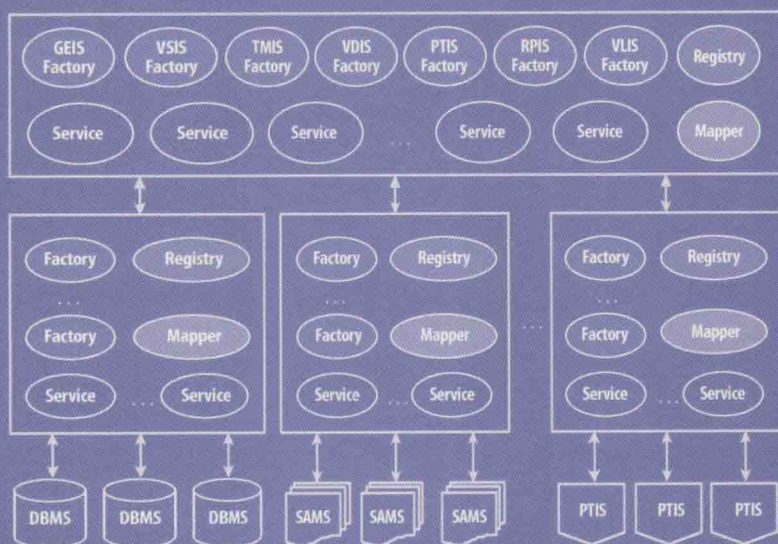


Pilar Herrero  
María S. Pérez  
Víctor Robles (Eds.)

# Scientific Applications of Grid Computing

First International Workshop, SAG 2004  
Beijing, China, September 2004  
Revised Selected and Invited Papers

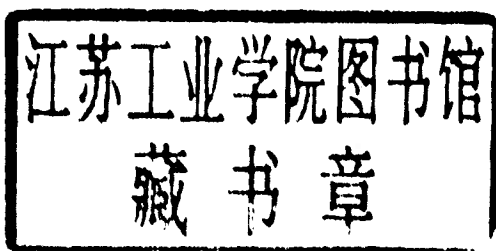


Springer

Pilar Herrero María S. Pérez  
Víctor Robles (Eds.)

# Scientific Applications of Grid Computing

First International Workshop, SAG 2004  
Beijing, China, September 20-24, 2004  
Revised Selected and Invited Papers



## Volume Editors

Pilar Herrero

María S. Pérez

Víctor Robles

Universidad Politécnica de Madrid

Facultad de Informática

Campus de Montegancedo S/N

28.660 Boadilla del Monte, Madrid, Spain

E-mail: {pherrero,mperez,vrobles}@fi.upm.es

Library of Congress Control Number: 2005925775

CR Subject Classification (1998): C.2, D.2.12, D.4.3-4, D.4.7, H.3, H.4, H.5.3

ISSN 0302-9743

ISBN-10 3-540-25810-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-25810-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

[springeronline.com](http://springeronline.com)

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper

SPIN: 11423287

06/3142

5 4 3 2 1 0

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*University of Dortmund, Germany*

Madhu Sudan

*Massachusetts Institute of Technology, MA, USA*

Demetri Terzopoulos

*New York University, NY, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Moshe Y. Vardi

