

Sixteen

RY

me

ing

sored by
Engineering F

Technical co-sponsor

Computer Society Technical Council on Software

IEE

SOC

EE

Andy Zait
Giuliano Antonior
Stéphane Ducasse

Proceedings

**16th Working Conference
on Reverse Engineering (WCRE 2009)**

Proceedings

16th Working Conference on Reverse Engineering (WCRE 2009)

13th – 16th October 2009 – Lille, France

Edited by

Andy Zaidman, Giuliano Antoniol and Stéphane Ducasee

Sponsored by



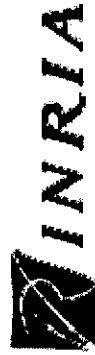
Reengineering Forum

Technical Co-sponsor



IEEE-CS Technical Council on
Software Engineering

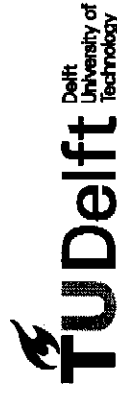
In cooperation with



INRIA Lille-Nord Europe



SOCCERLAB
Software Cost-effective Change
and Evolution Research Lab



Delft
University of
Technology



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 meetings 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 P3867
BMS Part Number CFP09090-PRT
ISBN 978-0-7695-3867-9
ISSN Number 1095-1350

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

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

Individual paper REPRINTS may be ordered at: <reprints@computer.org>

Editorial production by Bob Werner

Cover art production by Joe Daigle/Studio Productions

Printed in the United States of America by The Printing House

IEEE
computer
society



IEEE Computer Society
Conference Publishing Services (CPS)

<http://www.computer.org/cps>

2009 16th Working Conference on Reverse Engineering

WCRE 2009 Table of Contents

Message from the General Chair.....	ix
Message from the Program Chairs.....	x
Organizing Committee.....	xii
Steering Committee.....	xiii
Program Committee.....	xiv
Additional Reviewers.....	xv

Keynotes

Beyond the Lone Reverse Engineer: Insourcing, Outsourcing and Crowdsourcing	3
<i>Margaret-Anne D. Storey</i>	
Legacy and Future of Data Reverse Engineering	4
<i>Jean-Luc Hainaut</i>	

WCRE 1999 Most Influential Paper

Ten Years Later, Experiments with Clustering as a Software Remodularization Method	7
<i>Nicolas Anquetil and Timothy C. Lethbridge</i>	

Session I – Mining Software Repositories

Who are Source Code Contributors and How do they Change?	11
<i>Massimiliano Di Penta and Daniel M. German</i>	
A Study of the Time Dependence of Code Changes	21
<i>Omar Alam, Bram Adams, and Ahmed E. Hassan</i>	
Relating Identifier Naming Flaws and Code Quality: An Empirical Study	31
<i>Simon Butler, Michel Wermelinger, Yijun Yu, and Helen Sharp</i>	
Techniques for Identifying the Country Origin of Mailing List Participants	36
<i>Ran Tang, Ahmed E. Hassan, and Ying Zou</i>	

Session II – Dynamic Analysis

NTrace: Function Boundary Tracing for Windows on IA-32	43
<i>Johannes Passing, Alexander Schmidt, Martin von Löwis, and Andreas Polze</i>	
Recovering Views of Inter-System Interaction Behaviors	53
<i>Christopher Ackermann, Mikael Lindvall, and Rance Cleaveland</i>	
Mining Quantified Temporal Rules: Formalism, Algorithms, and Evaluation	62
<i>David Lo, Ganesan Ramalingam, Venkatesh Prasad Ranganath, and Kapil Vaswani</i>	

