2015 IEEE 11th International Conference on e-Science (e-Science 2015)

Munich, Germany
31 August – 4 September 2015

IEEE

IEEE Catalog Number: CFP1506A-POD
ISBN: 978-1-4673-9326-3

2015 IEEE 11th International Conference on eScience

eScience 2015

Table of Contents

Welcome from the General Chairs.....xii

Committees.....xiii

Program Committee.....xiv

Monday

Adventures of Categories: Modelling the Evolution of Categories During Scientific Investigation 1
Prashant Gupta, Mark Gahegan, and Gillian Dobbie

Beyond the Book: Linking Books to Wikipedia 12
Carlos Martinez-Ortiz, Marijn Koolen, Floor Buschenhenke, and Karina van Dalen-Oskam

HEEM, a Complex Model for Mining Emotions in Historical Text 22
Janneke M. van der Zwaan, Inger Leemans, Erika Kuijpers, and Isa Maks

Swedish eScience Education — A Graduate School in eScience 31
Anders Hast, Michael Hanke, and Hans O. Karlsson

Using Text Similarity to Detect Social Interactions not Captured by Formal Reply Mechanisms 36
Samuel Barbosa, Roberto M. Cesar-Jr, and Dan Cosley

Acceleration-as-a-Service: Exploiting Virtualised GPUs for a Financial Application 47
Blesson Varghese, Javier Prades, Carlos Reaño, and Federico Silla

An Integrated Approach to Porting Large Scientific Applications to GPUs 57
Ben van Werkhoven and Pieter Hijma

Application Support for Virtual GPGPUs in Grid Infrastructures 67
John Walsh and Jonathan Dukes

TemPSS: A Service Providing Software Parameter Templates and Profiles for Scientific HPC 78
Jeremy Cohen, Chris Cantwell, David Moxey, Jeremy Nowell, Peter Austing, Xu Guo, John Darlington, and Spencer Sherwin

Enabling SDMX-based Retrieval and Spatio-statistical Analysis of National Census and Related Datasets 88
Jane Hunter, Imran Azeezullah, Nigel Ward, Ross Barker, Tung-Kai Shyy, Chris Beer, Stuart Girvan, Alister Nairn, Merry Branson, Robert Stimson, James Dentrinos, Gerson Galang, Stewart Wallace, and Chris Pettit

Finding Pulsars in Real-Time 98
Alessio Sclocco, Henri E. Bal, and Rob V. van Nieuwpoort

From HPC Performance to Climate Modeling: Transforming Methods for HPC Predictions into Models of Extreme Climate Conditions 108
Ryan McKinney, Vivek K. Pallipuram, Rodrigo Vargas, and Michela Taufer

Multidisciplinary Collaboration to Facilitate Hypotheses Generation in Huntington's Disease 118
Eleni Mina, Mark Thompson, Kristina M. Hettne, Willeke van Roon-Mom, Rajaram Kaliyaperumal, Eelke van der Horst, Katherine Wolstencroft, Barend Mons, and Marco Roos

Bootstrapping Complex Workflow Middleware Systems into the Cloud 126
Karolina Vukojevic-Haupt, Florian Haupt, Frank Leymann, and Lukas Reinfurt

Cost-Aware Cloud Provisioning 136
Ryan Chard, Kyle Chard, Kris Bubendorfer, Lukasz Lacinski, Ravi Madduri, and Ian Foster

Deployment of a Multi-site Cloud Environment for Molecular Virtual Screenings 145
Anthony Nguyen, Andréa Matsunaga, Maurício Tsugawa, Susumu Date, Kohei Ichikawa, and Jason H. Haga

Federating Infrastructure as a Service Cloud Computing Systems to Create a Uniform E-infrastructure for Research 155
David C. H. Wallom, Matteo Turilli, Michel Drescher, Diego Scardaci, and Steven Newhouse

Tuesday

A Round Table for Multi-disciplinary Research on Geospatial and Climate Data 165
Romulo Goncalves, Milena Ivanova, Foteini Alvanaki, Jason Maassen, Kostis Kyzirakos, Oscar Martinez-Rubi, and Hannes Mühleisen

eScience through the Integration of Data and Models: A Biodiversity Scenario 171
Natalia Villanueva-Rosales, Luis Garnica Chavira, Nicholas del Rio, and Deana Pennington