*Rice University, Houston, TX, USA*

Gerhard Weikum

*Max-Planck Institute of Computer Science, Saarbruecken, Germany*

# Lecture Notes in Computer Science

For information about Vols. 1–3397

please contact your bookseller or Springer

- Vol. 3525: A.E. Abdallah, C.B. Jones, J.W. Sanders (Eds.), *Communicating Sequential Processes*. XIV, 321 pages. 2005.
- Vol. 3517: H.S. Baird, D.P. Lopresti (Eds.), *Human Interactive Proofs*. IX, 143 pages. 2005.
- Vol. 3516: V.S. Sunderam, G.D.v. Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2005*, Part III. LXIII, 1143 pages. 2005.
- Vol. 3515: V.S. Sunderam, G.D.v. Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2005*, Part II. LXIII, 1101 pages. 2005.
- Vol. 3514: V.S. Sunderam, G.D.v. Albada, P.M.A. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2005*, Part I. LXIII, 1089 pages. 2005.
- Vol. 3510: T. Braun, G. Carle, Y. Koucheryavy, V. Tsoulos (Eds.), *Wired/Wireless Internet Communications*. XIV, 366 pages. 2005.
- Vol. 3508: P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, M. Winikoff (Eds.), *Agent-Oriented Information Systems II*. X, 227 pages. 2005. (Subseries LNAI).
- Vol. 3503: S.E. Nikoletsas (Ed.), *Experimental and Efficient Algorithms*. XV, 624 pages. 2005.
- Vol. 3501: B. Kégl, G. Lapalme (Eds.), *Advances in Artificial Intelligence*. XV, 458 pages. 2005. (Subseries LNAI).
- Vol. 3500: S. Miyano, J. Mesirov, S. Kasif, S. Istrail, P. Pevzner, M. Waterman (Eds.), *Research in Computational Molecular Biology*. XVII, 632 pages. 2005. (Subseries LNBI).
- Vol. 3498: J. Wang, X. Liao, Z. Yi (Eds.), *Advances in Neural Networks – ISNN 2005*, Part III. L, 1077 pages. 2005.
- Vol. 3497: J. Wang, X. Liao, Z. Yi (Eds.), *Advances in Neural Networks – ISNN 2005*, Part II. L, 947 pages. 2005.
- Vol. 3496: J. Wang, X. Liao, Z. Yi (Eds.), *Advances in Neural Networks – ISNN 2005*, Part I. L, 1055 pages. 2005.
- Vol. 3495: P. Kantor, G. Muresan, F. Roberts, D.D. Zeng, F.-Y. Wang, H. Chen, R.C. Merkle (Eds.), *Intelligence and Security Informatics*. XVIII, 674 pages. 2005.
- Vol. 3494: R. Cramer (Ed.), *Advances in Cryptology – EUROCRYPT 2005*. XIV, 576 pages. 2005.
- Vol. 3492: P. Blache, E. Stabler, J. Busquets, R. Moot (Eds.), *Logical Aspects of Computational Linguistics*. X, 363 pages. 2005. (Subseries LNAI).
- Vol. 3489: G.T. Heineman, J.A. Stafford, H.W. Schmidt, K. Wallnau, C. Szyperski, I. Crnkovic (Eds.), *Component-Based Software Engineering*. XI, 358 pages. 2005.
- Vol. 3488: M.-S. Hacid, N.V. Murray, Z.W. Raś, S. Tsumoto (Eds.), *Foundations of Intelligent Systems*. XIII, 700 pages. 2005. (Subseries LNAI).
- Vol. 3483: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Lagana, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), *Computational Science and Its Applications – ICCSA 2005*, Part IV. XXVII, 1362 pages. 2005.
- Vol. 3482: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Lagana, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), *Computational Science and Its Applications – ICCSA 2005*, Part III. LXVI, 1340 pages. 2005.
- Vol. 3481: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Lagana, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), *Computational Science and Its Applications – ICCSA 2005*, Part II. LXIV, 1316 pages. 2005.
- Vol. 3480: O. Gervasi, M.L. Gavrilova, V. Kumar, A. Lagana, H.P. Lee, Y. Mun, D. Taniar, C.J.K. Tan (Eds.), *Computational Science and Its Applications – ICCSA 2005*, Part I. LXV, 1234 pages. 2005.
- Vol. 3479: T. Strang, C. Linnhoff-Popien (Eds.), *Location and Context-Awareness*. XII, 378 pages. 2005.
- Vol. 3477: P. Herrmann, V. Issarny (Eds.), *Trust Management*. XII, 426 pages. 2005.
- Vol. 3475: N. Guelfi (Ed.), *Rapid Integration of Software Engineering Techniques*. X, 145 pages. 2005.
- Vol. 3468: H.W. Gellersen, R. Want, A. Schmidt (Eds.), *Pervasive Computing*. XIII, 347 pages. 2005.
- Vol. 3467: J. Giesl (Ed.), *Term Rewriting and Applications*. XIII, 517 pages. 2005.
- Vol. 3465: M. Bernardo, A. Bogliolo (Eds.), *Formal Methods for Mobile Computing*. VII, 271 pages. 2005.
- Vol. 3463: M. Dal Cin, M. Kaâniche, A. Pataricza (Eds.), *Dependable Computing – EDCC 2005*. XVI, 472 pages. 2005.
- Vol. 3462: R. Boutaba, K. Almeroth, R. Puigjaner, S. Shen, J.P. Black (Eds.), *NETWORKING 2005*. XXX, 1483 pages. 2005.
- Vol. 3461: P. Urzyczyn (Ed.), *Typed Lambda Calculi and Applications*. XI, 433 pages. 2005.
- Vol. 3460: Ö. Babaoglu, M. Jelasity, A. Montresor, C. Fetzer, S. Leonardi, A. van Moorsel, M. van Steen (Eds.), *Self-star Properties in Complex Information Systems*. IX, 447 pages. 2005.
- Vol. 3459: R. Kimmel, N.A. Sochen, J. Weickert (Eds.), *Scale Space and PDE Methods in Computer Vision*. XI, 634 pages. 2005.
- Vol. 3458: P. Herrero, M.S. Pérez, V. Robles (Eds.), *Scientific Applications of Grid Computing*. X, 208 pages. 2005.
- Vol. 3456: H. Rust, *Operational Semantics for Timed Systems*. XII, 223 pages. 2005.