Session III – Empirical Software Engineering

An Exploratory Study of the Impact of Code Smells on Software Change-proneness	75
<i>Fouise Khomh, Massimiliano Di Penta, and Yann-Gaël Guéhéneuc</i>	
An Empirical Study on Inconsistent Changes to Code Clones at Release Level	85
<i>Nicolas Bettenburg, Weyi Shang, Walid Ibrahim, Bram Adams, Ying Zou, and Ahmed E. Hassan</i>	
Lexicon Bad Smells in Software	95
<i>Surafel Lemma Abebe, Sonia Haiduc, Paolo Tonella, and Andrian Marcus</i>	

Session IV – Remodularization and Reengineering

Automatic Package Coupling and Cycle Minimization	103
<i>Hani Abdeen, Stéphane Ducasse, Houari Sahraoui, and Ilham Alloui</i>	
Identifying Cycle Causes with Enriched Dependency Structural Matrix	113
<i>Jannik Laval, Simon Denier, Stéphane Ducasse, and Alexandre Bergel</i>	
The Logical Modularity of Programs	123
<i>Daniel Ratiu, Radu Marinescu, and Jan Jürjens</i>	
On the Use of ADM to Contextualize Data on Legacy Source Code for Software Modernization	128
<i>Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Orlando Ávila-García, and Mario Piattini</i>	

Session V - Change and Defect Proneness

On the Relationship Between Change Coupling and Software Defects	135
<i>Marco D'Ambros, Michele Lanza, and Romain Robbes</i>	
Tracking Design Smells: Lessons from a Study of God Classes	145
<i>Stéphane Vaucher, Fouise Khomh, Naouel Moha, and Yann-Gaël Guéhéneuc</i>	
Bug-Inducing Language Constructs	155
<i>Javed Ferzund, Syed Nadeem Ahsan, and Franz Wotawa</i>	
Design Patterns and Change Proneness: A Replication Using Proprietary C# Software	160
<i>Matt Gatrell, Steve Counsell, and Tracy Hall</i>	

Session VI – Static Analysis and Security

Automatic Static Unpacking of Malware Binaries	167
<i>Kevin Coogan, Saumya Debray, Tasneem Koochar, and Gregg Townsend</i>	
Computing the Structural Difference between State-Based Models	177
<i>Kirill Bogdanov and Neil Walkinshaw</i>	
Extraction of Inter-procedural Simple Role Privilege Models from PHP Code	187
<i>Dominic Letarte and Ettore Merlo</i>	

Session VII – Traceability

Traceability Recovery Using Numerical Analysis	195
<i>Giovanni Capobianco, Andrea De Lucia, Rocco Oliveto, Annibale Panichella, and Sebastiano Panichella</i>	
Benchmarking Lightweight Techniques to Link E-Mails and Source Code	205
<i>Alberto Bacchelli, Marco D'Ambros, Michèle Lanza, and Romain Robbes</i>	
Domain Feature Model Recovery from Multiple Applications Using Data Access Semantics and Formal Concept Analysis	215
<i>Yiming Yang, Xin Peng, and Wenyun Zhao</i>	

Session VIII - Program Comprehension

Characterizing Evolutionary Clusters	227
<i>Adam Yanya, Steven Klusener, Nico van Rooijen, and Hans van Yliet</i>	
Autumn Leaves: Curing the Window Plague in IDEs	237
<i>David Roethlisberger, Oscar Nierstrasz, and Stéphane Ducasse</i>	
Constructing a Resource Usage View of a Large and Complex Software-Intensive System	247
<i>Trosky Boris Callo Arias, Pierre America, and Paris Avgeriou</i>	

Session IX – Static Analysis

Static Detection of Disassembly Errors	259
<i>Nithya Krishnamoorthy, Saumya Debray, and Keith Fligg</i>	
Reverse Engineering Sequence Diagrams for Enterprise JavaBeans with Business Method Interceptors	269
<i>Alexander Serebrenik, Serguei Roubisov, Ella Roubisova, and Mark van den Brand</i>	
Computing Structural Types of Clone Syntactic Blocks	274
<i>Ettore Merlo and Thierry Lavoie</i>	
Reverse Engineering Existing Web Service Applications	279
<i>Houda El Bouhissi and Mimoun Malki</i>	

