Drabeim and Hildegard Rumetsbofer</i>	137
UNASSUMING VIEW-SIZE ESTIMATION TECHNIQUES IN OLAP - An Experimental Comparison <i>Kamel Aouiche and Daniel Lemire</i>	145
OLAP AGGREGATION FUNCTION FOR TEXTUAL DATA WAREHOUSE <i>Franck Ravat, Olivier Teste and Ronan Tournier</i>	151
REVISITING THE OLAP INTERACTION TO COPE WITH SPATIAL DATA AND SPATIAL DATA ANALYSIS <i>Rosa Matias and João Moura-Pires</i>	157
USING FUZZY DATACUBES IN THE STUDY OF TRADING STRATEGIES <i>M. Delgado Calvo-Flores, J. F. Nuñez Negrillo, E. Gibaja Galindo and C. Molina Fernández</i>	164
PIN: A PARTITIONING & INDEXING OPTIMIZATION METHOD FOR OLAP <i>Ricardo Jorge Santos and Jorge Bernardino</i>	170

FROM DATABASE TO DATAWAREHOUSE: A Design Quality Evaluation <i>Maurizio Pighin and Lucio Ieronutti</i>	178
IMPLEMENTING SPATIAL DATAWAREHOUSE HIERARCHIES IN OBJECT-RELATIONAL DBMSs <i>Elżbieta Malinowski and Esteban Zimányi</i>	186
A DOCUMENT REPOSITORY ARCHITECTURE FOR HETEROGENEOUS BUSINESS INFORMATION MANAGEMENT <i>Mohamed Mbarkei, Chantal Soulé-Dupuy and Nathalie Vallès-Parlangeau</i>	192
EXTRACTION AND TRANSFORMATION OF DATA FROM SEMI-STRUCTURED TEXT FILES USING A DECLARATIVE APPROACH <i>R. Raminhos and J. Moura-Pires</i>	199
DIMENSION HIERARCHIES UPDATES IN DATA WAREHOUSES - A User-driven Approach <i>Cécile Favre, Fadila Bentayeb and Omar Boussaid</i>	206
SECURE KNOWLEDGE EXCHANGE BY POLICY ALGEBRA AND ERML <i>Steve Barker and Paul Douglas</i>	212
EXPOSING WORKFLOWS TO LOAD BURSTS <i>Dmytro Dyachuk and Ralph Deters</i>	218
USING SEMANTIC WEB AND SERVICE ORIENTED TECHNOLOGIES TO BUILD LOOSELY COUPLED SYSTEMS: SWOAT – A Service and Semantic Web Oriented Architecture Technology <i>Bruno Caires and Jorge Cardoso</i>	226
THE CONCEPTUAL FRAMEWORK FOR BUSINESS PROCESS INNOVATION: Towards a Research Program on Global Supply Chain Intelligence <i>Charles Møller</i>	233
TRANSACTION SERVICE COMPOSITION - A Study of Compatibility Related Issues <i>Anna-Briith Arnisen and Randi Karlsen</i>	239
MEDIATION FRAMEWORK FOR ENTERPRISE INFORMATION SYSTEM INFRASTRUCTURES: Application-driven Approach <i>Leonid Kalinichenko, Dmitry Briukhov, Dmitry Martynov, Nikolay Skvortsov and Sergey Stupnikov</i>	246
A FRAMEWORK FOR SUPPORTING KNOWLEDGE WORK PROCESSES <i>Weidong Pan, Igor Hawryszkiemycz and Dongbei Xue</i>	252
A METHOD FOR EARLY CORRESPONDENCE DISCOVERY USING INSTANCE DATA <i>Indrakshi Ray and C. J. Michael Geisterfer</i>	258
PROBLEMS WITH NON-OPEN DATA STANDARDS IN SWEDISH MUNICIPALS - When Integrating and Adopting Systems <i>Benneth Christiansson and Fredrik Svensson</i>	264
CHANGE MANAGEMENT IN DATA INTEGRATION SYSTEMS <i>Rabee Ghurburn, Philippe Beaune and Hugues Solignac</i>	268
RELEVANT VALUES: NEW METADATA TO PROVIDE INSIGHT ON ATTRIBUTE VALUES AT SCHEMA LEVEL <i>Sonia Bergamaschi, Mirko Orsini, Francesco Guerra and Claudio Sartori</i>	274
LEGACY SYSTEM EVOLUTION – A Comparative Study of Modernisation and Replacement Initiation Factors <i>Irja Kankaanpää, Päivi Tübonen, Jarmo J. Ahonen, Jussi Koskinen, Tero Tilus and Henna Sivula</i>	280