- Vol. 3455: H. Treharne, S. King, M. Henson, S. Schneider (Eds.), *ZB 2005: Formal Specification and Development in Z and B*. XV, 493 pages. 2005.
- Vol. 3454: J.-M. Jacquet, G.P. Picco (Eds.), *Coordination Models and Languages*. X, 299 pages. 2005.
- Vol. 3453: L. Zhou, B.C. Ooi, X. Meng (Eds.), *Database Systems for Advanced Applications*. XXVII, 929 pages. 2005.
- Vol. 3452: F. Baader, A. Voronkov (Eds.), *Logic for Programming, Artificial Intelligence, and Reasoning*. XI, 562 pages. 2005. (Subseries LNAI).
- Vol. 3450: D. Hutter, M. Ullmann (Eds.), *Security in Pervasive Computing*. XI, 239 pages. 2005.
- Vol. 3449: F. Rothlauf, J. Branke, S. Cagnoni, D.W. Corne, R. Drechsler, Y. Jin, P. Machado, E. Marchiori, J. Romero, G.D. Smith, G. Squillero (Eds.), *Applications of Evolutionary Computing*. XX, 631 pages. 2005.
- Vol. 3448: G.R. Raidl, J. Gottlieb (Eds.), *Evolutionary Computation in Combinatorial Optimization*. XI, 271 pages. 2005.
- Vol. 3447: M. Keijzer, A. Tettamanzi, P. Collet, J.v. Hemert, M. Tomassini (Eds.), *Genetic Programming*. XIII, 382 pages. 2005.
- Vol. 3444: M. Sagiv (Ed.), *Programming Languages and Systems*. XIII, 439 pages. 2005.
- Vol. 3443: R. Bodik (Ed.), *Compiler Construction*. XI, 305 pages. 2005.
- Vol. 3442: M. Cerioli (Ed.), *Fundamental Approaches to Software Engineering*. XIII, 373 pages. 2005.
- Vol. 3441: V. Sassone (Ed.), *Foundations of Software Science and Computational Structures*. XVIII, 521 pages. 2005.
- Vol. 3440: N. Halbwachs, L.D. Zuck (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems*. XVII, 588 pages. 2005.
- Vol. 3439: R.H. Deng, F. Bao, H. Pang, J. Zhou (Eds.), *Information Security Practice and Experience*. XII, 424 pages. 2005.
- Vol. 3437: T. Gschwind, C. Mascolo (Eds.), *Software Engineering and Middleware*. X, 245 pages. 2005.
- Vol. 3436: B. Bouyssounouse, J. Sifakis (Eds.), *Embedded Systems Design*. XV, 492 pages. 2005.
- Vol. 3434: L. Brun, M. Vento (Eds.), *Graph-Based Representations in Pattern Recognition*. XII, 384 pages. 2005.
- Vol. 3433: S. Bhalla (Ed.), *Databases in Networked Information Systems*. VII, 319 pages. 2005.
- Vol. 3432: M. Beigl, P. Lukowicz (Eds.), *Systems Aspects in Organic and Pervasive Computing - ARCS 2005*. X, 265 pages. 2005.
- Vol. 3431: C. Dovrolis (Ed.), *Passive and Active Network Measurement*. XII, 374 pages. 2005.
- Vol. 3429: E. Andres, G. Damiand, P. Lienhardt (Eds.), *Discrete Geometry for Computer Imagery*. X, 428 pages. 2005.
- Vol. 3427: G. Kotsis, O. Spaniol (Eds.), *Wireless Systems and Mobility in Next Generation Internet*. VIII, 249 pages. 2005.
- Vol. 3423: J.L. Fiadeiro, P.D. Mosses, F. Orejas (Eds.), *Recent Trends in Algebraic Development Techniques*. VIII, 271 pages. 2005.
- Vol. 3422: R.T. Mittermeir (Ed.), *From Computer Literacy to Informatics Fundamentals*. X, 203 pages. 2005.
- Vol. 3421: P. Lorenz, P. Dini (Eds.), *Networking - ICN 2005, Part II*. XXXV, 1153 pages. 2005.
- Vol. 3420: P. Lorenz, P. Dini (Eds.), *Networking - ICN 2005, Part I*. XXXV, 933 pages. 2005.
- Vol. 3419: B. Faltings, A. Petcu, F. Fages, F. Rossi (Eds.), *Constraint Satisfaction and Constraint Logic Programming*. X, 217 pages. 2005. (Subseries LNAI).
- Vol. 3418: U. Brandes, T. Erlebach (Eds.), *Network Analysis*. XII, 471 pages. 2005.
- Vol. 3416: M. Böhlen, J. Gamper, W. Polasek, M.A. Wimmer (Eds.), *E-Government: Towards Electronic Democracy*. XIII, 311 pages. 2005. (Subseries LNAI).
- Vol. 3415: P. Davidsson, B. Logan, K. Takadama (Eds.), *Multi-Agent and Multi-Agent-Based Simulation*. X, 265 pages. 2005. (Subseries LNAI).
- Vol. 3414: M. Morari, L. Thiele (Eds.), *Hybrid Systems: Computation and Control*. XII, 684 pages. 2005.
- Vol. 3412: X. Franch, D. Port (Eds.), *COTS-Based Software Systems*. XVI, 312 pages. 2005.
- Vol. 3411: S.H. Myaeng, M. Zhou, K.-F. Wong, H.-J. Zhang (Eds.), *Information Retrieval Technology*. XIII, 337 pages. 2005.
- Vol. 3410: C.A. Coello Coello, A. Hernández Aguirre, E. Zitzler (Eds.), *Evolutionary Multi-Criterion Optimization*. XVI, 912 pages. 2005.
- Vol. 3409: N. Gueffi, G. Reggio, A. Romanovsky (Eds.), *Scientific Engineering of Distributed Java Applications*. X, 127 pages. 2005.
- Vol. 3408: D.E. Losada, J.M. Fernández-Luna (Eds.), *Advances in Information Retrieval*. XVII, 572 pages. 2005.
- Vol. 3407: Z. Liu, K. Araki (Eds.), *Theoretical Aspects of Computing - ICTAC 2004*. XIV, 562 pages. 2005.
- Vol. 3406: A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*. XVII, 829 pages. 2005.
- Vol. 3404: V. Diekert, B. Durand (Eds.), *STACS 2005*. XVI, 706 pages. 2005.
- Vol. 3403: B. Ganter, R. Godin (Eds.), *Formal Concept Analysis*. XI, 419 pages. 2005. (Subseries LNAI).
- Vol. 3402: M. Daydé, J.J. Dongarra, V. Hernández, J.M.L.M. Palma (Eds.), *High Performance Computing for Computational Science - VECPAR 2004*. XI, 732 pages. 2005.
- Vol. 3401: Z. Li, L.G. Vulkov, J. Waśniewski (Eds.), *Numerical Analysis and Its Applications*. XIII, 630 pages. 2005.
- Vol. 3400: J.F. Peters, A. Skowron (Eds.), *Transactions on Rough Sets III*. IX, 461 pages. 2005.
- Vol. 3399: Y. Zhang, K. Tanaka, J.X. Yu, S. Wang, M. Li (Eds.), *Web Technologies Research and Development - APWeb 2005*. XXII, 1082 pages. 2005.
- Vol. 3398: D.-K. Baik (Ed.), *Systems Modeling and Simulation: Theory and Applications*. XIV, 733 pages. 2005. (Subseries LNAI).