PhD Forum

Supporting Feature-Level Software Maintenance	287
<i>Meghan Revelle</i>	
Enabling the Evolution of J2EE Applications through Reverse Engineering and Quality Assurance	291
<i>Fabrizio Perin</i>	
Approximate Graph Matching in Software Engineering	295
<i>Ségla Kpodjedo</i>	
Evolving Software Systems Towards Adaptability	299
<i>Mehdi Amoui</i>	
SQUAD: Software Quality Understanding through the Analysis of Design	303
<i>Foutse Khomh</i>	

Tool Demonstrations

- PRECISO: A Reverse Engineering Tool to Discover Web Services from Relational Databases309
Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Ismael Caballero, Macario Polo, and Mario Piattini
- Recovering Class Models Stereotyped with Crosscutting Concerns311
Heitor Augustus Xavier Costa, Paulo Afonso Parreira Junior, Valtier Vieira de Camargo, and Rosângela Aparecida Dellosso Penteado
- SHINOBI: A Tool for Automatic Code Clone Detection in the IDE313
Shinji Kawaguchi, Takanobu Yamashina, Hidetake Uwano, Kyohei Fushida, Yasutaka Kamei, Masataka Nagura, and Hajimu Iida
- Enhancing Quality of Code Clone Detection with Program Dependency Graph315
Yoshiki Higo and Shinji Kusumoto
- JavaCompExt: Extracting Architectural Elements from Java Source Code317
Nicolas Anquetil, Jean-Claude Royer, Pascal André, Gilles Ardourel, Petr Hnětýnka, Tomáš Poch, Dragoş Petrascu, and Vladieiţa Petrascu
- ConAn: A Tool for the Identification of Crosscutting Concerns in Object Oriented Systems Based on Type Hierarchy Analysis319
Mario Luca Bernardi and Giuseppe Antonio Di Lucca

Workshops

- R.E.M. 2009 - International Workshop on Reverse Engineering Models from Software Artifacts323
Leon Moonen and Tarja Systä
- FAMOOSt 2009 - Workshop on FAMIX and Moose in Software Reengineering325
Simon Denier and Tudor Gîrba

- Author Index**327

Message from the Program Chairs

Welcome to WCRE, the Working Conference on Reverse Engineering! This year's conference is the most competitive in WCRE's history. The large number of high-quality submissions made the task of selecting papers very challenging. There were many good papers that we could unfortunately not include in the program, yet we hope that the authors will take advantage of the reviewers' feedback to make improvements. All submissions went through a rigorous review process as every paper was reviewed in detail by at least three Program Committee members from Industry and Academia. All decisions were open for discussion by the entire Program Committee before finalizing them. In the end, we selected 20 full papers (25.3% acceptance rate) and 11 short papers (39.2% acceptance rate) from a total of 79 submissions, originating from 27 countries.

Authors of the best papers will be invited to submit extended versions of their work to a special issue of the *Science of Computer Programming (SCP)*, published by Elsevier.

Given that WCRE is a working conference, papers that are likely to stir discussion and which touch upon non-traditional topics are encouraged and welcomed. Attendees of WCRE want to learn about the exciting and upcoming research opportunities in reverse engineering, not only the past innovative results.

Short papers help touch on important and up-coming areas in our growing and evolving community. Short papers are important to the community and that is why we are not putting the short papers in a separate session, but integrate them into the technical program, creating sessions that contain both short and full papers. This year's program includes sessions on exciting topics such as dynamic analysis, modularization and reengineering, program comprehension or empirical software engineering.

As always, WCRE technical sessions foster paper presentations followed by an open and energetic discussion. A trademark of WCRE is that there is much more than a recital of papers – there is vigorous and open discussion so please contribute your ideas and be part of a WCRE tradition!