A XML-BASED QUALITY MODEL FOR WEB SERVICES CERTIFICATION <i>J. Jorge Dias Jr., J. Adson O. G. da Cunha, Alexandre Álvaro, Roberto S. M. de Barros and Sílvia Meira</i>	288
AN INSERTION STRATEGY FOR A TWO-DIMENSIONAL SPATIAL ACCESS METHOD <i>Wendy Osborn and Ken Barker</i>	295
PTSM: A PORTLET SELECTION MODEL <i>M<sup>a</sup> Ángeles Moraga, Coral Calero, Mario Piattini and Oscar Díaz</i>	301
KNOWLEDGE-MASHUPS AS NEXT GENERATION WEBBASED SYSTEMS - Converging Systems via Self-explaining Services <i>Thomas Bopp, Bürger Kühnel, Thorsten Hampel, Christian Prpitsch and Frank Lützenkirchen</i>	307
ON CORRECTNESS CRITERIA FOR WORKFLOW <i>Belinda M. Carter and Maria E. Orlowska</i>	315
ENABLING CSCW SYSTEMS TO AUTOMATICALLY BIND EXTERNAL KNOWLEDGE BASES <i>Thomas Bopp, Jonas Schulte and Thorsten Hampel</i>	323
WFSelector - A Tool for Comparing and Selecting Workflow Engines <i>Karim Baïna</i>	330
INTEGRATING IDENTIFICATION CONSTRAINTS IN WEB ONTOLOGY <i>Thi Dieu Thu Nguyen and Nhan Le-Thanh</i>	338
EXTENSIBLE METADATA REPOSITORY FOR INFORMATION SYSTEMS AND ENTERPRISE APPLICATIONS <i>Ricardo Ferreira and João Moura-Pires</i>	344
AN EXTENSIBLE RULE TRANSFORMATION MODEL FOR XQUERY OPTIMIZATION - Rules Pattern for XQuery Tree Graph View <i>Nicolas Travers and Tuyêt Trâm Dang Ngoc</i>	351
DQXSD: AN XML SCHEMA FOR DATA QUALITY - An XSD for Supporting Data Quality in XML <i>Eugenio Verbo, Ismael Caballero and Mario Piattini</i>	359
ENTERPRISE INFORMATION SEARCH SYSTEMS FOR HETEROGENEOUS CONTENT REPOSITORIES <i>Trieu C. Chieu, Shyh-Kwei Chen and Shiva S. Fu</i>	365
EVIE - AN EVENT BROKERING LANGUAGE FOR THE COMPOSITION OF COLLABORATIVE BUSINESS PROCESSES <i>Tony O'Hagan, Shazîa Sadiq and Wasim Sadiq</i>	372
SEMANTIC ORCHESTRATION MERGING - Towards Composition of Overlapping Orchestrations <i>Clementine Nemo, Mireille Blay-Fornarino, Michel Riveill and Günter Kniessel</i>	378
MAINTENANCE COST OF A SOFTWARE DESIGN: A Value-Based Approach <i>Daniel Cabrero, Javier Garzías and Mario Piattini</i>	384
AN INFORMATION SYSTEMS AUDITOR'S PROFILE <i>Mariana Carroll and Alta van der Merwe</i>	390
SOFTWARE COST ESTIMATION USING ARTIFICIAL NEURAL NETWORKS WITH INPUTS SELECTION <i>Efi Papatheocharous and Andreas Andreou</i>	398
DOING THINGS RIGHT OR DOING THE RIGHT THINGS? Proposing a Documentation Scheme for Small to Medium Enterprises <i>Josephine Antoniou, Panagiotis Germanakos and Andreas S. Andreou</i>	408