# Preface

We wish to extend a warm welcome to the reader of this extended postproceedings publication of SAG 2004, the 1st International Workshop on Scientific Applications on Grid Computing. This workshop was held in September 2004, in conjunction with the 2004 IEEE/WIC/ACM International Joint Conference on Web Intelligence (WI 2004) and Intelligent Agent Technology (IAT 2004).

The WI and IAT conferences have provided, for several years, a leading international forum to bring together researchers and practitioners from diverse fields, such as computer science, information technology, business, education, human factors, systems engineering, and robotics, to explore the fundamental roles as well as practical impacts of artificial intelligence (AI) (e.g., knowledge representation, planning, knowledge discovery, and data mining, intelligent agents and social network intelligence) and advanced information technology (IT) (e.g., wireless networks, ubiquitous devices, social networks, the Wisdom Web, and data/knowledge grids), and to examine the design principles and performance characteristics of various approaches in intelligent agent technology.

In the last decade, Grid computing has become one of the most important topics to appear and one of the most widely developed fields. Research into Grid computing is making rapid progress, owing to the increasing necessity of computation resources in the resolution of complex applications. The great challenge is the complete integration of heterogeneous computing systems and data resources with the aim of providing a global computing space. The achievement of this goal will involve revolutionary changes in the field of computation, enabling seamless resource and data sharing across networks. SAG 2004 aimed to provide a forum for novel topics related to Grid computing, providing an opportunity for researchers to discuss and identify key aspects of this important area.

The set of technical papers presented in this volume comprises the SAG 2004 selected papers plus a further 8 invited papers. As for the invited papers, we can say that this extended postproceedings publication gave us the opportunity and the pleasure to introduce the work being carried out by some people who are very well known in the Grid community; and for the SAG 2004 selected papers, we can say that this selection was the result of a difficult and thorough review process. The SAG 2004 workshop received 29 submissions of high quality from which the 10 papers making up the technical program were selected. The number of submissions and the quality and diversity of the resulting program are testimony to the interest in this up-and-coming area.

