

2021 IEEE Workshop on Workflows in Support of Large-Scale Science (WORKS 2021)

**St. Louis, Missouri, USA
15 November 2021**



**IEEE Catalog Number: CFP21A54-POD
ISBN: 978-1-6654-1137-0**

**Copyright © 2021 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:	CFP21A54-POD
ISBN (Print-On-Demand):	978-1-6654-1137-0
ISBN (Online):	978-1-6654-1136-3

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

2021 IEEE Workshop on Workflows in Support of Large-Scale Science (WORKS) **WORKS 2021**

Table of Contents

Message from the Workshop Chairs	vi
Workshop Organization	vii

Scheduling and Resource Provisioning

A Recommender System for Scientific Datasets and Analysis Pipelines	1
<i>Mandana Mazaheri (Concordia University, Montreal, Canada), Gregory Kiar (Center for the Developing Brain, Child Mind Institute, USA), and Tristan Glatard (Concordia University, Montreal, Canada)</i>	
Intelligent Resource Provisioning for Scientific Workflows and HPC	9
<i>Benjamin T. Shealy (Clemson University, USA), Frank Feltus (Clemson University, USA), and Melissa Smith (Clemson University, USA)</i>	
Not All Tasks Are Created Equal: Adaptive Resource Allocation for Heterogeneous Tasks in Dynamic Workflows	17
<i>Thanh Son Phung (University of Notre Dame), Logan Ward (Argonne National Laboratory), Kyle Chard (University of Chicago), and Douglas Thain (University of Notre Dame)</i>	

In-situ Processing

Dynamic Heterogeneous Task Specification and Execution for In Situ Workflows	25
<i>Orcun Yildiz (Argonne National Laboratory, USA), Dmitriy Morozov (Lawrence Berkeley National Laboratory, CA), Bogdan Nicolae (Argonne National Laboratory, USA), and Tom Peterka (Argonne National Laboratory, USA)</i>	
An Adaptive Elasticity Policy For Staging Based In-Situ Processing	33
<i>Zhe Wang (Rutgers University), Matthieu Dorier (Argonne National Laboratory), Pradeep Subedi (University of Utah), Philip E. Davis (University of Utah), and Manish Parashar (University of Utah)</i>	

Scalability

The Benefits of Prefetching for Large-Scale Cloud-based Neuroimaging Analysis Workflows	42
<i>Valerie Hayot-Sasson (Concordia University, Canada), Tristan Glatard (Concordia University, Canada), and Ariel Rokem (University of Washington, USA)</i>	
ExaWorks: Workflows for Exascale	50
<i>Aymen Al-Saadi (The State University of New Jersey, USA), Dong H. Ahn (Lawrence Livermore National Laboratory, USA), Yadu Babuji (Argonne National Laboratory, USA; The University of Chicago, USA), Kyle Chard (Argonne National Laboratory, USA; The University of Chicago, USA), James Corbett (Lawrence Livermore National Laboratory, USA), Mihael Hategan (Argonne National Laboratory, USA; The University of Chicago, USA), Stephen Herbein (Lawrence Livermore National Laboratory, USA), Shantenu Jha (Brookhaven National Laboratory, USA; The State University of New Jersey, USA), Daniel Laney (Lawrence Livermore National Laboratory, USA), Andre Merzky (Brookhaven National Laboratory, USA), Todd Munson (Argonne National Laboratory, USA), Michael Salim (Argonne National Laboratory, USA), Mikhail Titov (Brookhaven National Laboratory, USA), Matteo Turilli (The State University of New Jersey, USA; Brookhaven National Laboratory, USA), Thomas Uram (Argonne National Laboratory, USA), and Justin M. Wozniak (Argonne National Laboratory, USA)</i>	