THE CHALLENGES FACING GLOBAL ERP SYSTEMS IMPLEMENTATIONS <i>Paul Hawking, Andrew Stein and Susan Foster</i>	415
OOPUS - A Production Planning Information System to Assure High Delivery Reliability under Short-term Demand Changes and Production Disturbances <i>Wilhelm Dangelmaier, Tobias Rust, Thomas Hermanowski, Daniel Brüggemann, Daniel Kaschula, Andre Döring and Thorsten Timm</i>	423
 <b>POSTERS</b>	
MODEL-DRIVEN DEVELOPMENT USING STANDARD TOOLS <i>Julián Garrido, M<sup>a</sup> Angeles Martos and Fernando Berzal</i>	433
DEVELOPMENT OF AN ACCOUNTING SYSTEM - Applying the Incrementally Modular Abstraction Hierarchy to a Complex System <i>Kenji Ohmori and Tosiyasu L. Kunii</i>	437
AN ORDER ALLOCATION MODEL IN VIRTUAL ENTERPRISES BASED ON INDUSTRIAL CLUSTERS <i>Fangqi Cheng, Feifan Ye and Jianguo Yang</i>	445
A WEB TOOL FOR WEB DOCUMENT AND DATA SOURCE SELECTION WITH SQLFI <i>Marlene Goncalves and Leonid Tineo</i>	449
A DATABASE INTEGRATION SYSTEM BASED ON GLOBAL VIEW GENERATION <i>Uchang Park and Ramon Lawrence</i>	453
A KOREAN SEARCH PATTERN IN THE LIKE OPERATION <i>Sung Chul Park, Eun Hyang Lo, Jong Chul Park and Young Chul Park</i>	457
ON THE SEMI-AUTOMATIC VALIDATION AND DECOMPOSITION OF TERNARY RELATIONSHIPS WITH OPTIONAL ELEMENTS <i>Ignacio-J. Santos, Paloma Martínez Fernández and Dolores Cuadra</i>	465
INDUCTION OF DATA QUALITY PROTOCOLS INTO BUSINESS PROCESS MANAGEMENT <i>Shazia Sadiq, Maria Orlowska and Wasim Sadiq</i>	473
MODELING OF AN ANALYTICAL DATABASE SYSTEM <i>Alex Sandro Romeu de Souza Poletto and Jorge Rady Almeida Junior</i>	477
USABILITY ISSUES IN SERVICE-ORIENTED ARCHITECTURE <i>Jaroslav Král and Michal Zemlicka</i>	482
MULTIDIMENSIONAL VECTOR ROUTING IN A P2P NETWORK <i>Laurent Yeb, Georges Gardarin and Florin Dragan</i>	486
THE HAV DATA INTEGRATION APPROACH: The Mapping in HAV <i>Fatima Boulçane</i>	490
TIMING BEHAVIOR ANOMALY DETECTION IN ENTERPRISE INFORMATION SYSTEMS <i>Matthias Rohr, Simon Giesecke and Wilhelm Hasselbring</i>	494
UNDERSTANDING THE DYNAMICS OF INFORMATION SYSTEMS <i>Abdelwahab Hamou-Lhadj</i>	498
ONE-TO-MANY DATA TRANSFORMATIONS - As Relational Operations <i>Paulo Carreira</i>	503

AN EXECUTIVE INFORMATION SYSTEM FOR SECURITIES BROKER'S RISK MANAGEMENT WITH DATA WAREHOUSING AND OLAP <i>Yung-Hsin Wang, Shing-Han Li and Kuo-Lung Sun</i>	508
STRATEGIC FRAMEWORK TO IMPLEMENT A TELECOMMUNICATIONS BUSINESS INTELLIGENCE SOLUTION IN A DEVELOPING COUNTRY <i>D. P. du Plessis and T. McDonald</i>	512
TOWARDS INDUSTRIAL SERVICE BUSINESS: Challenges in Designing ICT Support for the Networks of Companies <i>Sauli Häppavuori, Markus Hänninen, Samuli Pekkola and Kari Luostarinen</i>	516
MANAGING COMPLEX INFORMATION IN REACTIVE APPLICATIONS USING AN ACTIVE TEMPORAL XML DATABASE APPROACH <i>Essam Mansour, Kudakwashe Dube and Bing Wu</i>	520
ORGANIZATIONAL ISSUES ON COOPETTIVE FEDERATED INFORMATION SYSTEMS <i>Mirko Cesarini and Mario Mezzanzanica</i>	524
A DATABASE MANAGEMENT SYSTEM KERNEL FOR IMAGE COLLECTIONS <i>Liana Stanescu, Dumitru Burdescu, Cosmin Stoica and Marius Brezovan</i>	530
INTEGRATING ENTERPRISE DATA FOR DECISION SUPPORT IN CONSTRUCTION ORGANISATIONS <i>Tanko Ishaya, James Chadband and Lucy Grierson</i>	534
A NEW LOOK INTO DATA WAREHOUSE MODELLING <i>Nikolay Nikolov</i>	540
A MULTI-VIEWS REPOSITORY FOR MULTI-STRUCTURED DOCUMENTS <i>Karim Djemal</i>	544
INCENTIVES AND OBSTACLES IN IMPLEMENTING INTER-ORGANISATIONAL INTEROPERABILITY <i>Rajja Halonen and Veikko Halonen</i>	549
A FRAMEWORK FOR MODEL-DRIVEN PATTERN MATCHING <i>Ignacio García-Rodríguez de Guzmán, Macario Polo and Mario Piattini</i>	553
A METRICS PROPOSAL TO EVALUATE SOFTWARE INTERNAL QUALITY WITH SCENARIOS <i>Anna Grimán, María Pérez, Maryoly Ortega and Luis Mendoza</i>	558
A METHOD PROPOSAL FOR ARCHITECTURAL RELIABILITY EVALUATION <i>Anna Grimán, María Pérez, Luis E. Mendoza and Edumilis Méndez</i>	564
TOURISM INFORMATION AGGREGATION USING AN ONTOLOGY BASED APPROACH <i>Miguel Gouveia and Jorge Cardoso</i>	569
AN OVERVIEW OF THE OBJECT-ORIENTED DATABASE PROGRAMMING LANGUAGE DBPQL <i>Markus Kirchberg</i>	573
A DATA WAREHOUSE ARCHITECTURE FOR INTEGRATING FIELD-BASED DATA <i>Alberto Salguero, Francisco Araque and Ramón Carrasco</i>	577
TEXT ANALYTICS AND DATA ACCESS AS SERVICES - A Case Study in Transforming a Legacy Client-server Text Analytics Workbench and Framework to SOA <i>E. Michael Maximilien, Ying Chen, Ana Lelescu, James Rhodes, Jeffrey Kreulen and Scott Spangler</i>	581