FishGraph: A Network-Driven Data Analysis 177
Patrícia Cavoto, Victor Cardoso, Régine Vignes Lebbe, and André Santanchè

A Comparison of Background Subtraction Algorithms for Detecting Avian Nesting Events in Uncontrolled Outdoor Video 187
Kyle Goehner, Travis Desell, Rebecca Eckroad, Leila Mohsenian, Paul Burr, Nicolas Caswell, Alicia Andes, and Susan Ellis-Felege

Discovering Loose Group Movement Patterns from Animal Trajectories 196
Yuwei Wang, Ze Luo, Yan Xiong, Diann J. Prosser, Scott H. Newman, John Y. Takekawa, and Baoping Yan

WRF-ARW Model for Prediction of High Temperatures in South and South East Regions of Armenia 207
H. Astsatryan, A. Shakhnazaryan, V. Sahakyan, Yu. Shoukourian, V. Kotroni, Z. Petrosyan, R. Abrahamyan, and H. Melkonyan

Desktop as a Service Supporting Environmental 'omics 214
David C. H. Wallom, Timothy Booth, Andy Bowery, Ben Collier, Phillip Kershaw, Anurag Priyam, Yannick Wurm, and Dawn Field

Live ANDES: Mobile-Cloud Shared Workspace for Citizen Science and Wildlife Conservation 215
Cristian Bonacic, Andres Neyem, and Andrea Vasquez

VERCE Delivers a Productive E-science Environment for Seismology Research 224
Malcolm Atkinson, Michele Carpené, Emanuele Casarotti, Steffen Claus, Rosa Filgueira, Anton Frank, Michelle Galea, Tom Garth, André Gemünd, Heiner Igel, Iraklis Klampanos, Amrey Krause, Lion Krischer, Siew Hoon Leong, Federica Magnoni, Jonas Matser, Alberto Michelini, Andreas Rietbrock, Horst Schwichtenberg, Alessandro Spinuso, and Jean-Pierre Vilotte

Acoustic Feature Extraction Using Perceptual Wavelet Packet Decomposition for Frog Call Classification 237
Jie Xie, Michael Towsey, Philip Eichinski, Jinglan Zhang, and Paul Roe

Summer in the City: Forecasting and Mapping Human Thermal Comfort in Urban Areas 243
J. J. Attema, B. G. Heusinkveld, R. J. Ronda, G. J. Steeneveld, and A. A. M. Holtslag

Transforming Geodata for Immersive Visualisation 249
Markus Wiedemann, Christoph Anthes, Hans-Peter Bunge, Bernhard S.A. Schuberth, and Dieter Kranzlmülle

Poster Session

Data Access and Reproducibility in Privacy Sensitive eScience Domains 255
Stefan Pröell, Rudolf Mayer, and Andreas Rauber

Applying a Multi-dimensional Time-Series Similarity Method to Typhoon-track Prediction 259
Yu Fang, Kosuke Sugano, Kenta Oku, and Kyoji Kawagoe

Autocuration Cyberinfrastructure for Scientific Discovery and Preservation 264
Smruti Padhy, Edgar Black, Betsy Cowdery, Liana Diesendruck, Michael Dietze, Greg Jansen, Rob Kooper, Praveen Kumar, Jong Lee, Rui Liu, Richard Marciano, Luigi Marini, Dave Mattson, Barbara Minsker, Chris Navarro, Ankit Rai, Marcus Slavenas, William Sullivan, Jason Votava, Qina Yan, Inna Zharnitsky, and Kenton McHenry

Building an Institution Digital Repository: The ASI Science Data Center Bibliography Tool 268
Federica Moscato, Cristina Leto, and Paolo Giommi

Classifying Bone Trabecular Volumes Using 3D Shape Descriptors 271
Manuel Herrera, Carlos Martinez-Ortiz, Patricia Carvalho-Lobato, and Maria-Cristina Manzanares-Cespedes

CloudBench — A Framework for Distributed, Self-organizing, Continuous and Topology-aware IaaS Cloud Benchmarking with Super-peer Networks 273
Peter Krauss, Tobias Kurze, and Achim Streit

