

2009 Fifth IEEE International Conference on e-Science

(e-Science 2009)

**Oxford, United Kingdom
9 – 11 December 2009**



IEEE Catalog Number: CFP0906A-PRT
ISBN: 978-1-4244-5340-5

2009 Fifth IEEE International Conference on e-Science

e-Science 2009

Table of Contents

Preface

Organizing Committee

Reviewers

Arts, Humanities and Social Science 1

Alfalab: Construction and Deconstruction of a Digital Humanities Experiment	1
<i>Joris van Zundert, Douwe Zeldenrust, and Anne Beaulieu</i>	
Topic Maps in the eHumanities	6
<i>Lutz Maicher, Khalil Ahmed, Alida Isolani, Aki Kivelä, Sam Oh, Antony Pitts, and Salvatore Vassallo</i>	
An Image Processing Portal and Web-Service for the Study of Ancient Documents	14
<i>Ségolene M. Tarte, David C. H. Wallom, Pin Hu, Kang Tang, and Tiejun Ma</i>	

Arts, Humanities and Social Science 2

CLAROS - Bringing Classical Art to a Global Public	20
<i>Donna Kurtz, Greg Parker, David Shotton, Graham Klyne, Florian Schroff, Andrew Zisserman, and Yorick Wilks</i>	
Integrating Full-Text Search and Linguistic Analyses on Disperse Data for Humanities and Social Sciences Research Projects	28
<i>Marta Villegas and Carla Parra</i>	
Building Bridges between Islands of Data - An Investigation into Distributed Data Management in the Humanities	33
<i>Mike Jackson, Mario Antonioletti, Alastair Hume, Tobias Blanke, Gabriel Bodard, Mark Hedges, and Shrija Rajbhandari</i>	

Arts, Humanities and Social Science 3

An Ontology Based Framework for the Preservation of Interactive Multimedia Performances	40
<i>Kia Ng, Eleni Mikroyannidi, Bee Ong, and David Giaretta</i>	
The SweDat Project and Swedia Database for Phonetic and Acoustic Research	45
<i>Jonas Lindh and Anders Eriksson</i>	
Supporting the Running and Analysis of Trials of Web-Based Behavioural Interventions: The LifeGuide	50
<i>Yang Yang, Adrian Osmond, Xiaoyu Chen, Mark Weal, Gary Wills, David De Roure, Judith Joseph, and Lucy Yardley</i>	

Bioinformatics & Health 1

Phylogenetic Predictions on Grids	58
<i>Priyank Raj Katariya and Sathish S. Vadhiyar</i>	
Increasing the Efficiency of Data Storage and Analysis Using Indexed Compression	66
<i>Nathaniel Beagley, Chad Scherrer, Yan Shi, Brian H. Clowers, William F. Danielson, and Anuj R. Shah</i>	
User-Level Virtual Network Support for Sky Computing	72
<i>Maurício Tsugawa, Andréa Matsunaga, and José Fortes</i>	

Bioinformatics & Health 2

Strategies for Network Motifs Discovery	80
<i>Pedro Ribeiro, Fernando Silva, and Marcus Kaiser</i>	
An Architecture for Real Time Data Acquisition and Online Signal Processing for High Throughput Tandem Mass Spectrometry	88
<i>Anuj R. Shah, Navdeep Jaitly, Nino Zuljevic, Matthew E. Monroe, Andrei Liyu, Ashoka D. Polpitiya, Joshua N. Adkins, Mikhail E. Belov, Gordon A. Anderson, Richard D. Smith, and Ian Gorton</i>	
Sharing and Reusing Cancer Image Segmentation Algorithms Using Scientific Workflows: Pros and Cons	94
<i>M. S. Avila-Garcia, A. E. Trefethen, N. Joshi, F. Gleeson, and W. Ba-alawi</i>	