A CONTINUOUS DATA INTEGRATION METHODOLOGY FOR SUPPORTING REAL-TIME DATA WAREHOUSING <i>Ricardo Jorge Santos and Jorge Bernardino</i>	589
<b>SPECIAL SESSION ON NEW INFORMATION SYSTEM AND APPROACHES FOR PRODUCT MAINTENANCE</b>	
LOGISTICS TRACEABILITY FOR SUPPLY CHAIN IMPROVEMENT - Case Study of SMMART Project <i>Paulina Blaszkomska, Jana Pieriegud and Michal Wolanski</i>	599
A POLICY-BASED PRIVACY STORAGE APPROACH <i>Julien Nowalczyk and Frédérique Tastet-Cherel</i>	605
SUPPLY CHAIN IMPROVEMENT - Assessing Readiness for Change through Collaboration Evaluation <i>Olivier Zephir, Emilie Chapotot, Stéphanie Minel and Benoît Roussel</i>	609
ACTIVITY THEORY MODEL - Application in the Automotive Industry <i>Jon Aldazabal, Gaizka Ballesteros and Juan Antonio Martín</i>	615
AUTHOR INDEX	621



# DQXSD: AN XML SCHEMA FOR DATA QUALITY

## *An XSD for Supporting Data Quality in XML*

Eugenio Verbo, Ismael Caballero  
*Soluziona Consultancy and Technology*  
*UCLM-Soluziona Research and Development Institute*  
*Ronda de Toledo s/n – 13004 Ciudad Real, Spain*  
*{emverbo, icaballerom}@soluziona.com*

Mario Piattini  
*ALARCOS Research Group*  
*Information Systems and Technology Department*  
*UCLM-Soluziona Research and Development Institute*  
*Paseo de la Universidad 4 s/n – 13071 Ciudad Real, Spain*  
*Mario.Piattini@uclm.es*

Keywords: Data Quality, Data Quality Dimensions, Data Quality measures, XML, quality attributes.

Abstract: Traditionally, data quality management has mainly focused on both data source and data target. Increasingly, data processing to get a data product need raw data typically distributed among different data sources. However, if data quality is not preserved when transmitted, resulting data product and consequent information will not be of much value. It is necessary to improve exchange methods and means to get a better information process. This paper focus on that issue, proposing a new approach for assuring and transmitting data quality in the interchange. Using XML and related technologies, a document structure that considers data quality as a main topic is defined. The resulting schema is verified using several measures and comparing it to the data source.

## 1 INTRODUCTION

Nowadays, organizations structure is usually spread on different locations, which requires to distribute organizational data storage in order to achieve better performance. On the other hand, Service Oriented Architectures are being consolidated. These services typically provide data in an XML format in order to be easily transmitted. Both scenarios suppose new challenges as data replication and data integration. One of our concerns is to study these challenges from the point of view of data quality.

In (Strong, 1997) the ten main problems for data quality are outlined and justified. Two of them are directly involved in distributed systems: a) multiple sources of the same data produce different values and b) distributed heterogeneous systems lead to inconsistent definitions, formats, and values.

Last years, XML has been intensively used up to become the main standard technology for data exchanging between distributed systems. Using

XML, structured documents can be described, making their retrieval more efficient and effective. We could use it as a means/media for assessing and improving information quality, taking advantage of its related technologies as XSLT for easy processing of XML documents, and the restrictions model of XML Schema to define correct value patterns.

### 1.1 Data Quality Issues

A typical example of a data quality problem consists of having different values for the same data stored in several sources. Suppose a decision-support system which access to those sources and analyze them. With no additional information, the system only know that the data has different values but it cannot decide which is the correct one. If the system could assess the quality of each source and only accept those over a threshold, it would process the data with higher quality and, in consequence, produce better results.

