# 2009 Fifth IEEE International Conference on e-Science

## e-Science 2009

### Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>x</td>
</tr>
<tr>
<td>Organizing Committee</td>
<td>xii</td>
</tr>
<tr>
<td>Reviewers</td>
<td>xiii</td>
</tr>
</tbody>
</table>

### Arts, Humanities and Social Science 1

**Alfalab: Construction and Deconstruction of a Digital Humanities Experiment** ............................................1

*Joris van Zundert, Douwe Zeldenrust, and Anne Beaulieu*

**Topic Maps in the eHumanities** .................................................................6

*Lutz Maicher, Khalil Ahmed, Alida Isolani, Aki Kivelä, Sam Oh, Antony Pitts, and Salvatore Vassallo*

**An Image Processing Portal and Web-Service for the Study of Ancient Documents** ...........................................14

*Ségonèle M. Tarte, David C. H. Wallom, Pin Hu, Kang Tang, and Tiejun Ma*

### Arts, Humanities and Social Science 2

**CLAROS - Bringing Classical Art to a Global Public** .................................................................20

*Donna Kurtz, Greg Parker, David Shotton, Graham Klyne, Florian Schroff, Andrew Zisserman, and Yorick Wilks*

**Integrating Full-Text Search and Linguistic Analyses on Disperse Data for Humanities and Social Sciences Research Projects** .................................................................28

*Marta Villegas and Carla Parra*

**Building Bridges between Islands of Data - An Investigation into Distributed Data Management in the Humanities** .................................................................33

*Mike Jackson, Mario Antonioletti, Alastair Hume, Tobias Blanke, Gabriel Bodard, Mark Hedges, and Shrija Rajbhandari*
Arts, Humanities and Social Science 3

An Ontology Based Framework for the Preservation of Interactive Multimedia Performances

Kia Ng, Eleftheria Mikroyannidi, Bee Ong, and David Giaretta

The SweDat Project and Swedia Database for Phonetic and Acoustic Research

Jonas Lindh and Anders Eriksson

Supporting the Running and Analysis of Trials of Web-Based Behavioural Interventions: The LifeGuide

Yang Yang, Adrian Osmond, Xiaoyu Chen, Mark Weal, Gary Wills, David De Roure, Judith Joseph, and Lucy Yardley

Bioinformatics & Health 1

Phylogenetic Predictions on Grids

Priyank Raj Katariya and Sathish S. Vadhiyar

Increasing the Efficiency of Data Storage and Analysis Using Indexed Compression

Nathaniel Beagley, Chad Scherrer, Yan Shi, Brian H. Clowers, William F. Danielson, and Anuj R. Shah

User-Level Virtual Network Support for Sky Computing

Mauricio Tsugawa, Andre Matsumata, and Jose Fortes

Bioinformatics & Health 2

Strategies for Network Motifs Discovery

Pedro Ribeiro, Fernando Silva, and Marcus Kaiser

An Architecture for Real Time Data Acquisition and Online Signal Processing for High Throughput Tandem Mass Spectrometry


Sharing and Reusing Cancer Image Segmentation Algorithms Using Scientific Workflows: Pros and Cons

M. S. Avila-Garcia, A. E. Trefethen, N. Joshi, F. Gleeson, and W. Ba-alawi

Climate and Earth Sciences 1

A Cloud-Based Interactive Application Service

Nayden Markatchev, Roger Curry, Cameron Kiddle, Andrey Mirtchovski, Rob Simmonds, and Tingxi Tan

eScience for Sea Science: A Semantic Scientific Knowledge Infrastructure for Marine Scientists

Kristin Stock, Anne Robertson, Femke Reitsma, Tim Stojanovic, Mohamed Bishr, David Medyckyj-Scott, and Jens Ortmann
Fluxdata.org: Publication and Curation of Shared Scientific Climate and Earth Sciences Data ................................................................. 118
   Marty Humphrey, Deb Agarwal, and Catharine van Ingen

Digital Repositories and Data Management 1
Towards a Mobile Workbench for Researchers ................................................................. 126
   Andreas Hense, Florian Quadt, and Matthias Römer
The Role of OAIS Representation Information in the Digital Curation of Crystallography Data .......................................................................................... 132
   Manjula Patel, Simon Coles, David Giaretta, Stephen Rankin, and Brian McIlwrath
Towards a Universal, Quantifiable, and Scalable File Format Converter ................................................................. 140
   Kenton McHenry, Rob Kooper, and Peter Bajcsy