Climate and Earth Sciences 1

A Cloud-Based Interactive Application Service	102
<i>Nayden Markatchev, Roger Curry, Cameron Kiddle, Andrey Mirtchovski, Rob Simmonds, and Tingxi Tan</i>	
eScience for Sea Science: A Semantic Scientific Knowledge Infrastructure for Marine Scientists	110
<i>Kristin Stock, Anne Robertson, Femke Reitsma, Tim Stojanovic, Mohamed Bishr, David Medyckyj-Scott, and Jens Ortmann</i>	

Fluxdata.org: Publication and Curation of Shared Scientific Climate and Earth Sciences Data	118
<i>Marty Humphrey, Deb Agarwal, and Catharine van Ingen</i>	

Digital Repositories and Data Management 1

Towards a Mobile Workbench for Researchers	126
<i>Andreas Hense, Florian Quadt, and Matthias Römer</i>	
The Role of OAI Representation Information in the Digital Curation of Crystallography Data	132
<i>Manjula Patel, Simon Coles, David Giaretta, Stephen Rankin, and Brian McIlwrath</i>	
Towards a Universal, Quantifiable, and Scalable File Format Converter	140
<i>Kenton McHenry, Rob Kooper, and Peter Bajcsy</i>	

Digital Repositories and Data Management 2

Comparing METS and OAI-ORE for Encapsulating Scientific Data Products: A Protein Crystallography Case Study	148
<i>Charles Brooking, Stephen R. Shouldice, Gautier Robin, Bostjan Kobe, Jennifer L. Martin, and Jane Hunter</i>	
Stream Monitoring in Large-Scale Distributed Concealed Environments	156
<i>Mario Lassnig, Thomas Fahringer, Vincent Garonne, Angelos Molfetas, and Miguel Branco</i>	
Publish My Data: A Composition of Services from ANDS and ARCS	164
<i>Adrian Burton and Andrew Treloar</i>	

Digital Repositories and Data Management 3

A Protocol for Exchanging Scientific Citations	171
<i>Brian Matthews, Alastair Duncan, Catherine Jones, Cameron Neylon, Mark Borkum, Simon Coles, and Philip Hunter</i>	
Beyond the Document Library: Portal-Based Browsing and Exploration of Community Data Clouds	178
<i>Yong Liu, Kailash Kotwani, Alejandro Rodriguez, Joe Futrelle, Robert McGrath, and James Myers</i>	
Extracting and Ingesting DDI Metadata and Digital Objects from a Data Archive into the iRODS Extension of the NARA TPAP Using the OAI-PMH	185
<i>Jewel H. Ward, Antoine de Torcy, Mason Chua, and Jonathan Crabtree</i>	

Digital Repositories and Data Management 4

A Methodology for File Relationship Discovery	193
<i>Michal Ondrejcek, Jason Kastner, Rob Kooper, and Peter Bajcsy</i>	
ICAT: Integrating Data Infrastructure for Facilities Based Science	201
<i>Damian Flannery, Brian Matthews, Tom Griffin, Juan Bicarregui, Michael Gleaves, Laurent Lerusse, Roger Downing, Alun Ashton, Shoab Sufi, Glen Drinkwater, and Kerstin Kleese</i>	

Supporting Cloud Computing with the Virtual Block Store System	208
<i>Xiaoming Gao, Mike Lowe, Yu Ma, and Marlon Pierce</i>	

eScience Practice and Education 1

Some Challenges Facing Scientific Software Developers: The Case of Molecular Biology	216
<i>Chris Morris and Judith Segal</i>	

Problems and Approaches of Workflow Scheduling in MediGRID	223
<i>Dietmar Sommerfeld and Harald Richter</i>	

A Fresh Perspective on Developing and Executing DAG-Based Distributed Applications: A Case-Study of SAGA-Based Montage	231
<i>Andre Merzky, Katerina Stamou, Shantenu Jha, and Daniel S. Katz</i>	

Physical Sciences and Engineering 1

Virtual Microscopy and Analysis Using Scientific Workflows	239
<i>David Abramson, Blair Bethwaite, Minh Ngoc Dinh, Colin Enticott, Stephen Firth, Slavisa Garic, Ian Harper, Martin Lackmann, Hoang Nguyen, Tirath Ramdas, A. B. M. Russel, Stefan Schek, and Mary Vail</i>	