There are a lot of researching lines trying to explain what data quality is. Most of authors have drawn the conclusion that “fitness for use” is probably the best definition for the term. Fitness implies that a set of special and specific characteristics must be observed in the data as to say it can be used to get sound information. These characteristics have been named as Data Quality Dimensions. (Strong, 1997) propose a set of them which could be tailored for a wide range of contexts.

From these Data Quality Dimensions, several measures have to be defined to get a quantitative idea of how good a piece of data is. These measures (see nomenclature about SMO in (García, 2005)), also called metrics, must be defined to rightly manage data quality. (Lee, 2006) presents some of these measures. A data quality management team must use those measures to improve the data quality. The most used methodology for this goal is TDQM, mainly described in (Wang, 1998).

### 1.2 Addressing the Problems

Once brought to the context the main foundations of data quality, we want to address the problem we have posed. Supposing that data is currently stored in a database, and several data quality dimensions and measures have been defined. The problem is *How can data quality be assured when data flows from sources to targets?*

The answer comes with the technology on which data product mainly trips from sources to targets: XML. So the idea is to create an XML structure that can give the necessary support for transmitting data quality concerns used in the source to the targets in order to be used them to maintain the quality of the

data products.

The remainder of the paper is structured as follows: section 2 present the main foundations of our work and design keys to elaborate our proposal. Section 3 shows several training examples of the proposal. Finally, section 4 outcomes several conclusions and future researching lines.

## 2 DQXSD: AN XML SCHEMA FOR DATA QUALITY

Data quality, as being quality, can be also studied from two points of view:

- **Expected Data Quality:** users expect that raw data and product data have a set of data quality dimensions like accuracy, free-of-error and so on. This kind of quality could almost always be evaluated without user interaction, for instance from metadata and quantity of stored data for each data set.
- **Required Data Quality:** users need and require that raw data and product data present specific and context-dependant characteristics which can be only evaluated by taking into account the judgement of the user, so user interaction is required, for instance, to provide a value which can be used as a basis or threshold to determine whether a data is good or not. The provided values must have been stored anywhere, and it could be necessary to be transmitted together with the data product. The way in which databases must be prepared to accept these values is given for the Data Quality Requirements in (Wang, 1995).

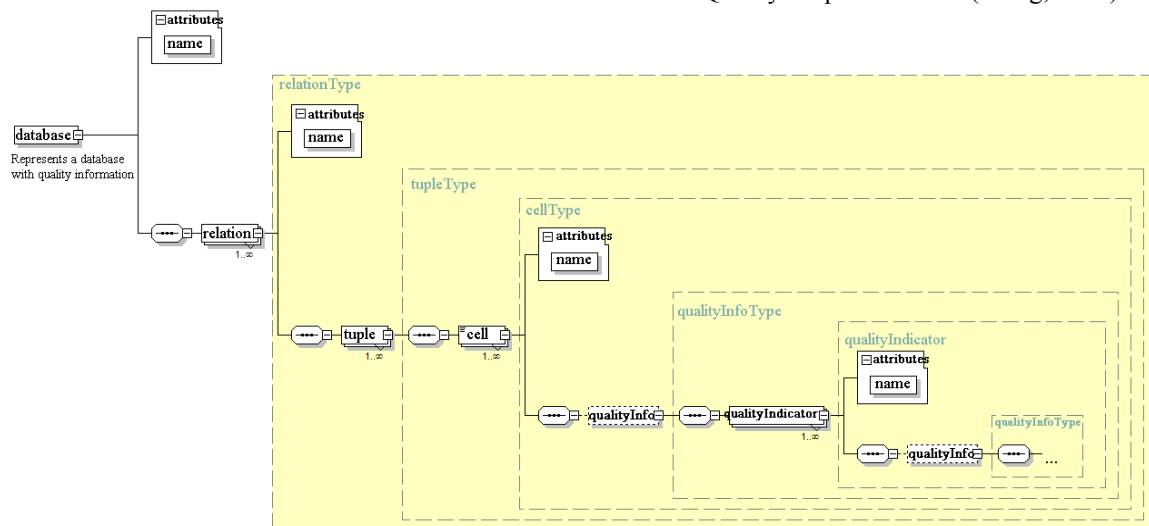


Figure 1: DQXSD Structure.

In order to grasp the special needs of XML documents quality, we are going to use more specific terms from now on:

- **External quality:** properties related to required data quality. It answers the question of what data is exchanged. Usually, an external agent must provide the information needed to address this issue.
- **Internal quality:** it deals with expected data quality and answers the question of how data is exchanged, i.e., the XML document structure. It is usually assessed applying measures on the document.

## 2.1 External Quality

In (Wang, 1995) the relational model is extended with an attribute-based data model to store data related to specific data quality dimensions which can let organizations achieve higher data quality. Since the attribute value of a cell is the basic unit of manipulation, it is necessary to tag quality information at the cell level.