Performance Characterization

A Performance Characterization of Scientific Machine Learning Workflows	58
<i>Patrycja Krawczuk (University of Southern California, USA), George Papadimitriou (University of Southern California, USA), Ryan Tanaka (University of Southern California, USA), Tu Mai Anh Do (University of Southern California, USA), Srujana Subramanya (University of Southern California, USA), Shubham Nagarkar (University of Southern California, USA), Aditi Jain (University of Southern California, USA), Kelsie Lam (University of Southern California, USA), Anirban Mandal (Renaissance Computing Institute, USA), Loïc Pottier (University of Southern California, USA), and Ewa Deelman (University of Southern California, USA)</i>	
Science Capsule: Towards Sharing and Reproducibility of Scientific Workflows	66
<i>Devarshi Ghoshal (Lawrence Berkeley National Laboratory, CA), Ludovico Bianchi (Lawrence Berkeley National Laboratory, CA), Abdelilah Essiari (Lawrence Berkeley National Laboratory, CA), Drew Paine (Lawrence Berkeley National Laboratory, CA), Sarah S. Poon (Lawrence Berkeley National Laboratory, CA), Michael Beach (University of Washington, WA), Alpha T. N'Diaye (Lawrence Berkeley National Laboratory, CA), Patrick Huck (Lawrence Berkeley National Laboratory, CA), and Lavanya Ramakrishnan (Lawrence Berkeley National Laboratory, CA)</i>	

Community

Emerging Frameworks for Advancing Scientific Workflows Research, Development, and Education	74
<i>Henri Casanova (University of Hawaii, USA), Ewa Deelman (University of Southern California, USA), Sandra Gesing (University of Illinois Chicago, USA), Michael Hildreth (University of Notre Dame, USA), Stephen Hudson (Argonne National Laboratory, USA), William Koch (University of Hawaii, USA), Jeffrey Larson (Argonne National Laboratory, USA), Mary Ann McDowell (University of Notre Dame, USA), Natalie Meyers (University of Notre Dame, USA), John-Luke Navarro (Argonne National Laboratory), George Papadimitriou (University of Southern California, USA), Ryan Tanaka (University of Southern California, USA), Ian Taylor (University of Notre Dame, USA), Douglas Thain (University of Notre Dame, USA), Stefan M. Wild (Argonne National Laboratory, USA), Rosa Filgueira (Heriot-Watt University, UK), and Rafael Ferreira da Silva (Oak Ridge National Laboratory, USA)</i>	
A Community Roadmap for Scientific Workflows Research and Development	81
<i>Rafael Ferreira da Silva (Oak Ridge National Laboratory, USA; University of Southern California, USA), Henri Casanova (University of Hawaii, USA), Kyle Chard (Argonne National Laboratory, USA; The University of Chicago, USA), Ilkay Altintas (University of California, USA), Rosa M Badia (Barcelona Supercomputing Center, Spain), Bartosz Balis (AGH University of Science and Technology, Poland), Tainā Coleman (University of Southern California, USA), Frederik Coppens (University of Hawaii, Honolulu, USA; VIB Center for Plant Systems Biology, Belgium), Frank Di Natale (Lawrence Livermore National Lab, USA), Bjoern Enders (Lawrence Berkeley National Lab, USA), Thomas Fahringer (University of Innsbruck, Austria), Rosa Filgueira (Heriot-Watt University, UK), Grigori Fursin (OctoML, USA), Daniel Garijo (Universidad Politécnica de Madrid, Spain), Carole Goble (The University of Manchester, UK), Dorrán Howell (Tweag, Switzerland), Shantenu Jha (Brookhaven National Laboratory, USA), Daniel S. Katz (University of Illinois at Urbana-Champaign, USA), Daniel Laney (Lawrence Livermore National Lab, USA), Ulf Leser (Humboldt-Universität zu Berlin, Germany), Maciej Malawski (AGH University of Science and Technology, Poland), Kshitij Mehta (Oak Ridge National Laboratory, USA), Loïc Pottier (University of Southern California, USA), Jonathan Ozik (Argonne National Laboratory, USA; The University of Chicago, USA), J. Luc Peterson (Lawrence Livermore National Lab, USA), Lavanya Ramakrishnan (Lawrence Berkeley National Lab, USA), Stian Soiland-Reyes (The University of Manchester, UK; University of Amsterdam, The Netherlands), Douglas Thain (University of Notre Dame, USA), and Matthew Wolf (Oak Ridge National Laboratory, USA)</i>	
Author Index	91