A Topological Framework for the Interactive Exploration of Large Scale Turbulent Combustion	247
<i>Peer-Timo Bremer, Gunther H. Weber, Julien Tierny, Valerio Pascucci, Marcus S. Day, and John B. Bell</i>	

Enabling Computational Steering with an Asynchronous-Iterative Computation Framework	255
<i>Alexandre di Costanzo, Chao Jin, Carlos A. Varela, and Rajkumar Buyya</i>	

Physical Sciences and Engineering 2

Robust Asynchronous Optimization for Volunteer Computing Grids	263
<i>Travis Desell, Malik Magdon-Ismail, Boleslaw Szymanski, Carlos Varela, Heidi Newberg, and Nathan Cole</i>	

The MOSFET Virtual Organisation: Grid Computing for Simulation in Nanoelectronics	271
<i>R. Valin, N. Seoane, M. Aldegunde, and A. Garcia-Loureiro</i>	

A High-Performance Hybrid Computing Approach to Massive Contingency Analysis in the Power Grid	277
<i>Ian Gorton, Zhenyu Huang, Yousu Chen, Benson Kalahar, Shuangshuang Jin, Daniel Chavarría-Miranda, Doug Baxter, and John Feo</i>	

Research Tools, Workflow and Systems 1

Rapid Chemistry Portals through Engaging Researchers	284
<i>Jos Koetsier, Andrew Turner, Patricia Richardson, and Jano van Hemert</i>	

Running Parallel Applications with Topology-Aware Grid Middleware	292
<i>Pavel Bar, Camille Coti, Derek Groen, Thomas Herault, Valentin Kravtsov, Assaf Schuster, and Martin Swain</i>	

Scheduling Multiple Parameter Sweep Workflow Instances on the Grid	300
<i>Sucha Smanchat, Maria Indrawan, Sea Ling, Colin Enticott, and David Abramson</i>	

Research Tools, Workflow and Systems 2

A Virtual Connectivity Layer for Grids	307
<i>Jefferson Tan, David Abramson, and Colin Enticott</i>	

A New Fault Tolerance Heuristic for Scientific Workflows in Highly Distributed Environments Based on Resubmission Impact	313
<i>Kassian Plankensteiner, Radu Prodan, and Thomas Fahringer</i>	

Building Reliable Data Pipelines for Managing Community Data Using Scientific Workflows	321
<i>Yogesh Simmhan, Catharine van Ingen, Alex Szalay, Roger Barga, and Jim Heasley</i>	

Research Tools, Workflow and Systems 3

DryadLINQ for Scientific Analyses	329
<i>Jaliya Ekanayake, Thilina Gunarathne, Geoffrey Fox, Atilla Soner Balkir, Christophe Poulain, Nelson Araujo, and Roger Barga</i>	

CHIC - Converting Hamburgers into Cows	337
<i>Joseph A. Townsend, Jim Downing, and Peter Murray-Rust</i>	

Expressive Reusable Workflow Templates	344
<i>Yolanda Gil, Paul Groth, Varun Ratnakar, and Christian Fritz</i>	

Research Tools, Workflow and Systems 4

Towards Lensfield - Data Management, Processing and Semantic Publication for Vernacular e-Science	352
<i>Nick Day, Jim Downing, Lezan Hawizy, Nico Adams, and Peter Murray-Rust</i>	

Enabling Advanced Visualization Tools in a Web-Based Simulation Monitoring System	358
<i>Emanuele Santos, Julien Tierny, Ayla Khan, Brad Grimm, Lauro Lins, Juliana Freire, Valerio Pascucci, Cláudio T. Silva, Scott Klasky, Roselyne Barreto, and Norbert Podhorszki</i>	

An Autonomic Approach to Integrated HPC Grid and Cloud Usage	366
<i>Hyunjoo Kim, Yaakoub el-Khamra, Shantenu Jha, and Manish Parashar</i>	

Author Index