

2023 IEEE 19th International Conference on e-Science (e-Science 2023)

**Limassol, Cyprus
9-13 October 2023**



**IEEE Catalog Number: CFP2306A-POD
ISBN: 979-8-3503-2224-8**

**Copyright © 2023 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:	CFP2306A-POD
ISBN (Print-On-Demand):	979-8-3503-2224-8
ISBN (Online):	979-8-3503-2223-1
ISSN:	2325-372X

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

2023 IEEE 19th International Conference on e-Science (e-Science)

eScience 2023

Table of Contents

Message from the IEEE eScience 2023 Conference leadership
Organising Committee
Technical Programme Committee
Keynotes

The complete list of papers and presentations follows. The list is organized by full papers, invited talks and poster extended abstracts, all in alphabetical order by the primary author's last name. Next are papers submitted to the workshops: INSTIL, ReWorDS and AI4Health, in the same alphabetical order.

Full Papers

I/O Burst Prediction for HPC Clusters using Darshan Logs.....	1
<i>Engin Arslan, Ehsan Saeedizade and Roya Taheri</i>	
Demystifying the Performance of Data Transfers in High-Performance Research Networks.....	11
<i>Engin Arslan, Ehsan Saeedizade and Bing Zhang</i>	
Online and Scalable Data Compression Pipeline with Guarantees on Quantities of Interest.....	22
<i>Tania Banerjee, Jaemoon Lee, Jong Choi, Qian Gong, Jieyang Chen, Choongseok Chang, Scott Klasky, Anand Rangarajan and Sanjay Ranka</i>	
Search in Archival Facsimile Documents for Digital History.....	32
<i>David Banyasz, Sebastian Hofstätter and Allan Hanbury</i>	
Asynchronous Decentralized Bayesian Optimization for Large Scale Hyperparameter Optimization.....	42
<i>Romain Egele, Isabelle Guyon, Venkatram Vishwanath and Prasanna Balaprakash</i>	
LivePublication: The Science Workflow Creates and Updates the Publication.....	52
<i>Augustus Ellerm, Mark Gahegan and Benjamin Adams</i>	

Quantum Computer Simulations at Warp Speed: Assessing the Impact of GPU Acceleration.....	62
<i>Jennifer Faj, Ivy Peng, Jacob Wahlgren and Stefano Markidis</i>	
Spatiotemporally Adaptive Compression for Scientific Dataset with Feature Preservation – a Case Study on Simulation Data with Extreme Climate Events Analysis.....	72
<i>Qian Gong, Chengzhu Zhang, Xin Liang, Viktor Reshniak, Jieyang Chen, Anand Rangarajan, Sanjay Ranka, Lipeng Wan, Nicolas Vidal, Paul Ullrich, Norbert Podhorszki, Robert Jacob and Scott Klasky</i>	
FlyPaw: Optimized Route Planning for Scientific UAV Missions.....	82
<i>Andrew Grote, Eric Lyons, Komal Thareja, Georgios Papadimitriou, Ewa Deelman, Anirban Mandal, Prasad Calyam and Michael Zink</i>	
PSI/J: A Portable Interface for Submitting Monitoring, and Managing Jobs.....	92
<i>Mihael Hategan-Marandiuc, Andre Merzky, Nicholson Collier, Ketan Maheshwari, Jonathan Ozik, Matteo Turilli, Andreas Wilke, Justin M. Wozniak, Kyle Chard, Ian Foster, Rafael Ferreira da Silva, Shantenu Jha and Daniel Laney</i>	
Reversible Network Covert Channel by Payload Modulation in Streams of Decimal Sensor Values.....	102
<i>Carina Heßeling, Jörg Kelle and Sebastian Litzinger</i>	
Entangled, a Bidirectional System for Sustainable Literate Programming.....	110
<i>Johan Hidding</i>	
Truck Speed Detection through Video Streams.....	119
<i>Zuo Huang, Richard O. Sinnott and Krista A. Ehinger</i>	
Achieving Decentralized Authority for Collaborative Data Sharing with Consensus.....	129
<i>Kamil Jarosz, Dawid Dębowski, Radosław Szuma, Łukasz Opióła, Łukasz Dutka, Renata G. Słota and Jacek Kitowski</i>	
In-Situ Techniques on GPU-Accelerated Data-Intensive Applications.....	139
<i>Yi Ju, Mingshua Li, Adalberto Perez, Laura Bellentani, Niclas Jansson, Stefano Markidis, Philipp Schlatter and Erwin Laure</i>	
Lazy Python Dependency Management in Large-scale systems.....	149
<i>Alok Kamatar, Mansi Sakarvadia, Valerie Hayot-Sasson, Kyle Chard and Ian Foster</i>	
OCSRL: An Model-Based Reinforcement Learning Approach to Optimize Energy Consumption of Cooling Systems.....	159
<i>Keshav Kaushik and Vinayak Naik</i>	
Reconstruct Geospatial Data from Ultra Sparse Inputs to Predict Climate Events...	169
<i>Marco Landt-Hayen, Peer Kröger, Willi Rath and Martin Claus</i>	
A Dense Retrieval System and Evaluation Dataset for Scientific Computational Notebooks.....	179
<i>Na Li, Yangjun Zhang and Zhiming Zhao</i>	