This publication could not have taken place without considerable enthusiasm, support and encouragement as well as sheer hard work. Many people have earned the thanks of those who attended and organized SAG 2004. In particular, we would like to thank:

- The many supporters of WI and IAT 2004 for their contributions to the conference. Many of these people have been involved with the WI and IAT 2004 conferences for several years.
- The members of the workshop Program Committee who gave their time and energy to ensure that the conference maintained its high technical quality and ran smoothly. The many individuals we owe our thanks to are listed in this volume.
- All the invited authors for their great effort, hard work and support: Konstantinos Karasavvas, Mario Antonioletti, Malcolm Atkinson, Neil Chue Hong, Tom Sugden, Alastair Hume, Mike Jackson, Amrey Krause, Charaka Palansuriya, M. Nedim Alpdemir, Arijit Mukherjee, Anastasios Gounaris, Norman W. Paton, Alvaro A.A. Fernandes, Rizos Sakellariou, Paul Watson, Peter Li, Ilkay Altintas, Adam Birnbaum, Kim Baldridge, Wibke Sudholt, Mark Miller, Celine Amoreira, Yohan Potier, Bertram Ludaescher, Georgousopoulos Cristos, Omer F. Rana, M. Cannataro, M. Comin, C. Ferrari, C. Guerra, A. Guzzo, and P. Veltri, Jose M. Perez, Felix Garcia, Jesus Carretero, Jose D. Garcia, Soledad Escolar, J. Herrera, E. Huedo, R.S. Montero, I.M. Llorente, and Jemal H. Abawajy.
- All those who submitted to the workshop. The standard set was higher than our expectations and reflected well on the research work in the community.

We would also like to acknowledge the organizers of the WI and IAT 2004 conferences, as well as Alfred Hofmann, from Springer, for the support and encouragement they extended to this publication. This volume is the result of a close cooperation and hopefully will allow us to contribute to the growth of this research community.

Pilar Herrero, María S. Pérez, Víctor Robles



# **1st International Workshop On Scientific Applications on Grid Computing (SAG 2004)**

## **Program Committee**

### **Steering Committee Co-chairs**

Pilar Herrero, Universidad Politécnica de Madrid, Spain

María S. Pérez, Universidad Politécnica de Madrid, Spain

### **General Co-chairs**

Víctor Robles, Universidad Politécnica de Madrid, Spain

Milena Radenkovic, University of Nottingham, UK

### **Program Committee**

Abawajy, Jemal, Faculty of Science and Technology, Deakin University,  
Victoria, Australia

Antic, Dragan, University of Electronic Engineering, Nis, Serbia and  
Montenegro

Baker, Mark, University of Portsmouth, UK

Benford, Steve, University of Nottingham, UK

Bosque, José Luis, URJC, Madrid, Spain

Buyya, Rajkumar, University of Melbourne, Australia

Carretero, Jesús, UC3M, Madrid, Spain

Corsaro, Angelo, Washington University in St. Louis, USA

Cortes, Toni, UPC, Barcelona, Spain

Del Peso, José, UAM, Madrid, Spain

Dongarra, Jack, University of Tennessee, Knoxville, USA

Dopico, Antonio G., UPM, Madrid, Spain

García, Félix, UC3M, Madrid, Spain

Greenhalgh, Chris, University of Nottingham, UK

Humble, Jan, University of Nottingham, UK

Lord, Phil, University of Manchester, UK

Martín, Ignacio, INTA, Madrid, Spain

Martín, Vicente, UPM, Madrid, Spain

Menasalvas, Ernestina, UPM, Madrid, Spain

Miles, Simon, University of Southampton, UK

Omicini, Andrea, Università di Bologna, Bologna, Italy

## VIII Organization

Peña, José María, UPM, Madrid, Spain

Rana, Omer, Cardiff University, UK

Rosales, Francisco, UPM, Madrid, Spain

Sánchez, Alberto, UPM, Madrid, Spain

Segovia, Javier, UPM, Madrid, Spain

Stockinger, Heinz, CERN, Geneva, Switzerland

Tari, Zahir, RMIT University, Melbourne, Australia

Zhong, Ning, Maebashi Institute of Technology, Maebashi, Japan

# Table of Contents

## Data-Based Applications

### Introduction to OGSA-DAI Services

<i>Konstantinos Karasavvas, Mario Antonioletti, Malcolm Atkinson, Neil Chue Hong, Tom Sugden, Alastair Hume, Mike Jackson, Amrey Krause, Charaka Palansuriya</i> .....	1
--	---

### Using OGSA-DQP to Support Scientific Applications for the Grid

<i>M. Nedim Alpdemir, Arijit Mukherjee, Anastasios Gounaris, Norman W. Paton, Alvaro A.A. Fernandes, Rizos Sakellariou, Paul Watson, Peter Li</i> .....	13
---	----

### Mobile Agent-Based Service Provision in Distributed Data Archives

<i>Christos Georgousopoulos, Omer F. Rana</i> .....	25
---	----

### A Proxy Service for the xrootd Data Server

<i>Andrew Hanushevsky, Heinz Stockinger</i> .....	38
---	----

### A Flexible Two-Level I/O Architecture for Grids

<i>Alberto Sánchez, María S. Pérez, Víctor Robles, José M. Peña, Pilar Herrero</i> .....	50
--	----

### Data Driven Infrastructure and Policy Selection to Enhance Scientific Applications in Grid

<i>Jose M. Perez, Felix Garcia, Jesus Carretero, Jose D. Garcia, Soledad Escolar</i> .....	59
--	----