Digital Repositories and Data Management 2
Comparing METS and OAI-ORE for Encapsulating Scientific Data Products: A Protein Crystallography Case Study .......................................................................................... 148
   Charles Brooking, Stephen R. Shouldice, Gautier Robin, Bostjan Kobe, Jennifer L. Martin, and Jane Hunter
Stream Monitoring in Large-Scale Distributed Concealed Environments .......................................................................................... 156
   Mario Lassnig, Thomas Fahninger, Vincent Garonne, Angelos Molfetas, and Miguel Branco
Publish My Data: A Composition of Services from ANDS and ARCS .......................................................................................... 164
   Adrian Burton and Andrew Treloar

Digital Repositories and Data Management 3
A Protocol for Exchanging Scientific Citations ............................................................................. 171
   Brian Matthews, Alastair Duncan, Catherine Jones, Cameron Neylon, Mark Borkum, Simon Coles, and Philip Hunter
Beyond the Document Library: Portal-Based Browsing and Exploration of Community Data Clouds .......................................................................................... 178
   Yong Liu, Kailash Kotwani, Alejandro Rodriguez, Joe Futrelle, Robert McGrath, and James Myers
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
   Jewel H. Ward, Antoine de Torcy, Mason Chua, and Jonathan Crabtree

Digital Repositories and Data Management 4
A Methodology for File Relationship Discovery ............................................................................. 193
   Michal Ondrejcek, Jason Kastner, Rob Kooper, and Peter Bajcsy
ICAT: Integrating Data Infrastructure for Facilities Based Science .......................................................................................... 201
   Damian Flannery, Brian Matthews, Tom Griffin, Juan Bicarregui, Michael Gleaves, Laurent Lerousse, Roger Downing, Alun Ashton, Shoaib Sufi, Glen Drinkwater, and Kerstin Kleese
Running Parallel Applications with Topology-Aware Grid Middleware ...........................................292
   Pavel Bar, Camille Coti, Derek Groen, Thomas Herault, Valentin Kravtsov,
   Assaf Schuster, and Martin Swain
Scheduling Multiple Parameter Sweep Workflow Instances on the Grid ...........................................300
   Sucha Smanchat, Maria Indrawan, Sea Ling, Colin Enticott, and David Abramson

Research Tools, Workflow and Systems 2
A Virtual Connectivity Layer for Grids ..................................................................................................307
   Jefferson Tan, David Abramson, and Colin Enticott
A New Fault Tolerance Heuristic for Scientific Workflows in Highly Distributed Environments Based on Resubmission Impact ..........................................................313
   Kassian Plankensteiner, Radu Prodan, and Thomas Fahringer
Building Reliable Data Pipelines for Managing Community Data Using Scientific Workflows ........................................................................................................321
   Yogesh Simmhan, Catharine van Ingen, Alex Szalay, Roger Barga,
   and Jim Heasley

Research Tools, Workflow and Systems 3
DryadLINQ for Scientific Analyses .......................................................................................................329
   Jaliya Ekanayake, Thilina Gunarathne, Geoffrey Fox, Atilla Soner Balkir,
   Christophe Poulain, Nelson Araujo, and Roger Barga
CHIC - Converting Hamburgers into Cows .........................................................................................337
   Joseph A. Townsend, Jim Downing, and Peter Murray-Rust
Expressive Reusable Workflow Templates .......................................................................................344
   Yolanda Gil, Paul Groth, Varun Ratnakar, and Christian Fritz

Research Tools, Workflow and Systems 4
Towards Lensfield - Data Management, Processing and Semantic Publication for Vernacular e-Science .................................................................................................352
   Nick Day, Jim Downing, Lezan Hawizy, Nico Adams, and Peter Murray-Rust
Enabling Advanced Visualization Tools in a Web-Based Simulation Monitoring System ........................................................................................................358
   Emanuele Santos, Julien Tierny, Ayla Khan, Brad Grimm, Lauro Lins,
   Juliana Freire, Valerio Pascucci, Cláudio T. Silva, Scott Klasky,
   Roselyne Barreto, and Norbert Podhorszki
An Autonomic Approach to Integrated HPC Grid and Cloud Usage ...........................................366
   Hyunjoo Kim, Yaakoub el-Khamra, Shantenu Jha, and Manish Parashar

Author Index ..................................................................................................................................375