Its principles are built over the notion of quality indicator. A quality indicator provides objective information about the characteristics of data and its manufacturing process.

It develops a mechanism to facilitate the linkage between an attribute and its immediate quality indicators. This mechanism is developed through the *quality key* concept. An attribute in a relation scheme is expanded into an ordered pair, called a *quality attribute*, consisting of the *attribute* and a *quality key*. The quality key is a reference to the underlying quality indicators. It also allows to detail a quality indicator linking it to a set of quality indicators. To achieve data integrity, an attribute value and its corresponding quality indicators must be treated as an atomic unit.

Suppose an organization has a database schema with associated quality information and they want to interchange its content with their partners. It would be much easier if there were a standard format to do so. In this point is when DQXML comes into play. Our work provides support for that model in XML.

Since XML is the preferred technology for data exchanging, it becomes the optimum choice. DQXML structure is defined with XML Schema and captures the requirements of the attribute-based approach. Figure 1 shows graphically the structure of the schema.

Table 1: Database and DQXML measures comparison.

Measure meaning	Database measure name	DQXML measure name
Effort necessary to retrieve all the information related to a component	DRT	DDQT
Information fragmentation degree	RD	RD
Quantity of information about a component directly accessible.	NA	NA
Cohesion of the system	COS	COS
Wasted communication bandwidth	-	NEE, NEA
Size of the system	-	NN
Number of references between the components of the system	-	NArc
Structural complexity of the system	-	SC <sub>XML</sub>

The main element is *database*, which represents the whole database we want to exchange. It is composed by a sequence of *relations* elements. Each *relation* models a table of the DB scheme and is, in turn, formed by a set of *tuples*, which are divided in *cells*. In addition to its respective values, inside a *cell* there can be a *qualityInfo* element that specifies quality information and is detailed with a set of *qualityIndicator* subelements that model a single quality indicator value. Note that a quality indicator can have nested quality information as well.

This schema defines what we have called **XSD for Data Quality (DQXSD)**. It preserves external quality as data associated with quality indicators is carefully treated in the document scheme. Moreover, it is especially useful in the sense that there are already several database management systems with XML support in their queries. They also allows to specify an XSD that defines the structure that the results must follow so the translation of the data stored in the database to DQXSD structure would be quite immediate.

We would like to highlight the fact that quality indicators are intended to contain useful data for assessing data quality on a certain quality dimension. The purpose of this model is acquisition and transmission of data. It should not include any measurement result because measurement methods may vary depending on the role of the system user

and its context. We will treat measurement in the next section.

From now on, we will write DQXML to refer to an XML document validated against DQXSD.

## 2.2 Internal Quality

After extracting database records according to the DQXSD structure, we obtain an XML document that preserves external quality from the original data source but, *what happens to the efficiency and accuracy of the data representation, i.e., internal quality?* In order to assess this issue objectively we have applied a measurement approach.

In (Ivan, 1998) a general definition for data measures is given along with examples of use. (Piattini, 2001) proposes some internal measures to measure relational databases which influence its complexity. Centered in XML documents, in (Diaz, 2003) a set of measures are proposed and implemented in a measurement tool. However, there is not much research work in data quality measurement oriented to XML documents.

Our idea is to adapt validated database measures to XML documents and compare them in order to demonstrate that the results are similar. In (Piattini, 2001) and (Calero, 2001) the following measures are defined:

- *Depth of the referential tree*: the DRT of a table A ( $DRT(A)$ ), is the length of the longest referential path from the table A, counted as the number of arcs on the path and considering cycles only once.
- *Referential degree*: the RD of a table A ( $RD(A)$ ), is the number of foreign keys in the table A.
- *Number of attributes*: the NA of a table A ( $NA(A)$ ), is the number of attributes of the table A.
- *Cohesion of the Schema*: the COS of a schema S, is the sum of the square of the number of tables in each not connected component in the schema graph.

We have adapted those measures to XML documents with DQXSD structure:

- a. *Depth of the DQXML tree*: the DDQT of a DQXML ( $DDQT(D)$ ), is the number of nested *qualityInfo* elements inside a *cell* plus one. It is equivalent to DRT for relational databases.
- b. *Referential degree*: the RD of a DQXML ( $RD(D)$ ), is the number of *qualityInfo* subelements that *cell* elements of a *relation* contains.

- c. *Number of attributes*: the NA of a *relation* element R in a DQXML ( $NA(R)$ ), is the maximum number of *cell* elements that each *tuple* element contains.
- d. *Cohesion of the Schema*: the COS of a DQXML ( $COS(D)$ ), is the square of the number of *relation* elements that are not connected in the schema graph.