## BioApplications

### Modelling a Protein Structure Comparison Application on the Grid Using PROTEUS

<i>Mario Cannataro, Matteo Comin, Carlo Ferrari, Concettina Guerra, Antonella Guzzo, Pierangelo Veltri</i> .....	75
--	----

### Grid Services Complemented by Domain Ontology Supporting Biomedical Community

<i>Maja Hadzic, Elizabeth Chang</i> .....	86
---	----

**Applications Architecture, Frameworks and Models**

A Generic Architecture for Sensor Data Integration with the Grid  
*Jan Humble, Chris Greenhalgh, Alastair Hamsphire,  
Henk L. Muller, Stefan Rennick Egglestone* ..... 99

Embarrassingly Distributed and Master-Worker Paradigms on the Grid  
*J. Herrera, E. Huedo, R.S. Montero, I.M. Llorente* ..... 108

A Framework for the Design and Reuse of Grid Workflows  
*Ilkay Altintas, Adam Birnbaum, Kim K. Baldridge,  
Wibke Sudholt, Mark Müller, Celine Amoreira, Yohann Potier,  
Bertram Ludaescher* ..... 120

Towards Peer-to-Peer Access Grid  
*Milena Radenkovic, Igor Miladinovic* ..... 134

A Service Oriented Architecture for Integration of Fault Diagnostics  
*Xiaoxu Ren, Max Ong, Geoffrey Allan, Visakan Kadirkamanathan,  
Haydn Thompson, Peter Fleming* ..... 146

GAM: A Grid Awareness Model for Grid Environments  
*Pilar Herrero, María S. Pérez, Víctor Robles* ..... 158

**Accounting and Market-Based Architecture**

Grid Accounting Service Infrastructure for Service-Oriented Grid  
Computing Systems  
*Jemal H. Abawajy* ..... 168

Mercatus: A Toolkit for the Simulation of Market-Based Resource  
Allocation Protocols in Grids  
*Daniel Grosu, Umesh Kant* ..... 176

**Resource and Information Management in Grid**

A Resource Monitoring and Management Middleware Infrastructure for  
Semantic Resource Grid  
*Fawad Nazir, Hafiz Farooq Ahmad, Hamid Abbas Burki,  
Tallat Hussain Tarar, Arshad Ali, Hiroki Suguri* ..... 188

A Service-Oriented Framework for Traffic Information Grid  
*Guozhen Tan, Chengxu Li, Jiankun Wu* ..... 197

**Author Index** ..... 207

# Introduction to OGSA-DAI Services

Konstantinos Karasavvas<sup>1</sup>, Mario Antonioletti<sup>2</sup>, Malcolm Atkinson<sup>1</sup>,  
Neil Chue Hong<sup>2</sup>, Tom Sugden<sup>2</sup>, Alastair Hume<sup>2</sup>, Mike Jackson<sup>2</sup>,  
Amrey Krause<sup>2</sup>, and Charaka Palansuriya<sup>2</sup>

<sup>1</sup> National e-Science Centre, University of Edinburgh, Edinburgh EH8 9AA, UK  
{kostas, mpa}@nesc.ac.uk

<sup>2</sup> EPCC, University of Edinburgh, JCMB, The King's Buildings,  
Mayfield Road, Edinburgh EH9 3JZ, UK  
{mario, neilc, tom, ally, michaelj, amrey, charaka}@epcc.ed.ac.uk

**Abstract.** In today's large collaborative environments, potentially composed of multiple distinct organisations, uniform controlled access to data has become a key requirement if these organisations are to work together as *Virtual Organisations*. We refer to such an integrated set of *data resources*<sup>1</sup> as a *virtual data warehouse*. The *Open Grid Services Architecture - Data Access and Integration* (OGSA-DAI) project was established to produce a common middleware solution, aligned with the Global Grid Forum's (GGF) OGSA vision [OGSA] to allow uniform access to data resources using a service based architecture. In this paper the service infrastructure provided by OGSA-DAI is presented providing a snapshot of its current state, in an evolutionary process, which is attempting to build infrastructure to allow easy integration and access to distributed data using grids or web services. More information about OGSA-DAI is available from the project web site: [www.ogsadai.org](http://www.ogsadai.org).

**Keywords:** Data, Databases, Grid, DAIS, OGSA-DAI, Open Grid Services Architecture, Web Services.

## 1 Introduction

Access to and the sharing of data across organisational boundaries is an important requirement for a large number of UK national and international collaborative projects. Instead of requiring each of these projects to individually solve the same data access problem, the OGSA-DAI project was established to produce a common middleware solution that allows uniform access to data resources using a service-based architecture. The initial objectives of the project have concentrated on developing the base data access platform and now, gradually, to

---

<sup>1</sup> A data resource here is taken to mean any entity that can act as a source and/or sink of data together with its associated management framework. Although the framework being developed at the moment works mainly with databases the scope is more general and could encompass file systems and streams.

focus on more sophisticated functionality that offers data integration capabilities, such as *distributed query processing* provided by the OGSA-DQP project [OGSA-DQP] using OGSA-DAI services. Up to the advent of OGSA-DAI, provision for uniform data access through service-based interfaces was absent from the then available Grid toolkits, such as the Globus Toolkit 3.0<sup>2</sup> (GT3) and Unicore.

