

# **2012 IEEE 8th International Conference on E-Science**

**(e-Science 2012)**

**Chicago, Illinois  
8-12 October 2012**



**IEEE Catalog Number: CFP1206A-PRT**  
**ISBN: 978-1-4673-4467-8**

# 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 Gitmans, 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 Aiftimei, 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**

Mina Cintho, Roberto Marcondes Cesar, João Eduardo Ferreira

## Semantics

### **RightField: Semantic Enrichment of Systems Biology Data using Spreadsheets**

Katherine Wolstencroft, Stuart Owen, Carole Goble, Quyen Nguyen, Olga Krebs, Wolfgang Müller

### **MIM: A Minimum Information Model Vocabulary and Framework for Scientific Linked Data**

Matthew Gamble, Jun Zhao, Graham Klyne, Carole Goble

### **From Scripts Towards Provenance Inference**

Mohammad Rezwanaul 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**

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**

Jurandy Almeida, Jefersson Alex Dos Santos, Bruna Alberton, Ricardo Torres, Leonor Morellato

### **Flashes in a Star Stream: Automated Classification of Astronomical Transient Events**

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**

Irfan Azeezullah, Friska Pambudi, Tung-Kai Shyy, Nigel Ward, Jane Hunter, Robert Stimson

## Runtime & Middleware

### **P\*: A Model of Pilot-Abstractions**

Andre Luckow, Mark Santcroos, Ole Weidner, Andre Merzky, Pradeep Kumar Mantha, Shantenu Jha

### **MARISSA: MapReduce Implementation for Streaming Science Applications**

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**

Ryousei Takano, Hidemoto Nakada, Takahiro Hirofuchi, Yoshio Tanaka, Tomohiro Kudoh

### **Using Promethee Methods for Multi-Criteria Pull-based Scheduler on DCIs**

Mircea Moca, Gilles Fedak

## Data Infrastructures

### **A Data-driven Urban Research Environment for Australia**

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** &- %

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, Estanislao Gonzalez, Sebastian Denvil, Mark Morgan, Sandro Fiore, Zed Pobre, Roland Schweitzer

**CINET: A CyberInfrastructure for Network Science** "" \$%

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** "" \$-

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** "" %

Hui Zhang, Wenjun Wu, Zhenan Li

**Partial Replica Selection for Spatial Datasets** "" &+

Yun Tian, Philip Rhodes

**Image Retrieval in the Unstructured Data Management System AUDR** "" ' +

Junwu Luo, Bo Lang, Danchen Zhang, Chao Tian

## Application Systems & Frameworks

**A System for Management of Computational Fluid Dynamics Simulations for Civil Engineering** "" ((

Peter Sempolinski, Douglas Thain, Daniel Wei, Ahsan Kareem

**BIGS: A Framework for Large-Scale Image Processing and Analysis Over Distributed and Heterogeneous Computing Resources** "" ) &

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** "" \* \$

Christian Haas, Simon Caton, Daniel Trumpp, Christof Weinhardt

## Workflow 2

**Quality of Data Driven Simulation Workflows** "" \*,

Michael Reiter, Uwe Breitenbücher, Dimka Karastoyanova, Oliver Kopp

**A new Optimization Phase for Scientific Workflow Management Systems** "" +\*

Sonja Holl, Olav Zimmermann, Martin Hofmann-Apitius

**Discovering Drug Targets for Neglected Diseases Using a Pharmacophylogenomic Cloud Workflow**

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**

Kalev Leetaru

**Happy or Not : Generating Topic-based Geospatial Emotional Heatmaps for Culturomics using CyberGIS**

Eric Shook, Kalev Leetaru, Guofeng Cao, Anand Padmanabhan, Shaowen Wang

**High-Performance Computing without commitment**

Kevin Jorissen, John J. Rehr, Fernando D. Vila, William Johnson

**Lessons Learned from Galaxy, a Web-based Platform for High-throughput Genomic Analyses**

Jeremy Goecks, Galaxy Team, Anton Nekrutenko, James Taylor

**High-Performance Data Management For Genome Sequencing Centers Using Globus Online: A Case Study**

Dinanath Sulakhe, Rajkumar Kettimuthu, Utpal Dave

**FRED Navigator: An Interactive System for Visualizing Results from Large-scale Epidemic Simulations**

Jack Paparian, Shawn Brown, Donald Burke, John Grefenstette

**Digitization and Search: A Non-Traditional Use of HPC**

Liana Diesendruck, Luigi Marini, Rob Kooper, Mayank Kejriwal, Kenton Mchenry

**Large Scale Video Analytics: On-demand, iterative inquiry for moving image research**

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**

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**

Roscoe Bartlett, Micheal Heroux, James Willenbring

**IPOL: Reviewed Publication and Public Testing of Research Software**

Nicolas Limare, Laurent Oudre

**On Realizing the Concept Study ScienceSoft of the European Middleware Initiative - Open Software for Open Science**

Alberto Di Meglio, Morris Riedel

## eSoN 12

**ExSciTech: Expanding Volunteer Computing to Explore Science, Technology, and Health**

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**

Ryan Chard, Kris Bubendorfer, Kyle Chard

**A Satellite Data Portal Developed for Crowdsourcing Data Analysis and Interpretation**

Zhenghui Hu, Wenjun Wu

**g-Social: Enhancing Integrated eScience Tools with Social Networking Functionality**

Stylianou Andriani, Louloudes Nicholas, Dikaiakos Marios D.

**RunMyCode.org: a novel dissemination and collaboration platform for executing published computational results**

Victoria Stodden, Christophe Hurlin, Christophe Pérignon

**Towards a quantitative academic internationalization assessment of Brazilian research groups**

Evelyn Perez Cervantes, Jesús Mena-Chalco, Roberto Cesar

**Towards the Gamification of Well-being Measures**

Margaret 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**

Jan Bot, Migiel de Vos, Sander Boele, Marcel Reinders, Joost Kok

**Pilot Abstractions for Compute, Data, and Network**

Mark Santcroos, Delgado Olabarriga, Daniel S. Katz, Shantenu Jha

**Verification and User Experience of High Data Rate Bandwidth-on-Demand networks**

Jimmy Cullen, Richard Hughes-Jones, Ralph Spencer

**Scientific Workflows using Dynamically Provisioned High-Speed Circuits**

Paul Ruth, Anirban Mandal, Yufeng Xin, Ilia Badine, Chris Heerman, Jeff Chase

**Evaluation of RoCE over a 40 Gbps Transcontinental Layer-2 Circuit**

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** ( ) ( %

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** ( ) ) \$

Darren Thompson, Alex Khassapov, Yakov Nesterets, Timur Gureyev, John Taylor

**Verification and User Experience of High Data Rate Bandwidth-on-Demand networks** ( ) ) +

Chris Myers, Michael DeSilva

**GridFTP based real-time data movement architecture for x-ray photon correlation spectroscopy at the Advanced Photon Source** ( ) ) -

Suresh Narayanan, Timothy Madden, Alec Sandy, Raj Kettimuthu, Michael Link

**Data intensive science at synchrotron based 3D x-ray imaging facilities** ( ) ) \* +

Francesco De Carlo, Xianghui Xiao, Kamel Fezzaa, Steve Wang, Chris Jacobsen, Nikhilesh Chawla, Florian Füsseis

**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** ( ) ) + \$

Richard Farnsworth, Scott Benes

**Web Applications for Experimental Control in CLS** ( ) ) +'

Dong Liu, Elder Mattias