

2016 IEEE 12th International Conference on e-Science (e-Science 2016)

**Baltimore, Maryland, USA
23-27 October 2016**



**IEEE Catalog Number: CFP1606A-POD
ISBN: 978-1-5090-4274-6**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP1606A-POD
ISBN (Print-On-Demand):	978-1-5090-4274-6
ISBN (Online):	978-1-5090-4273-9

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2016 IEEE 12th International Conference on e-Science e-Science 2016

Table of Contents

Welcome Message from the Chairs.....	i
Committees.....	ii
Steering Committee.....	iii
Program Committee.....	iv
Sponsors	vii
Author Index	450

Paper Session 1: Cloud Computing

Campus Compute Co-Operative (CCC): A Service Oriented Cloud Federation	1
<i>Andrew Grimshaw, Md Anindya Prodhan, Alexander Thomas, Craig Stewart, Richard Knepper</i>	
Albatross: An Efficient Cloud-Enabled Task Scheduling and Execution Framework Using Distributed Message Queues	11
<i>Iman Sadooghi, Geet Kumar, Ke Wang, Dongfang Zhao, Tonglin Li, Ioan Raicu</i>	
Prediction of Workflow Execution Time Using Provenance Traces: Practical Applications in Medical Data Processing.....	21
<i>Hugo Hiden, Simon Woodman, Paul Watson</i>	

Paper Session 2: Curation

Accelerating Data-Driven Discovery with Scientific Asset Management.....	31
<i>Robert E. Schuler, Carl Kesselman, Karl Czajkowski</i>	
Cooperative Human-Machine Data Extraction from Biological Collections	41
<i>Icaro Alzuru, Andréa Matsunaga, Maurício Tsugawa, José A.B. Fortes</i>	
Generating Knowledge Networks from Phenotypic Descriptions	51
<i>Fagner Leal Pantoja, Patrícia Cavoto, Julio Cesar dos Reis, André Santanchè</i>	

Paper Session 3: Reproducibility

PRUNE: A Preserving Run Environment for Reproducible Scientific Computing.....	61
<i>Peter Ivie, Douglas Thain</i>	
Converting Scripts into Reproducible Workflow Research Objects	71
<i>Lucas A.M.C. Carvalho, Khalid Belhajjame, Claudia Bauzer Medeiros</i>	
A Framework for Scientific Workflow Reproducibility in the Cloud	81
<i>Rawaa Qasha, Jacek Cała, Paul Watson</i>	
Conducting Reproducible Research with Umbrella: Tracking, Creating, and Preserving Execution Environments.....	91
<i>Haiyan Meng, Douglas Thain, Alexander Vyushkov, Matthias Wolf, Anna Woodard</i>	

Paper Session 4: Algorithms

Fast Window Aggregate on Array Database by Recursive Incremental Computation.....	101
<i>Li Jiang, Hideyuki Kawashima, Osamu Tatebe</i>	
Representation of Continuously Changing Data Over Time and Space: Modeling the Shape of Spatiotemporal Phenomena.....	111
<i>José Moreira, Paulo Dias, Pedro Amaral</i>	
An n-Gram Cache for Large-Scale Parallel Extraction of Multiword Relevant Expressions with LocalMaxs.....	120
<i>Carlos Goncalves, Joaquim F. Silva, Jose C. Cunha</i>	
DOT-K: A Distributed Online Top-K Elements Algorithm Using Extreme Value Statistics	130
<i>Nicholas Carey, Tamás Budavári, Yanif Ahmad, Alexander S. Szalay</i>	

Paper Session 5: Workflow

Budget Distribution Strategies for Scientific Workflow Scheduling in Commercial Clouds	137
<i>Vahid Arabnejad, Kris Bubendorfer, Bryan Ng</i>	
Datatrack: An R Package for Managing Data in a Multi-Stage Experimental Workflow — Data Versioning and Provenance Considerations in Interactive Scripting.....	147
<i>Philip Eichinski, Paul Roe</i>	
Starting Workflow Tasks Before They're Ready	155
<i>Wladislaw Gusew, Björn Scheuermann</i>	

