Table of Contents

Conference

Workflow 1

Common Motifs in Scientific Workflows: An Empirical Analysis  1
Daniel Garijo, Pinar Alper, Khalid Belhajjame, Oscar Corcho, Yolanda Gil, Carole Goble

Why workflows break - Understanding and combating decay in Taverna workflows  9
Jun Zhao, Jose Manuel Gomez-Perez, Khalid Belhajjame, Graham Klyne, Esteban Garcia-Cuesta, Aleix Garrido, Kristina Hettne, Marco Roos, David De Roure, Carole Goble

WorkflowSim: A Toolkit for Simulating Scientific Workflows in Distributed Environments  18
Weiwei Chen, Ewa Deelman

Ecology

Towards Semantically-enabled Exploration and Analysis of Environmental Ecosystems  26
Ping Wang, Linyun Fu, Evan Patton, Deborah McGuinness, F. Joshua Dein, R. Sky Bristol

Generalized Representation and Mapping for Social-Ecological Data: Freeing Data from the Database  34
Scott Jensen, Beth Plale, Xiaozhong Liu, Miao Chen, David Leake, Julie England

Automated Data Verification in a Large-scale Citizen Science Project: a Case Study  42
Jun Yu, Steve Kelling, Jeff Gerbracht, Weng-Keen Wong

Process of eScience

Enabling Scientific Data Sharing and Re-use  50
Robert Darby, Simon Lambert, Brian Matthews, Michael Wilson, Kathrin Gitsmans, Sunje Dallmeier-Tiessen, Salvatore Mele, Jari Suhonen

Collaborative Information Management In Scientific Research Process  58
Shirley Crompton, Brian Matthews, Erica Yang, Cameron Neylon, Simon Coles

Towards next generations of software for distributed infrastructures: the European Middleware Initiative  65
Cristina Aiftimiel, Alberto Aimar, Andrea Ceccanti, Marco Cecchi, Alberto Di Meglio, Florida Estrella, Patrick Fuhrmann, Emidio Giorgio, Balázs Kónya, Jon Kerr Nilsen, Morris Riedel, John White

Workflow and the User

Service-based Integration of Human Users in Workflow-driven Scientific Experiments  75
Dimka Karastoyanova, Dimitrios Dentsas, David Schumm, Mirko Sonntag, Lina Sun, Karolina Vukojevic
WorkWays: Interactive Workflow-based Science Gateways  83
Hoang Nguyen, David Abramson

Integration of modern data management practice with scientific workflows  91
Neil Killeen, Jason Lohrey, Michael Farrell, Wilson Liu, Slavisa Garic, David Abramson, Gary Egan

Applications & Data

Data-Oriented Research for Bioresource Utilization: a Case Study to Investigate Water Uptake in Cellulose Using Principal Components  99
Yi Liu, Carlos Driemeier, Roberto Cesar-Jr

Adaptive Sampling of Streaming Signals  106
Tristan Wietsma, Barbara Minsker

Calibration of Watershed Models using Cloud Computing  113
Marty Humphrey

Technologies

Folding Proteins at 500 ns/hour using Work Queue  121
Badi' Abdul-Wahid, Li Yu, Dinesh Rajan, Haoyun Feng, Eric Darve, Douglas Thain, Jesus A. Izaguirre

Fast Confidential Search for Bio-Medical Data Using Bloom Filters and Homomorphic Cryptography  129
Henning Perl, Yassene Mohammed, Michael Brenner, Matthew Smith

Virtual Simulation Objects Concept as a Framework for System-Level Simulation  137
Sergey Kovalchuk, Pavel Smirnov, Sergey Kosukhin, Alexander Boukhanovsky

Data Provenance

Temporal Representation for Scientific Data Provenance  145
Peng Chen, Beth Plale, Mehmet Aktas

Provenance Analysis: Towards Quality Provenance  153
You-Wei Cheah, Beth Plale

Scientific Workflow Rewriting while Preserving Provenance  161
Sarah Cohen-Boulakia, Christine Froidevaux, Jiuqiang Chen