We would like to extend our sincere thanks to the two keynote speakers; we are very fortunate this year to have great keynotes given by two top software engineering researchers: Dr. Margaret-Anne Storey from the University of Victoria ("*Beyond the Lone Reverse Engineer: Insourcing, Outsourcing and Crowdsourcing*"), and Jean-Luc Hainaut from the Facultés Universitaires Notre-Dame de la Paix, Namur ("*Legacy and Future of Data Reverse Engineering*").

WCRE is one of the highly respected and well-cited conferences in Software Engineering. And over the past 16 years our community has come up with some great ideas and innovations. That is why, since WCRE 2008, we are looking back at what the reverse engineering community did in the past, by awarding the most influential paper award from 10 years ago.

This year we are looking back at WCRE 1999, held in Atlanta, Georgia, USA, and with a large consensus, the WCRE 2009 Program Committee elected the WCRE 1999 paper titled "*Experiments with Clustering as a Software Remodularization Method*" as the most influential paper. Nicolas Anquetil and Timothy Lethbridge have provided enormous contributions to our community during their career. We look forward to hear Nicolas Anquetil and Timothy Lethbridge reflect on their previous work in their talk entitled "*Ten Years Later: Experiments with Clustering as a Software Remodularization Method*."

The conference program also includes a tool demonstration track featuring six tool demos, a Ph.D. symposium where five promising young reverse engineering scientists will present their work. The WCRE program this year also includes two workshops: the *International Workshop on Reverse Engineering Models from Software Artifacts (R.E.M. 2009)* and *FAMOOSr: 2009 Workshop on FAMIX and Moose in Software Reengineering*.

These events will be highly interactive and build the basis for strong research works and new collaborations among the participants. We would like to thank the organizer of these workshops for choosing to co-locate with

Message from the General Chair

Welcome to Lille, France, for the 16th international Working Conference on Reverse Engineering! This is a real and personal pleasure to get WCRE to Lille and this for several reasons.

First, because Lille is a crossroad in Europe, crossroad between France, Belgium and the UK and to a larger circle with the western part of Germany and the Netherlands which are reachable in a couple of hours.

The second reason is that since 2008 INRIA has established a new research laboratory at Lille: INRIA Lille Nord Europe. INRIA is (one of) the top research institutes in applied mathematics and computer science in France. Within the INRIA Lille Nord Europe laboratory, INRIA supported the creation of the RMoD Team (remodularization and modular software). The RMoD team is one of the too few teams in France working on reverse engineering and reengineering. Our team is working in two areas: reengineering and new dynamic languages (traits, for example, which are used in SUN Microsystem's Fortress new scientific high performance language and may also be included in the next Javascript standard). We are contributing to a new open-source and modular object-oriented language named Pharo (<http://www.pharo-project.org>), descendant of Smalltalk. On the reengineering side we are continuously improving the Moose open-source reengineering environment, whose development started in 1996 and of which we rebuilt a large part last year (<http://moose.unibe.ch/>).

Third, Lille is located in the "Industries du commerce" region, meaning that large companies such as Auchan, Carrefour or Decathlon have their headquarters in the area of Lille. And as we know, such companies have legacy systems. Now in this moment where the society is getting more and more concerned about ecology, we have a new challenge: how can we help them producing software whose creation and evolution is preserving our environment? We should show that reuse and maintenance are CO² free. In the past in some wealthy and large companies, project managers could just allocate 50 software engineers to rebuild software that already existed. We need to open new spaces to define CO² free software.

I would like to thank the program co-chairs, Andy Zaidman and Giuliano Antoniol, the Ph.D. Forum chairs, Rainer Koschke and Spiros Mancoridis, the Workshop chairs, Yann-Gaël Guéhéneuc and Damien Pollet, the Tool Demonstration chairs, Martin Pinzger and Bram Adams, and all the members of the program committee for their good job assembling an exciting program.