Paper Session 6: Potpourri and Science of Science

Automatic Glomerulus Extraction in Whole Slide Images Towards Computer Aided Diagnosis	165
<i>Yan Zhao, Edgar F. Black, Luigi Marini, Kenton McHenry, Norma Kenyon, Rachana Patil, Andre Balla, Amelia Bartholomew</i>	
Improving Data Provenance Reconstruction via a Multi-Level Funneling Approach.....	175
<i>Subha Vasudevan, William Pfeffer, Delmar Davis, Hazeline Asuncion</i>	
Plans and Performances: Parallels in the Production of Science and Music	185
<i>David De Roure, Graham Klyne, Kevin R. Page, John Pybus, David M. Weigl, Matthew Wilcoxson, Pip Willcox</i>	

Paper Session 7: New Platforms

MOHA: Many-Task Computing Meets the Big Data Platform.....	193
<i>Jik-Soo Kim, Cao Nguyen, Soonwook Hwang</i>	
Globus Auth: A Research Identity and Access Management Platform	203
<i>Steven Tuecke, Rachana Ananthakrishnan, Kyle Chard, Mattias Lidman, Brendan McCollam, Stephen Rosen, Ian Foster</i>	
Management, Analysis, and Visualization of Experimental and Observational Data — The Convergence of Data and Computing	213
<i>E. Wes Bethel, Martin Greenwald, Kerstin Kleese van Dam, Manish Parashar, Stefan M. Wild, H. Steven Wiley</i>	
Developing a Citizen Science Web Portal for Manual and Automated Ecological Image Detection.....	223
<i>Marshall Mattingly III, Andrew Barnas, Susan Ellis-Felege, Robert Newman, David Iles, Travis Desell</i>	

Paper Session 8: Machine Learning Applications

Applying Data Mining Methods for the Analysis of Stable Isotope Data in Bioarchaeology	233
<i>Markus Mauder, Eirini Ntoutsis, Peer Kröger, Christoph Mayr, Gisela Grupe, Anita Toncala, Stefan Hölzl</i>	
Data-Oriented Neuron Classification from Their Parts	243
<i>Evelyn Perez Cervantes, Cesar Henrique Comin, Roberto Marcondes Cesar Junior, Luciano da Fontoura Costa</i>	
Detecting Wildlife in Uncontrolled Outdoor Video Using Convolutional Neural Networks	251
<i>Connor Bowley, Alicia Andes, Susan Ellis-Felege, Travis Desell</i>	
Using LSTM Recurrent Neural Networks to Predict Excess Vibration Events in Aircraft Engines	260
<i>AbdElRahman ElSaid, Brandon Wild, James Higgins, Travis Desell</i>	

Invited Short Papers: Hot Topics 1

SIMP: Secure Interoperable Multi-Granular Provenance Framework	270
<i>Amani Abu Jabal, Elisa Bertino</i>	
A Comprehensive Scenario Agnostic Data LifeCycle Model for an Efficient Data Complexity Management	276
<i>Amir Sinaeepourfard, Jordi Garcia, Xavier Masip-Bruin, Eva Marín-Tordera</i>	
Semantic Accountable Matchmaking for E-Science Resource Sharing	282
<i>Zeqian Meng, John Brooke, Rizos Sakellariou</i>	
Apache Airavata Security Manager: Authentication and Authorization Implementations for a Multi-Tenant eScience Framework	287
<i>Supun Nakandala, Hasini Gunasinghe, Suresh Marru, Marlon Pierce</i>	

Paper Session 9: Social Analytics and Population Simulation

Privacy-Protected Social Media User Trajectories Calibration	293
<i>Shuo Wang, Richard Sinnott, Surya Nepal</i>	
A Machine Learning Analysis of Twitter Sentiment to the Sandy Hook Shootings	303
<i>Nan Wang, Blesson Varghese, Peter D. Donnelly</i>	
A Hybrid Approach to Population Construction for Agricultural Agent-Based Simulation	313
<i>Peng Chen, Tom Evans, Michael Frisby, Eduardo Izquierdo, Beth Plale</i>	
A Parallel Microsimulation Package for Modelling Cancer Screening Policies	323
<i>Andreas Karlsson, Niten Olofsson, Erwin Laure, Mark Clements</i>	