Biology

An integrated science portal for collaborative compute and data intensive protein structure studies  170
Ian Stokes-Rees, Piotr Sliz

Prediction of Protein Solubility in E. coli  178
Taghrid Samak, Dan Gunter, Zhong Wang

Data-intensive analysis of HIV mutations  186
## Semantics

**RightField: Semantic Enrichment of Systems Biology Data using Spreadsheets**  193  
Katherine Wolstencroft, Stuart Owen, Carole Goble, Quyen Nguyen, Olga Krebs, Wolfgang Müller

**MIM: A Minimum Information Model Vocabulary and Framework for Scientific Linked Data**  201  
Matthew Gamble, Jun Zhao, Graham Klyne, Carole Goble

**From Scripts Towards Provenance Inference**  209  
Mohammad Rezwanul Huq, Peter M.G. Apers, Yoshihide Wada, Ludovicus P.H. van Beek, Andreas Wombacher

## Data Mining & Machine Learning

**Mining Hidden Mixture Context With ADIOS-P To Improve Predictive Pre-fetcher Accuracy**  217  
Jong Youl Choi, Mohammad H. Abbasi, David Pugmire, Scott Klasky, Judy Qiu, Geoffrey Fox

**Remote Phenology: Applying Machine Learning to Detect Phenological Patterns in a Cerrado Savanna**  225  
Jurandy Almeida, Jefersson Alex Dos Santos, Bruna Alberton, Ricardo Torres, Leonor Morellato

**Flashes in a Star Stream: Automated Classification of Astronomical Transient Events**  233  
S. George Djorgovski, Ashish Mahabal, Ciro Donalek, Matthew Graham, Andrew Drake, Baback Moghaddam, Michael Turmon

**Statistical Analysis and Visualization Services for Spatially Integrated Social Science Datasets**  241  
Irfan Azeezullah, Friska Pambudi, Tung-Kai Shyy, Nigel Ward, Jane Hunter, Robert Stimson

## Runtime & Middleware

**P*: A Model of Pilot-Abstractions**  249  
Andre Luckow, Mark Santcroos, Ole Weidner, Andre Merzky, Pradeep Kumar Mantha, Shantenu Jha

**MARISSA: MapReduce Implementation for Streaming Science Applications**  259  
Elif Dede, Zacharia Fadika, Jessica Hartog, Madhusudhan Govindaraju, Lavanya Ramakrishnan, Dan Gunter, Richard Shane Canon

**Cooperative VM Migration for a Virtualized HPC Cluster with VMM-Bypass I/O devices**  267  
Ryousei Takano, Hidemoto Nakada, Takahiro Hirofuchi, Yoshio Tanaka, Tomohiro Kudoh

**Using Promethee Methods for Multi-Criteria Pull-based Scheduler on DCIs**  275  
Mircea Moca, Gilles Fedak

## Data Infrastructures

**A Data-driven Urban Research Environment for Australia**  283  
Richard O. Sinnott, Christopher Bayliss, Gerson Galang, Phillip Greenwood, George Koetsier, Damien Mannix, Luca Morandini, Marcos Nino-Ruiz, Chris Pettit, Martin Tomko, Muhammed Sarwar, Robert
Stimson, William Voorsluys, Ivo Widjaja

The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geo-Spatial Data  291
Luca Cinquini, Dan Crichton, Chris Mattmann, Gavin Bell, Bob Drach, Dean Williams, John Harney, Galen Shipman, Feiyi Wang, Philip Kershaw, Stephen Pascoe, Rachana Ananthakrishnan, Neill Miller, Estanislaw Gonzalez, Sebastian Denvil, Mark Morgan, Sandro Fiore, Zed Pobre, Roland Schweitzer

CINET: A CyberInfrastructure for Network Science  301
Sherif Abdelhamid, Richard Alo, S M Arifuzzaman, Pete Beckman, Md Hasanuzzaman Bhuiyan, Keith Bisset, Edward Fox, Geoffrey Fox, Kevin Hall, S.M.Shamimul Hasan, Anurodh Joshi, Maleq Khan, Chris Kuhlman, Spencer Lee, Jonathan Leidig, Hemanth Makkapati, Madhav Marathe, Henning Mortveit, Judy Qiu, Ravi S.S., Zalia Shams, Ongard Sirisaengtaksin, Rajesh Subbiah, Samarth Swarup, Nick Trebon, Anil Vullikanti, Zhao Zhao