Computer-Assisted Workflow Composition Based on Virtual Simulation Objects Technology 279
Pavel A. Smirnov and Sergey V. Ivanov

Constructing Scalable Data-Flows on Hadoop with Legacy Components 283
Rainer Schmidt, Matthias Rella, and Sven Schlarb

Forecasting Auroral Substorms from Observed Data with a Supervised Learning Algorithm 284
Takanori Tanaka, Daisuke Kitao, Yuka Sato, Yoshimasa Tanaka, and Daisuke Ikeda

Interactive Analytic Systems for Understanding the Scholarly Impact of Large-Scale E-science Cyberenvironments 288
Krishna Madhavan, Daniel F. Mejia, Hanjun Xian, Lynn K. Zentner, Victoria A. Farnsworth, Swaroop Samek, and Gerhard Klimeck

Negotiation Protocol for Agile and Reliable E-science Collaboration 292
Zeqian Meng and John Brooke

Presenting the Cloud as Files and Directories 296
Vitalian A. Danciu

Provenance-driven Representation of Crowdsourcing Data for Efficient Data Analysis 300
Carlos Martinez-Ortiz, Lora Aroyo, Oana Inel, Stavros Champilomatis, Anca Dumitrache, and Benjamin Timmermans

Shape Optimisation of Architectural Shell with Algae Photoreactors 304
Natalia Buzalo and Pavel Ermachenko

Web-Based Manipulation of Multiresolution Micro-CT Images 308
Lasse Wollatz, Simon Cox, and Steven Johnston

Wednesday

A Quantitative Study on the Re-executability of Publicly Shared Scientific Workflows 312
Rudolf Mayer and Andreas Rauber

Data Analytics in Bioinformatics: Data Science in Practice for Genomics Analysis Workflows 322
Kary A. C. S. Ocaña, Vítor Silva, Daniel de Oliveira, and Marta Mattoso

Scaling Up Bioinformatics Workflows with Dynamic Job Expansion: A Case Study Using Galaxy and Makeflow 332
Nicholas Hazekamp, Joseph Sarro, Olivia Choudhury, Sandra Gesing, Scott Emrich, and Douglas Thain

Streaming Algorithms for Halo Finders 342
Zaoxing Liu, Nikita Ivkin, Lin Yang, Mark Neyrinck, Gerard Lemson, Alexander Szalay, Vladimir Braverman, Tamas Budavari, Randal Burns, and Xin Wang

A Task-Centered Framework for Computationally-Grounded Science Collaborations 352
Yolanda Gil, Felix Michel, Varun Ratnakar, Matheus Hauder, Christopher Duffy, Hilary Dugan, and Paul Hanson

Cloud-based E-Infrastructure for Scheduling Astronomical Observations 362
James Wetter, Özgur Akgun, Adam Barker, Martin Dominik, Ian Miguel, and Blesson Varghese

Managing Complexity in Distributed Data Life Cycles Enhancing Scientific Discovery 371
Richard Grunzke, Alvaro Aguilera, Wolfgang E. Nagel, Jens Krüger, Sonja Herres-Pawlis, Alexander Hoffmann, and Sandra Gesing

Monitoring of Upper Limb Rehabilitation and Recovery after Stroke: An Architecture for a Cloud-Based Therapy Platform 381
Simon Woodman, Hugo Hiden, Mark Turner, Stephen Dowsland, and Paul Watson

Genome Analysis in a Dynamically Scaled Hybrid Cloud 391
Christopher Smowton, Georgiana Copil, Hong-Linh Truong, Crispin Miller, and Wei Xing

Globus Data Publication as a Service: Lowering Barriers to Reproducible Science 401
Kyle Chard, Jim Pruyne, Ben Blaiszik, Rachana Ananthakrishnan, Steven Tuecke, and Ian Foster

Identification of Biomarkers and Signatures in Protein Data 411
Torbjörn E. M. Nordling, Narendra Padhan, Sven Nelander, and Lena Claesson-Welsh

Porting Ordinary Applications to Blue Gene/Q Supercomputers 420
Ketan Maheshwari, Justin M. Wozniak, Timothy G. Armstrong, Daniel S. Katz, T. Andrew Binkowski, Xiaoliang Zhong, Olle Heinonen, Dmitry Karpeyev, and Michael Wilde