Invited Short Papers: Hot Topics 2

OntoSoft: A Distributed Semantic Registry for Scientific Software	331
<i>Yolanda Gil, Daniel Garijo, Saurabh Mishra, Varun Ratnakar</i>	
A Secure Data Enclave and Analytics Platform for Social Scientists	337
<i>Yadu N. Babuji, Kyle Chard, Aaron Gerow, Eamon Duede</i>	
Reproducibility in Computer Vision: Towards Open Publication of Image Analysis Experiments as Semantic Workflows	343
<i>Ricky J. Sethi, Yolanda Gil</i>	
Crossing Analytics Systems: A Case for Integrated Provenance in Data Lakes	349
<i>Isuru Suriarachchi, Beth Plale</i>	
Interactive Provenance Summaries for Reproducible Science	355
<i>Xiang Li, Xiaoyang Xu, Tanu Malik</i>	

Paper Session 10: High-Performance Computing

ExTASY: Scalable and Flexible Coupling of MD Simulations and Advanced Sampling Techniques	361
<i>Vivekanandan Balasubramanian, Iain Bethune, Ardita Shkurti, Elena Breitmoser, Eugen Hruska, Cecilia Clementi, Charles Laughton, Shantenu Jha</i>	
Leveraging Burst Buffer Coordination to Prevent I/O Interference	371
<i>Anthony Kougkas, Matthieu Dorier, Rob Latham, Rob Ross, Xian-He Sun</i>	
A Fast Algorithm for Neutrally-Buoyant Lagrangian Particles in Numerical Ocean Modeling	381
<i>Renske Gelderloos, Alexander S. Szalay, Thomas W.N. Haine, Gerard Lemson</i>	
Data Management and Simulation Support Accelerating Carbon Capture Through Computing.....	389
<i>You-Wei Cheah, Joshua Boverhof, Abdelrahman Elbashandy, Deb Agarwal, Jim Leek, Thomas Epperly, John Eslick, David Miller</i>	

ECW 2016 — Environmental Computing Workshop

Session 1: Tools and Platforms

OMUSE: Oceanographic Multipurpose Software Environment.....	399
<i>Inti Pelupessy, Ben van Werkhoven, Arjen van Elteren, Jan Viebahn, Adam Candy, Simon Portegies Zwart, Henk Dijkstra</i>	
Automating Environmental Computing Applications with Scientific Workflows.....	400
<i>Rafael Ferreira da Silva, Ewa Deelman, Rosa Filgueira, Karan Vahi, Mats Rynge, Rajiv Mayani, Benjamin Mayer</i>	
Utilising Amazon Web Services to Provide an On Demand Urgent Computing Facility for climateprediction.net.....	407
<i>Peter Uhe, Friederike E.L. Otto, Md Mamunur Rashid, David C.H. Wallom</i>	

Session 2: Data

Automatic Fire Perimeter Determination Using MODIS Hotspots Information	414
<i>N. Chiaraviglio, T. Artés, R. Bocca, J. López, A. Gentile, J. San Miguel Ayanz, A. Cortés, T. Margalef</i>	
A Columnar Architecture for Modern Risk Management Systems	424
<i>Romulo Goncalves, Sisi Zlatanova, Kostis Kyzirakos, Pirouz Nourian, Foteini Alvanaki, Willem van Hage</i>	
The eWaterCycle Project.....	430
<i>Niels Drost, Rolf Hut, Maarten van Meersbergen, Edwin Sutanudjaja, Marc Bierkens, Nick van de Giesen</i>	
The IQmulus High Volume Fusion and Analysis Platform for Geospatial Point Clouds, Featuring a Marine Bathymetry Use Case	431
<i>Quillon Harpham</i>	

Session 3: Case Studies

Manufacturing of Weather Forecasting Simulations on High Performance Infrastructures.....	432
<i>V. Šipková, L. Hluchý, M. Dobrucký, J. Bartok, B.M. Nguyen</i>	
The Climate Change for Flood and Debris Mitigation After Typhoon Morakot 2009 in Taiwan.....	440
<i>Harold Yih-Chi Tan</i>	
Performance Evaluation of Environmental Applications Using TELEMAC-MASCARET on Virtual Platforms.....	441
<i>Minh Thanh Chung, Manh-Thin Nguyen, Nhu-Y Nguyen-Huynh, Thong Nguyen, Nam Thoai</i>	
Environmental Computing — Applications, Tools and Data Solutions.....	449
<i>Matti Heikkurinen, Dieter Kranzlmüller</i>	