A Framework to Access Handwritten Information within Large Digitized Paper Collections  309
Liana Diesendruck, Luigi Marini, Rob Kooper, Mayank Kejriwal, Kenton Mchenry

Data Access & Management

Open Social based Group Access Control Framework for e-Science Data Infrastructure  319
Hui Zhang, Wenjun Wu, Zhenan Li

Partial Replica Selection for Spatial Datasets  327
Yun Tian, Philip Rhodes

Image Retrieval in the Unstructured Data Management System AUDR  337
Junwu Luo, Bo Lang, Danchen Zhang, Chao Tian

Application Systems & Frameworks

A System for Management of Computational Fluid Dynamics Simulations for Civil Engineering  344
Peter Sempolinski, Douglas Thain, Daniel Wei, Ahsan Kareem

BIGS: A Framework for Large-Scale Image Processing and Analysis Over Distributed and Heterogeneous Computing Resources  352
Raúl Ramos-Pollán, Fabio Augusto González Osorio, Juan Carlos Caicedo Rueda, Ángel Cruz-Roa, Jorge Eliécer Camargo Mendoza, Jorge Andrés Vanegas, John Edilson Arévalo Ovalle, Paola Katherine Rozo Bernal, Santiago Alonso Pérez Rubiano, Jose David Bermeo López, Juan Sebastián Otálora Montenegro

A Simulator for Social Exchanges and Collaborations - Architecture and Case Study  360
Christian Haas, Simon Caton, Daniel Trumpp, Christof Weinhardt

Workflow 2

Quality of Data Driven Simulation Workflows  368
Michael Reiter, Uwe BreitenbÜcher, Dimka Karastoyanova, Oliver Kopp

A new Optimization Phase for Scientific Workflow Management Systems  376
Discovering Drug Targets for Neglected Diseases Using a Pharmacophylogenomic Cloud Workflow 384
Kary Ocaña, Daniel Oliveira, Jonas Dias, Eduardo Ogasawara, Marta Mattoso

Workshops

XHPC

Towards HPC For the Digital Humanities, Arts, and Social Sciences: Needs and Challenges of Adapting Academic HPC for Big Data 392
Kalev Leetaru

Happy or Not : Generating Topic-based Geospatial Emotional Heatmaps for Culturomics using CyberGIS 398
Eric Shook, Kalev Leetaru, Guofeng Cao, Anand Padmanabhan, Shaowen Wang

High-Performance Computing without commitment 404
Kevin Jorissen, John J. Rehr, Fernando D. Vila, William Johnson

Lessons Learned from Galaxy, a Web-based Platform for High-throughput Genomic Analyses 410
Jeremy Goecks, Galaxy Team, Anton Nekrutenko, James Taylor

High-Performance Data Management For Genome Sequencing Centers Using Globus Online: A Case Study 416
Dinanath Sulakhe, Rajkumar Kettimuthu, Utpal Dave

FRED Navigator: An Interactive System for Visualizing Results from Large-scale Epidemic Simulations 422
Jack Paparian, Shawn Brown, Donald Burke, John Grefenstette

Digitization and Search: A Non-Traditional Use of HPC 427
Liana Diesendruck, Luigi Marini, Rob Kooper, Mayank Kejriwal, Kenton Mchenry

Large Scale Video Analytics: On-demand, iterative inquiry for moving image research 433
Virginia Kuhn, Ritu Arora, Alan Craig, Kevin Franklin, Dave Bock, Michael Simeone, Luigi Marini

Incorporating Circulation Data in Relevancy Rankings for Search Algorithms in Library Collections 438
Harriett Green, Kirk Hess, Richard Hislop

SoftwarePractice 2012

Overview of the TriBITS Lifecycle Model: A Lean/Agile Software Lifecycle Model for Research-based Computational Science and Engineering and Applied Mathematical Software 444
Roscoe Bartlett, Micheal Heroux, James Willenbring