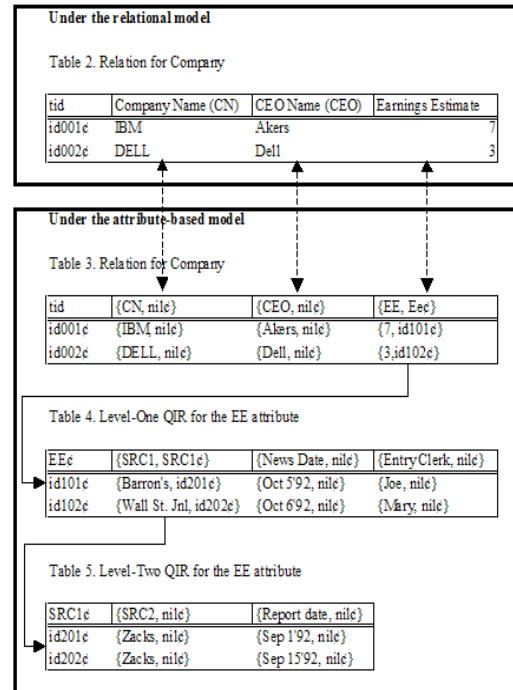


Figure 2: Database with quality information - taken from (Wang, 1995).

In addition to the previous measures, some specific ones for XML from (Díaz, 2003) has been included:

- e. *Number of empty elements*: the NEE of an element A ( $NEE(A)$ ) is the number of empty elements that are children of the element A.
- f. *Number of empty attributes*: the NEA of an element A ( $NEA(A)$ ) is the number of empty attributes of the element A.
- g. *Number of nodes*: the NN of a DQXML ( $NN(D)$ ) is the number of nodes needed to represent the document graph considering as a node any element, attribute or element value.
- h. *Number of arcs*: the NArc of a DQXML ( $Narc(D)$ ) is the number of arcs needed to represent the document graph. An arc is a relation between parent and children elements, element attributes and element values.
- i. *Structural complexity*: the  $SC_{XML}$  of a DQXML is:

$$SC_{XML} = N_{Arc} - NN + 1 \quad (1)$$

Taking advantage of XML related technologies, XSLT can be applied to calculate many of this measures, simplifying the construction of a management tool. Consequently, a second XML document with the measurement results would be generated for later viewing or processing. The most remarkable benefits of this approach are portability, interoperability and programming language independence.

A brief summary of the measures applied to databases and to DQXMLs is shown in Table 1.

### 3 EXAMPLES

#### 3.1 External Quality

To illustrate DQXSD usage, we have borrowed the theoretical training example shown in Figure 2 from (Wang, 1995) with the aim of adapting it to our model.

Suppose an organization has a database schema that contains a table like Table 2 (Tables 2-5 are embedded in Figure 2) with attributes like company name, CEO name and earnings estimate. Data may be collected over a period of time and come from a variety of sources. If the organization wanted to assess the believability of the data, the previous database schema should be adapted to the new quality requirements.

As a result, the original Table 2 is expanded into Table 3, which consists of the ordered pairs ( $\{CN, nil\}$ ,  $\{CEO, nil\}$ ,  $\{EE, EE\}$ ). The “ $nil$ ” indicates that no quality indicators are associated with attributes CN and CEO; whereas  $EE\phi$  indicates that EE has quality indicators associated.

Table 4 is a quality indicator relation for the attribute *Earnings Estimate* in Table 3 and Table 5 is a quality indicator relation for *SRC1* in Table 4.

First of all, we translate Table 3 to the DQXSD format without including quality indicators (text in normal font in Figure 3). In the resulting DQXML, there is only one *relation* element, which contains the two tuples that specifies companies data.

Later, we include the first level of quality indicators inside the *cell* element named “EE” (text in bold and italic in Figure 3). And, finally, we include the second level of quality indicators into the first level quality indicator *SRC1* (text in italic in

Figure 3). For shortening, only one tuple and one cell have been included in Figure 3.

#### 3.2 Internal Quality

The results of applying the measures explained in Section 2.2 can be consulted in Table 6.

First of all, we can see that equivalent measures DRT and DDQT have a similar value, 3.

```

<database>
  <relation>
    <tuple>
      <cell name="EE">
        7
        <qualityInfo>
          <qualityIndicator
            name="SRC1">
            Barron's
            <qualityInfo>
              <qualityIndicator
                name="SRC2">
                Zacks
              </qualityIndicator>
              <qualityIndicator
                name="Report
Date">
                1992/09/1
              </qualityIndicator>
            </qualityInfo>
          </qualityIndicator>
          <qualityIndicator
            name="News Date">
            1992/10/5
          </qualityIndicator>
          <qualityIndicator
            name="entryClerk">
            Joe
          </qualityIndicator>
        </qualityInfo>
      </cell>
      ...More cell elements...
    </tuple>
    ...More tuple elements...
  </relation>
</database>

```

Figure 3: DQXML with complete quality data.