I would like to thank Marie-Bénédicte Demoncourt who took care of the local organization. Since none of the RMod team members are from Lille or North of France we had no clue about what to propose to you and organize but Marie-Bénédicte was there.

Particular thanks to all our sponsors: The Reengineering Forum and the supporting organizations: INRIA, TU Delft, Soccer Lab of the École Polytechnique de Montréal and IEEE Computer Society's Technical Council on Software Engineering. I would like to thanks Cadbury and Caroline Hirsbein-Frommer for the excellent Carambars they nicely offered to us! Last but not least much gratitude is owed to Elliot Chikofsky for his advisory work on organizing WCRE this year.

Finally I hope that you will really enjoy the lively presentations, workshops, the city and the Carambars, of course.

Stéphane Ducasse
WCRE 2009 General Chair

WCRE this year and we trust that attendees will benefit from increased opportunity of interaction and discussion that these workshops offer.

We would like to thank all the people who helped us to organize this program: Rainer Koschke and Spiros Mancoridis (Ph.D. Forum Chairs), Yann-Gaël Guéhéneuc and Damien Pollet (Workshop Chairs), Martin Pinzger and Bram Adams (Tool Demonstration Chairs), Elliot Chikofsky (Finance and Registration Chair, and father of WCRE), Chanchal Roy (Publicity Chair), and Bob Werner from the IEEE Computer Society for managing the publication process. We would especially like to thank the Program Committee and the additional reviewers. They provided detailed and constructive reviews in time and actively participated in the discussion. Last, but not least, we thank the authors for submitting to WCRE 2009 a set of very high quality papers.

Enjoy the conference!

Andy Zaidman and Giulio Antoniol
WCRE 2009 Program Chairs

Organizing Committee

General Chairs

Stéphane Ducasse – *INRIA Lille, France*

Program Chair

Andy Zaidman – *Delft University of Technology, The Netherlands*
Giuliano Antoniol – *École Polytechnique de Montréal, Canada*

PhD Forum Chairs

Rainer Koschke – *University of Bremen, Germany*
Spiros Mancoridis – *Drexel University, Philadelphia, USA*

Workshop Chairs

Yann-Gaël Guéhéneuc – *École Polytechnique de Montréal, Canada*
Damien Pollet – *University of Lugano, Switzerland*

Tool Demonstration Chairs

Martin Pinzger – *Delft University of Technology, The Netherlands*
Bram Adams – *Queen's University, Canada*

Finance and Registration Chair

Elliot Chikofsky – *Engineering Management & Integration, USA*

Publicity Chair

Chantal K. Roy – *Queen's University, Canada*

Local Arrangements Chair

Marie-Bénédicte Demoncourt – *INRIA Lille, France*

Web Chair

Ferdaous Boughanmi – *École Polytechnique de Montréal, Canada*

Steering Committee

Michael Blaha, *OMT Associates, USA*
Gerardo Canfora, *RCOST - University of Samnio, Italy*
Elliott Chikofsky, *Engineering Management & Integration, USA*
Arie van Deursen (*chair*), *Delft University of Technology, The Netherlands*
Massimiliano Di Penta, *RCOST - University of Samnio, Italy*
Stéphane Ducasse, *INRIA Lille, France*
Rainer Koschke, *Universität Bremen, Germany*
Andy Zaidman, *Delft University of Technology, The Netherlands*

Valued Alumni of the WCRE Steering Committee

Leon Moonen, *Simula Research Laboratory, Norway*
Richard Waters, *Mitsubishi Electric Research, USA*
Ira Baxter, *Semantic Designs, USA*
Chris Verhoef, *University of Amsterdam, Netherlands*
Alex Quilici, *University of Hawaii at Manoa, USA*
Linda Wills, *Georgia Institute of Technology, USA*
Melody Moore, *Georgia State University, USA*
Cristina Cifuentes, *Sun Microsystems Labs, USA*
Peter Aiken, *Institute for Data Research, USA*
James Cross, *Auburn University, USA*
Margaret-Anne Storey, *University of Victoria, Canada*