In moving towards these ends the development of OGSA-DAI has been guided by a set of key design principles, mainly:

- *Avoid unnecessary communication between a service and its clients.* In order to minimise the number of message exchanges between a service and its clients, multiple interactions are abstracted into a set of *activities* which are then contained in a single document, referred to as a *perform document*<sup>3</sup> requiring a single message exchange.
- *Avoid unnecessary data movement.* Wherever possible move the computation to the data. Capabilities already available to a data resource, together with additional capabilities implemented at the service layer such as data transformations or third party data delivery, are exposed through the *activities* which can be linked as a series of pipelined tasks through which data flows, all within the same perform document. These activities are then executed within the scope of a single service interaction near or at the data source.
- *Provide an extensible activity framework.* It is unlikely that OGSA-DAI will provide all the base functionality, implemented as activities, that a given project might require. Thus, from the outset, perform documents and the activity *engine* have been designed to be extensible. New functionality can be implemented as activities and incorporated to work within the existing framework.
- *OGSA-DAI does not provide a complete virtualisation of the underlying data resource.* It is still necessary to know what the underlying data resource is and target suitable queries for that type of data resource. The infrastructure does not automatically do this; OGSA-DAI is not attempting to create a new universal query language suitable for all types of data resource.
- *Provide an extensible architecture.* Allow the OGSA-DAI framework to be customised or extended, e.g to add a stronger security model.
- *Build the middleware using existing standards and, where these do not exist, try to produce standards to fill in the gaps.* Up to release 5, the OGSA-DAI middleware has been based on the GGF *Open Grid Services Infrastructure* recommendation [OGSI] and its dependencies. At the time there were no

<sup>2</sup> From GT3.2 OGSA-DAI became a contributed component to the Globus Toolkit; an endorsement that such functionality was required. It is also still distributed independently of Globus through the OGSA-DAI project web site, from which there have been over 3300 downloads by approximately a 1000 registered users.

<sup>3</sup> An additional advantage of using a document based interface is that only one operation is required at the interface, *perform*. Functionality can easily be added without requiring a change to the interface, as this can be done inside the document.



standards for service-based interfaces to access data in databases so a GGF Working Group (WG) was established. The *Database Access and Integration* (DAIS) WG, since GGF 5, has been attempting to standardise this in a manner consistent with the OGSA vision. The OGSA-DAI team has been providing a strong lead in this process and aim to produce one of the two interoperable implementations required by GGF for proposed recommendations to become a full GGF recommendation. DAIS and OGSA-DAI are not currently aligned due to the initial rapid movement of the proposed DAIS specifications.

- *Establish a standard, efficient way of moving data between services.* This has become a key concern within OGSA-DAI now that the focus is moving towards data integration where more complex service-to-service interactions are required. No generic solution exists as yet, although the project has developed a *Grid Data Transport* portType<sup>4</sup> to achieve this. The scope of service-to-service data transport is much wider than just within OGSA-DAI and may require further standardisation effort within GGF.

Using the above principles, together with other guiding criteria, five major releases of OGSA-DAI have been produced at fairly regular intervals since early 2003. Each of these releases has increased the functionality, performance, and robustness of the product, as well as keeping abreast of other supporting middleware. The current release, release 5, is based on the GGF 7 version of the DAIS specification [DAIS-GGF7] and the GGF-defined OGSi recommendation. However, within this time frame OGSi has been deprecated and the DAIS specifications have radically changed. Discussion of the implications of this is postponed to section 4. The current release of OGSA-DAI uses the Globus Toolkit 3.2 OGSi implementation (GT3 Core)<sup>5</sup>. A version with a subset of the release 5 functionality that runs over WS-I+ [WS-I+], implemented as the OMII platform<sup>6</sup>, is also available. The next few sections of this paper describe the components and operation of OGSA-DAI as found in release 5.

## 2 Architecture

OGSA-DAI has thus far relied on three types of services to provide its functionality: to publish and discover information about services and data resources; to provide persistent Grid proxies for data resources and a service to access and manipulate data resources. The current base OGSA-DAI services are thus:

- **Data Access and Integration Service Group Registry (DAISGR)** – a service allowing other services to publish metadata about any data resources they represent and the capabilities they expose. A client can thus use a

<sup>4</sup> A term used within WSDL [WSDL] to collect a set of operations.

<sup>5</sup> Only a modest subset of the facilities of GT3.2 are required for OGSA-DAI and a large number of projects run successfully on this platform.

<sup>6</sup> <http://www.omii.ac.uk>

DAISGR to identify, by querying its registered metadata, a resource provider that best satisfies its needs.