For RD, the difference is in the measure method. For the quality database, we have three values: the result is 2 for the table “Company”; the result is also 2 for the table with the first level of quality indicators for attribute EE; and the result is 0 for the table with the second level of quality indicators for attribute EE. However, for DQXML we only have one value, 4, for the entire document. This occurs

because the quality information embedded in cell elements of the DQXML is divided in three tables in the database. If we put together the values of every table of the schema, we get the same result, 4.

For the measure *number of attributes* the result is different as well. The explanation is that in the database, the storage medium does not differentiate between quality and raw data while DQXSD treats quality data adding semantic value that it did not have when stored in a database.

Table 6. Measurements results

Measure	Relational database	DQXML
DRT	3	
DDQT		3
RD	RD(Comp)=2 RD(E <sub>E1</sub> )=2 RD(E <sub>E2</sub> )=0	RD(R <sub>1</sub> )=4
NA	NA(Comp)=4 NA(E <sub>E1</sub> )=4 NA(E <sub>E2</sub> )=3	NA(D)=4
COS	0	0
NEE	-	0
NEA	-	0
NN	-	73
NArc	-	72
SC <sub>XML</sub>	-	0

The results of the measures *NEE* and *NEA* shows that the DQXML has high quality because it has no empty elements or attributes that waste bandwidth. Lastly, *NN* and *NArc* with their low values indicates that the DQXML has no excessive complexity, statement confirmed by *SC<sub>XML</sub>*.

## 4 CONCLUSIONS AND FUTURE WORKS

Traditionally, data quality has been only applied to data stored in databases as being raw data for manufacturing data products. This approach is clearly out of date because data exchanging is continuously getting more important in parallel to the consolidation of Service Oriented Architectures.

Static data quality issues must also be propagated when transmitted. To give the necessary support to this goal, we define a new document structure, DQXSD based on the most important technology for information exchanging, XML. To define it, XML Schema is used.

DQXSD helps to capture quality data stored in a database schema and translate it to a proper format ready to be transmitted.

To prove the data quality preservation through that process, several measures for DQXML documents have been developed and compared to database equivalents getting satisfactory results.

Although the results presented in this paper are oriented to capture quality data stored in relational databases, DQXSD could be easily adapted to other storage models due to the flexibility of the technologies used for its definition.

## ACKNOWLEDGMENTS

This research is part of the FAMOSO and ESFINGE projects supported by the Dirección General de Investigación of the Spanish Ministerio de Ciencia y Tecnología (Ministry of Science and Technology)(TIC2003-07804-C05-03).

## REFERENCES

- Calero, C., Piattini, M. & Genero, M., 2001. *Metrics for controlling Databases Complexity*, Becker, S.
- Díaz, E., 2003. *Herramienta para la gestión de métricas en documentos XML*. Departamento de Tecnologías y Sistemas de Información, Escuela Superior de Informática de Ciudad Real, Universidad de Castilla-La Mancha.
- Fran, W. & Simeon, J., 2003. Integrity constraints for XML. *Journal of Computer and System Sciences*.
- García, F., Bertoa, M. F., Calero, C., Vallecillo, A., Ruiz, F., Piattini, M. & Genero, M., 2005. Toward a consistent terminology for software measurement. *Information and Software Technology*, 48, 631-644.
- Ivan, I., Parlog, O., Oprea, P., Nosca, G. & Ivan, A.-A., 1998. Data Metrics. In *IQ 1998, Conference on Information Quality*.
- Klettke, Shneider, M. L. & Heuer, A., 2002. *Metrics for XML Document Collections*. Database Research Group, University of Rostock, Germany.
- Lee, Y. W., Pipino, L. L., Funk, J. D. & Wang, R. Y., 2006. *Journey to Data Quality*, The MIT Press.
- Piattini, M., Calero, C. & Genero, M., 2001. Table Oriented Metrics for Relational Databases. *Software Quality Journal*.
- Strong, D. M., Lee, Y. W. & Wang, R. Y., 1997. Data Quality in Context. *Communications of the ACM*.
- Strong, D. M., Lee, Y. W. & Wang, R. Y., 1997. 10 Potholes in the Road to Information Quality. *IEEE Computer*.
- Wang, R. Y., 1998. A Product Perspective on Total Data Quality Management. *Communications of the ACM*.
- Wang, R. Y., Reddy, M. P. & Kon, H. B., 1995. Toward quality data: An attribute-based approach. *Decision Support Systems*.



Proceedings of the  
9th International Conference on Enterprise Information Systems - ICEIS 2007  
ISBN: 978-972-8865-88-7  
<http://www.iceis.org>