Program Committee

Ira Baxter, *Semantic Designs, USA*
Dirk Beyer, *Simon Fraser University, Canada*
Michael Blaha, *Modelsoft Consulting Corporation, Canada*
Michael Collard, *The University of Akron, USA*
Jim Cordy, *School of Computing, Queen's University, Canada*
Marco D'Ambros, *Faculty of Informatics, University of Lugano, Italy*
Andrea De Lucia, *University of Salerno, Italy*
Serge Demeyer, *University of Antwerp, Belgium*
Simon Denier, *INRIA Lille-Nord Europe, France*
Arie van Deursen, *Delft University of Technology, The Netherlands*
Massimiliano Di Penta, *University of Sannio - Benevento, Italy*
Beat Fluri, *University of Zurich, Switzerland*
Harald Gall, *University of Zurich, Switzerland*
Keith Gallagher, *Durham University, UK*
Mohammed El Ramly, *The British University in Egypt, Egypt*
Tudor Gîrba, *University of Bern, Switzerland*
Mike Godfrey, *University of Waterloo, Canada*
Mark Harman, *Kings College London, UK*
Ahmed Hassan, *Queen's University Kingston, Canada*
George E. Kalb, *Johns Hopkins University, USA*
Sung Kim, *The University of Hong Kong Science and Technology, Hong Kong*
Jens Knodel, *Fraunhofer ISE, Germany*
Jens Krinke, *Kings College London, UK*
Michele Lanza, *University of Lugano, Switzerland*
Marin Litoiu, *IBM Canada, Canada*
Jonathan Maletic, *Kent State University, USA*
Andrian Marcus, *Wayne State University, USA*
Radu Marinescu, *Politehnica University of Timisoara, Romania*
Nabor Mendonca, *University of Fortaleza, Brazil*
Ali Mesbah, *Delft University of Technology, The Netherlands*
Leon Moonen, *Simula Research Laboratory, Norway*
Hausi Muller, *University of Victoria, Canada*
Liam O'Brian, *National ICT Australia Limited (NICTA), Australia*
Rocco Oliveto, *University of Salerno, Italy*
Denys Poshyvanyk, *The College of William and Mary, USA*
Filipo Ricca, *University of Genova, Italy*
Martin Robillard, *McGill University, Canada*
Gregorio Robles, *Universidad Rey Juan-Carlos, Spain*
Kamran Sartipi, *McMaster University, Canada*
Susan Sim, *University of California, Irvine, USA*
Eleni Stroulia, *University of Alberta, Canada*
Tajja Systä, *Technical University of Tampere, Finland*
Ladan Tahvildari, *University of Waterloo, Canada*
Paolo Tonella, *Fondazione Bruno Kessler - IRST, Italy*
Vassilios Tzerpos, *York University, Canada*
Jingwei Wu, *Oracle, USA*
Ying Zou, *Queen's University, Canada*

Additional Reviewers

Manar Alalfi
Asil Almonaies
Nicolas Bettenburg
Eric Bouwers
Amancio Bouza
Bas Cornelissen
Barthélemy Dagenais
Ron Desmarais
Vincenzo Deufemia
Fausto Fasano
Malcom Gethers
Hamoun Ghanbari
Carmine Gravino
Michaela Greiler
Sonia Haiduc
Maen Hammad
Zef Hemel
Zhen Ming Jiang
Cory Kapsner
Holger Kienle
Patrick Knab
Jannik Laval
Mircea Lungu
Collin McMillan
Fabrizio Perin
Damien Pollet
Michele Risi
Chanchal Roy
Alexey Rudkovskiy
Giuseppe Scanniello
Bonita Sharif
Mark Shtern
Andrew Sutton
Aiko Fallas Yamashita
Anis Yousefi

PRECISO: A Reverse Engineering Tool to Discover Web Services from Relational Databases

Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Ismael Caballero, Macario Polo and Mario Piattini

Alarcos Research Group, University of Castilla-La Mancha
Paseo de la Universidad, 4 13071, Ciudad Real, Spain

{Ricardo.PdelCastillo, Ignacio.GRodriguez, Ismael.Caballero, Mario.Piattini}@uclm.es

Abstract—There is a real need for SOA principles, such as those offering software as services, in the software industry. In this effort, databases (one of the most important artefacts in Information Systems) can be also seen as a set of services offering access to the stored information. This paper presents PRECISO, a reverse engineering tool to discover and generate Web Services automatically from relational databases. PRECISO makes it possible to modernize legacy databases by introducing them in an SOA context by means of the generated services. This tool was used in a real-life case study in the context of a software company.

Keywords: Database Reverse Engineering, Web Services, MDA, SOA and Pattern Matching.

1. INTRODUCTION

Today, organizations are increasingly forced to share more and more information as part of the basic activity in their daily operations. However, the heterogeneity of Information Systems (IS) is growing every day due to the appearance of new technological paradigms, standards, and environments, making it more and more difficult to share information [3]. Due to these facts and in order to keep their competitiveness level throughout their IS, organizations must be involved in a process of continuous renewal. Therefore, IS developers must constantly and quickly develop and maintain their products in order to meet market requirements [3]. Among all the artefacts that compose information systems, databases are possibly one of the most important elements since they contain all the organizational information and form the basis of decision-making.

This paper presents PRECISO [1], a tool for database reverse engineering following the MDA (Model-Driven Architecture) principles [5] to extract Web Services (WS) that show the database as a set of services, offering easy access to the information. PRECISO offers several benefits: (i) it minimizes heterogeneity problems since databases can be integrated in SOA environments; (ii) it advocates the reuse of legacy databases, thus extending the lifecycle of databases; and (iii) it shortens development time because the WS generation is automatic and instantaneous.

II. THE RECOVERY PROCESS

Figure 1 depicts the database re-engineering process, which consists of three main activities broken down into several tasks. *DMR* is the first activity to create a PSM (*Platform-Specific Model*) which represents the input relational database. The *DMR-1* task recovers metadata from the database and builds the PSM according to an SQL-92 metamodel based on [2]. The *DMR-2* task simultaneously discovers the potential services by means of pattern matching [4]. Figure 2 shows the patterns recognized in the database schema and the associated service templates.

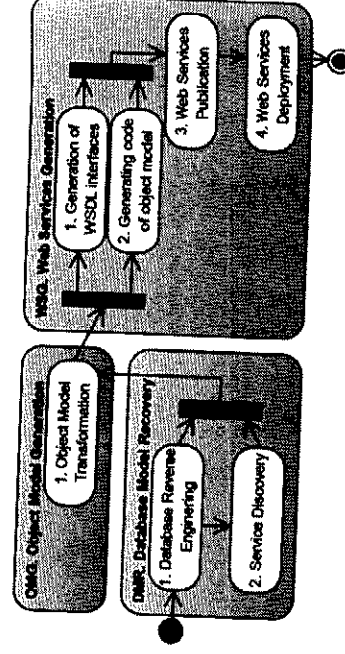


Figure 1. The Web Services recovery process

The second activity is *OMG*, which generates an object model from the previous database schema model. Thus, the *OMG-1* task transforms the PSM into a PIM (*Platform-Independent Model*) according to the UML2 metamodel [6]. Finally, the third activity is *WSG*, which generates the WS to manage the input database. The *WSG-1* task builds the service interfaces in WSDL (*Web Services Description Language*) from services discovered in *DMR-2*. This set of interfaces is the target PSM since it represents the WS platform. At the same time, *WSG-2* generates the source code related to the object model. This code supports the business layer of WS. Then *WSG-3* writes out the source code of the WS. After that, *WSG-4* carries out the Web Services deployment. Finally, the WS are ready to be used.