Searching the Human Genome for Snail and Slug with DNA@Home 429
Kristopher Zarns, Travis Desell, Sergei Nechaev, and Archana Dhasarathy

A Multipath Controller for Accelerating GridFTP Transfer over SDN 439
Che Huang, Chawanat Nakasan, Kohei Ichikawa, and Hajimu Iida

B2SHARE: An Open eScience Data Sharing Platform 448
Sarah Berenji Ardestani, Carl Johan Håkansson, Erwin Laure, Ilja Livenson, Pavel Straňák, Emanuel Dima, Dennis Blommesteijn, and Mark van de Sanden

dispel4py: An Agile Framework for Data-Intensive eScience 454
Rosa Filgueira, Amrey Krause, Malcolm Atkinson, Iraklis Klampanos, Alessandro Spinuso, and Susana Sanchez-Exposito

MemEFS: An Elastic In-memory Runtime File System for eScience Applications 465
Alexandru Uta, Andreea Sandu, Stefania Costache, and Thilo Kielmann

SODA: Science-Driven Orchestration of Data Analytics 475
Jai Dayal, Jay Lofstead, Greg Eisenhauer, Karsten Schwan, Matthew Wolf, Hasan Abbasi, and Scott Klasky

First International Workshop on Interoperable Infrastructures for Interdisciplinary Big Data Sciences (IT4RIs 15)

Towards Sustainable Curation and Preservation: The SEAD Project's Data Services Approach 485
James Myers, Margaret Hedstrom, Dharma Akmon, Sandy Payette, Beth A. Plale, Inna Kouper, Scott McCaulay, Robert McDonald, Isuru Suriarachchi, Aravindh Varadharaju, Praveen Kumar, Mostafa Elag, Jong Lee, Rob Kooper, and Luigi Marini

Enhanced Usability of Managing Workflows in an Industrial Data Gateway 495
Gary A. McGilvary, Malcolm Atkinson, Sandra Gesing, Alvaro Aguilera, Richard Grunzke, and Eva Sciacca

A Semantic Peer to Peer Network to Support e-Science 503
Tadeu Classe, Regina Braga, Fernanda Campos, and José Maria Nazar David

Open Information Linking for Environmental Research Infrastructures 513
Paul Martin, Paola Grosso, Barbara Magagna, Herbert Schentz, Yin Chen, Alex Hardisty, Wouter Los, Keith Jeffery, Cees de Laat, and Zhiming Zhao

WIP: Provenance Support for Interdisciplinary Research on the North Creek Wetlands 521
Qi Zhang, Jonathan Mason, Morteza Chini, Karen Potts, Nathan Duncan, Delmar B. Davis, and Hazeline U. Asuncion

Interoperability Oriented Architecture: The Approach of EPOS for Solid Earth e-Infrastructures 529
Daniele Bailo, Keith G. Jeffery, Alessandro Spinuso, and Giuseppe Fiameni

A Community Contribution Framework for Sharing Materials Data with Materials Project 535
Patrick Huck, Anubhav Jain, Dan Gunter, Donald Winston, and Kristin Persson

Multi-node Multi-agent Cloud Simulation: Approximating Synchronisation 542
Antonio Giardina, Yun Yang, Hai Vu, and Rajesh Vasa

Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures 551
Zhiming Zhao, Paul Martin, Paola Grosso, Wouter Los, Cees de Laat, Keith Jeffrey, Alex Hardisty, Alex Vermeulen, Donatella Castelli, Yannick Legré, and Werner Kutsch

First Workshop on E-science ReseaRch leading tO negative Results (ERROR)

Scientific Workflow Interchanging through Patterns: Reversals and Lessons Learned 557
Bruno Fernandes Bastos, Regina Maria Maciel Braga, and Antônio Tadeu Azevedo Gomes

From Thread to Transcontinental Computer: Disturbing Lessons in Distributed Supercomputing 565
Derek Groen and Simon Portegies Zwart

Early Experiences with Separate Caches for Private and Shared Data 572
Juan M. Cebrián, Alberto Ros, Ricardo Fernández-Pascual, and Manuel E. Acacio

A Ping Too Far: Real World Network Latency Measurement 580
Gary Jackson, Pete Keleher, and Alan Sussman

Author Index 589