# 2022 IEEE/ACM Workshop on Workflows in Support of Large-Scale Science (WORKS 2022)

## Dallas, Texas, USA
## 13 – 18 November 2022

IEEE

**IEEE Catalog Number:** CFP22A54-POD
**ISBN:** 978-1-6654-5192-5

**Copyright © 2022 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: CFP22A54-POD
ISBN (Print-On-Demand): 978-1-6654-5192-5
ISBN (Online): 978-1-6654-5191-8

**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

# 2022 IEEE/ACM Workshop on Workflows in Support of Large-Scale Science (WORKS)

# WORKS 2022

## Table of Contents

**Message from the Workshop Chairs** ..... **vi**
**Workshop Organization** ..... **vii**

### Session 1: Provenance

Automatic, Efficient and Scalable Provenance Registration for FAIR HPC Workflows ..... 1
*Raül Sirvent (Barcelona Supercomputing Center), Javier Conejero (Barcelona Supercomputing Center), Francesc Lordan (Barcelona Supercomputing Center), Jorge Ejarque (Barcelona Supercomputing Center), Laura Rodríguez-Navas (Barcelona Supercomputing Center), José M. Fernández (Barcelona Supercomputing Center), Salvador Capella-Gutiérrez (Barcelona Supercomputing Center), and Rosa M. Badia (Barcelona Supercomputing Center)*

Challenges of Provenance in Scientific Workflow Management Systems ..... 10
*Khairul Alam (University of Saskatchewan, Canada) and Banani Roy (University of Saskatchewan, Canada)*

A Domain-Specific Composition Environment for Provenance Query of Scientific Workflows ..... 19
*Muhammad Mainul Hossain (University of Saskatchewan, Canada), Banani Roy (University of Saskatchewan, Canada), Chanchal Roy (University of Saskatchewan, Canada), and Kevin Schneider (University of Saskatchewan, Canada)*

### Session 2: Heterogeneous Executions, Anomaly Detection, and Scheduling

RADICAL-Pilot and Parsl: Executing Heterogeneous Workflows on HPC Platforms ..... 27
*Aymen Alsaadi (Rutgers, the State University of New Jersey, USA), Logan Ward (Data Science and Learning Division, Argonne National Laboratory, USA), Andre Merzky (Rutgers, the State University of New Jersey, USA), Kyle Chard (Data Science and Learning Division, Argonne National Laboratory, USA; University of Chicago, USA), Ian Foster (Data Science and Learning Division, Argonne National Laboratory, USA; University of Chicago, USA), Shantenu Jha (Rutgers, the State University of New Jersey, USA; Brookhaven National Laboratory, USA), and Matteo Turilli (Rutgers, the State University of New Jersey, USA; Brookhaven National Laboratory, USA)*

Workflow Anomaly Detection with Graph Neural Networks ..... 35
*Hongwei Jin (Argonne National Laboratory, United States), Krishnan Raghavan (Argonne National Laboratory, United States), George Papadimitriou (Information Sciences Institute, University of Southern California, United States), Cong Wang (RENCI, University of North Carolina, USA), Anirban Mandal (RENCI, University of North Carolina, USA), Patrycja Krawczuk (University of Southern California, United States), Loïc Pottier (University of Southern California, United States), Mariam Kiran (Energy Sciences Network (ESnet), United States), Ewa Deelman (University of Southern California, United States), and Prasanna Balaprakash (Argonne National Laboratory, United States)*

Co-Scheduling Ensembles of In Situ Workflows ..... 43
*Tu Mai Anh Do (Information Sciences Institute, University of Southern California, United States of America), Loïc Pottier (Information Sciences Institute, University of Southern California, United States of America), Rafael Ferreira da Silva (Oak Ridge National Laboratory, United States of America), Frédéric Suter (Oak Ridge National Laboratory, United States of America), Silvina Caíno-Lores (University of Tennessee at Knoxville, United States of America), Michela Taufer (University of Tennessee at Knoxville, United States of America), and Ewa Deelman (Information Sciences Institute, University of Southern California, United States of America)*

Events as a Basis for Workflow Scheduling ..... 52
*David Marchant (Copenhagen University, Denmark)*

## Session 3: Applications

An Automated Cryo-EM Computational Environment on the HPC System using Pegasus WMS ..... 60
*Tomasz Osinski (University of Southern California, USA), Mats Rynge (University of Southern California, USA), James K. Hong (University of Southern California, USA), Karan Vahi (University of Southern California, USA), Ruilin Chu (University of Southern California, USA), Cesar Sul (University of Southern California, USA), Ewa Deelman (University of Southern California, USA), and Byoung-Do Kim (University of Southern California, USA)*

Cross-Facility Workflows: Case Studies with Active Experiments ..... 68
*Nicholas Tyler (Lawrence Berkeley National Laboratory, USA), Robert Knop (Lawrence Berkeley National Laboratory, USA), Deborah Bard (Lawrence Berkeley National Laboratory, USA), and Peter Nugent (Lawrence Berkeley National Laboratory, USA)*

CardioHPC: Serverless Approaches for Real-Time Heart Monitoring of Thousands of Patients ..... 76
*Marjan Gusev (Ss. Cyril and Methodius University in Skopje, North Macedonia), Sashko Ristov (University of Innsbruck, Austria), Andrei Amza (University of Klagenfurt, Austria), Armin Hohenegger (University of Klagenfurt, Austria), Radu Prodan (University of Klagenfurt, Austria), Dimitar Mileski (Innovation Dooel, North Macedonia), Pano Gushev (Innovation Dooel, North Macedonia), and Goran Temelkov (Innovation Dooel, North Macedonia)*

## Session 4: Lightning Talks

Novel Proposals for FAIR, Automated, Recommendable, and Robust Workflows ..... 84
*Ishan Abhinit (Indiana University, USA), Emily K. Adams (Indiana University, USA), Khairul Alam (University of Saskatchewan, Canada), Brian Chase (Indiana University, USA), Ewa Deelman (University of Southern California, USA), Lev Gorenstein (Purdue University, USA), Stephen Hudson (Argonne National Laboratory, USA), Tanzima Islam (Texas State University, USA), Jeffrey Larson (Argonne National Laboratory, USA), Geoffrey Lentner (Purdue University, USA), Anirban Mandal (Renaissance Computing Institute, USA), John-Luke Navarro (Argonne National Laboratory, USA), Bogdan Nicolae (Argonne National Laboratory, USA), Line Pouchard (Brookhaven National Laboratory, USA), Rob Ross (Argonne National Laboratory, USA), Banani Roy (University of Saskatchewan, Canada), Mats Rynge (University of Southern California, USA), Alexander Serebrenik (Eindhoven University of Technology, Netherlands), Karan Vah (University of Southern California, USA), Stefan Wild (Argonne National Laboratory, USA), Yufeng Xin (Renaissance Computing Institute, USA), Rafael Ferreira da Silva (Oak Ridge National Laboratory, USA), and Rosa Filgueira (University of St Andrews, UK)*

**Author Index** ..... **93**