- **Grid Data Service Factory (GDSF)** – acts as a persistent access point to a data resource and contains additional related metadata that may not be available at a DAISGR. A GDSF creates GDSs to access and manipulate data resources.
- **Grid Data Service (GDS)** – acts as a transient access point to a data resource. It is through a GDS that a client interacts with a data resource.

The GDS and GDSF services were specified in the GGF 7 DAIS draft recommendation. The DAISGR is not specified by DAIS but is based on interfaces specified in OGSi. A typical OGSA-DAI usage pattern is presented in figure 1 below.

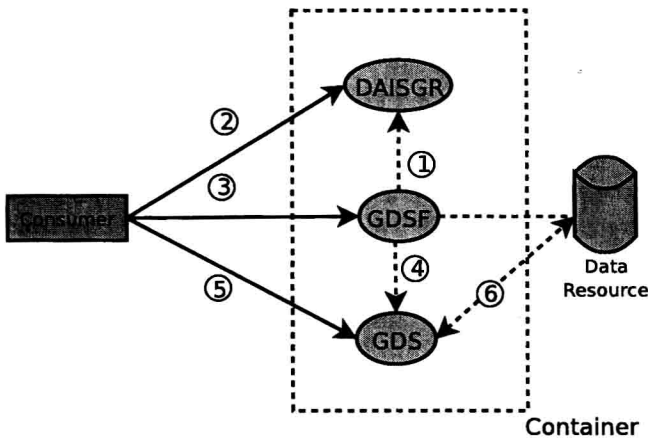


Fig. 1. Typical OGSA-DAI interaction

In figure 1 one or more persistent DAISGRs are instantiated at container start-up. Any service implementing one or more of the OGSA-DAI portTypes can register with a DAISGR. There is no requirement for the DAISGR to be co-located in the same container as the services that register with it. Services registering with a DAISGR may register their capabilities, as well as metadata about the data resource's information content.

GDSFs act as a point of presence for one particular data resource on a Grid. The current distribution of OGSA-DAI provides support for a number of different types of data resources amongst which are the relational databases: MySQL, DB2, Oracle, PostgreSQL, SQLServer, Derby (formerly Cloudscape); the XML databases: Xindice (eXist has been shown to work but is not currently supported); as well as some initial support for accessing the content of semi-structured files and file collections. Other types of data resources currently not directly supported by OGSA-DAI have also been employed, such as IBM content

manager in the e-Diamond project [eDiamond], and data streams [Data-Streams]. GDSFs are also persistent services configured at container start-up. More than one GDSF can be used to represent the same data resource if necessary but a GDSF can only expose one data resource. Currently GDSFs cannot be dynamically created or configured. On creation, a GDSF may register its service handle with a DAISGR, along with sufficient metadata and capability information to allow service/resource discovery to take place, see (1) in figure 1. Clients can obtain information about available resources (represented by GDSFs) by querying a DAISGR as in (2). They can then ask for detailed information, e.g. the schema of the resource, at a particular GDSF of interest – step (3). A GDSF, in effect, acts as a persistent Grid-enabled wrapper for a data resource but does not provide direct access to that data resource. Access to a data resource requires the creation of a GDS through the GDSF's Factory portType as specified in OGSF – step (4).

GDSs are transient services created at the request of clients who wish to access a data resource. Data resource access is done through the previously mentioned single document-based operation provided by the GDS. A client submits a perform document to the GDS – step (5) – which contains the sequence of activities to be executed on that data resource or the resulting data – step (6). The activities that can be executed by the GDS are defined when a GDSF is configured. The inner workings of a GDS are examined in more detail in the next section.

## 2.1 GDS Low-Level Interactions

The internals of a GDS are schematically shown in figure 2. The *activity engine* is the core component of a GDS. It is responsible for processing requests and generating responses. A request takes the form of a perform document – see section 2.3 for an example, which is an XML document containing one or more activities. An activity dictates the action that the activity engine must

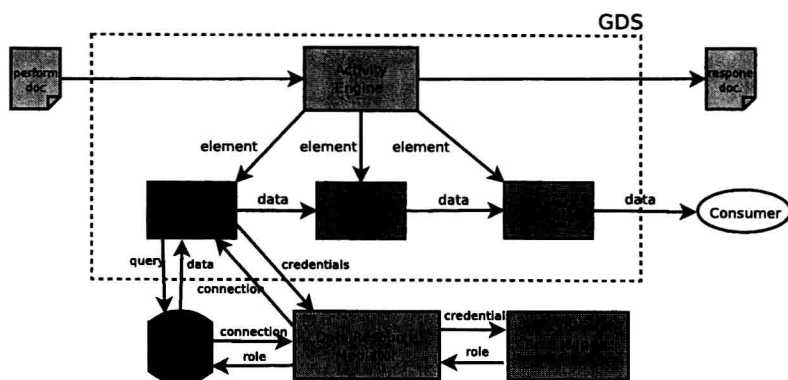


Fig. 2. GDS perform document processing and interactions