IPOL: Reviewed Publication and Public Testing of Research Software 452
Nicolas Limare, Laurent Oudre
On Realizing the Concept Study ScienceSoft of the European Middleware Initiative - Open Software for Open Science  460
Alberto Di Meglio, Morris Riedel

eSoN 12

ExSciTecH: Expanding Volunteer Computing to Explore Science, Technology, and Health  468
Michael Matheny, Samuel Schlachter, Lindsey Crouse, Edward Kimmel, Trilce Estrada, Marcel Schumann, Roger Armen, Gary Zoppetti, Michela Taufer

Experiences in the Design and Implementation of a Social Cloud for Volunteer Computing  476
Ryan Chard, Kris Bubendorfer, Kyle Chard

A Satellite Data Portal Developed for Crowdsourcing Data Analysis and Interpretation  484
Zhenghui Hu, Wenjun Wu

g-Social: Enhancing Integrated eScience Tools with Social Networking Functionality  492
Stylianou Andriani, Loulloudes Nicholas, Dikaiakos Marios D.

RunMyCode.org: a novel dissemination and collaboration platform for executing published computational results  500
Victoria Stodden, Christophe Hurlin, Christophe Pérignon

Towards a quantitative academic internationalization assessment of Brazilian research groups  508
Evelyn Perez Cervantes, Jesús Mena-Chalco, Roberto Cesar

Towards the Gamification of Well-being Measures  516
Margeret Hall, Steven O. Kimbrough, Christian Haas, Christof Weinhardt, Simon Caton

Collaborative research using eScience infrastructure and high speed networks (NECS 2012)

Enabling large genomic data transfers using nationwide and international dynamic lightpaths  524
Jan Bot, Migiel de Vos, Sander Boele, Marcel Reinders, Joost Kok

Pilot Abstractions for Compute, Data, and Network  526
Mark Santcroo, Delgado Olabarriaga, Daniel S. Katz, Shantenu Jha

Verification and User Experience of High Data Rate Bandwidth-on-Demand networks  528
Jimmy Cullen, Richard Hughes-Jones, Ralph Spencer

Scientific Workflows using Dynamically Provisioned High-Speed Circuits  530
Paul Ruth, Anirban Mandal, Yufeng Xin, Ilia Badine, Chris Heerman, Jeff Chase

Evaluation of RoCE over a 40 Gbps Transcontinental Layer-2 Circuit  532
Brian Tierney, Ezra Kissel, Martin Swany, Eric Pouyoul

eScience Meets the Instrument
Velo and REXAN - Integrated Data Management and High Speed Analysis for Experimental Facilities 541
Kerstin Kleese van Dam, James Carson, Abigail Corrigan, Daniel Einstein, Zoe Guillen, Brandi Heath, Andrew Kuprat, Ingela Lanekoff, Carina Lansing, Julia Laskin, Dongheng Li, Yan Liu, Matt Marshall, Erin Miller, Galya Orr, Paulo Pinheiro da Silva, Seun Ryu, Craig Szymanski, Mathew Thomas

X-ray imaging software tools for HPC clusters and the Cloud 550
Darren Thompson, Alex Khassapov, Yakov Nesterets, Timur Gureyev, John Taylor

Verification and User Experience of High Data Rate Bandwidth-on-Demand networks 557
Chris Myers, Michael DeSilva

GridFTP based real-time data movement architecture for x-ray photon correlation spectroscopy at the Advanced Photon Source 559
Suresh Narayanan, Timothy Madden, Alec Sandy, Raj Kettimuthu, Michael Link

Data intensive science at synchrotron based 3D x-ray imaging facilities 567
Francesco De Carlo, Xianghui Xiao, Kamel Fezzaa, Steve Wang, Chris Jacobsen, Nikhilesh Chawla, Florian Fusseis

IRMIS: The care and feeding of a generalized relatively relational database for accelerator components with a connection to the real time EPICS Input output controllers 570
Richard Farnsworth, Scott Benes

Web Applications for Experimental Control in CLS 573
Dong Liu, Elder Mattias