Mapping the Repository Landscape: Harnessing Similarity with RepoSim and RepoSnipy.....	189
<i>Zihao Li and Rosa Filgueira</i>	
FAIRSECO: An Extensible Framework for Impact Measurement of Research Software.....	199
<i>Deekshitha M, Siamak Farshidi, Jason Maassen, Rena Bakhshi, Rob van Nieuwpoort and Slinger Jansen</i>	
Detection and Tracking of Dynamic Ocean Carbon Uptake Regimes Built upon Spatial Target-Driver Relationships via Adaptive Hierarchical Clustering.....	209
<i>Sweety Mohanty, Daniyal Kazempour, Lavinia Patara and Peer Kröger</i>	
A Method for Constructing Research Data Provenance in High-Performance Computing Systems.....	219
<i>Yuta Namiki, Takeo Hosomi, Hideyuki Tanushi, Akihiro Yamashita and Susumu Date</i>	
TrueTrees: A Scalable Workflow for the Integration of Airborne LiDAR Scanning Data into Fuel Models for Prescribed Fire Simulations.....	229
<i>Daniel Roten, Lucas Wells, Daniel Crawl, Russell Parsons, Anthony Marcozzi, Rodman Linn, Kevin Hiers and Ilkay Altintas</i>	
Building Lightweight Semantic Search Engines.....	239
<i>Michael Rovatsos and Rosa Filgueira</i>	
Online Boosted Gaussian Learners for in-situ Detection and Characterization of Protein Folding States in Molecular Dynamics Simulations.....	249
<i>Harshita Sahni, Hector Carrillo-Cabada, Ekaterina Kots, Silvina Caino-Lores, Jack Marquez, Ewa Deelman, Michel Cuendet, Harel Weinstein, Michela Taufer and Trilce Estrada</i>	
Virtual Patient Platform and Data Space for Sharing, Learning, Discussing, and Researching.....	259
<i>André Santanchè, Heitor Mattosinho, Marcos Mota, Fagner Pantoja, Gabriel Leite, Ana Claudia Tonelli, Fernando Valente, Juliana Martins, Sandro Queirós, Tiago Grangeia and Marco Antonio Carvalho Filho</i>	
Database Evolution, by Scientists, for Scientists: A Case Study.....	269
<i>Robert Schuler, Jitin Singla, Brinda Vallat, Kate L. White, Helen M. Berman and Carl Kesselman</i>	
Real-Time Vehicle Passenger Detection through Deep Learning.....	279
<i>Richard Sinnott and Yuhan Gu</i>	
Can Automated Metadata Extraction make Scientific Data More Navigable?	289
<i>Tyler Skluzacek, Kyle Chard and Ian Foster</i>	
Towards Lightweight Data Integration using Multi-workflow Provenance and Data Observability.....	299
<i>Renan Souza, Tyler J. Skluzacek, Sean R. Wilkinson, Maxim Ziatdinov and Rafael Ferreira da Silva</i>	

Reproducible Query Processing and Data Citation of in situ Soil Moisture Data.....	309
<i>Moritz Staudinger, Tobias Hajszan, Tomasz Miksa, Irene Himmelbauer, Daniel Aberer, Andreas Rauber and Wouter Dorigo</i>	
Driving Next-Generation Workflows from the Data Plane.....	319
<i>Frederic Suter, Rafael Ferreira da Silva, Ana Gainaru and Scott Klasky</i>	
Tomo2Mesh: Fast Porosity Mapping and Visualization for Synchrotron Tomography.....	329
<i>Aniket Tekawade, Viktor Nikitin, Yashas Satapathy, Zhengchun Liu, Xuan Zhang, Peter Kenesei, Weijian Zheng, Francesco De Carlo, Ian Foster and Rajkumar Kettimuthu</i>	
A Network Synthesis and Analytics Pipeline with Applications to Sustainable Energy in Smart Grid.....	339
<i>Swapna Thorve, Aparna Kishore, Dustin Machi, S. S. Ravi and Madhav Marathe</i>	
Contract-Driven Design of Scientific Data Analysis Workflows.....	349
<i>Anh Duc Vu, Jan Arne Sparka, Ninon de Mecquenem, Timo Kehrer, Ulf Leser and Lars Grunske</i>	
RepoGraph: A Novel Semantic Code Exploration Tool for Python Repositories based on Knowledge Graphs and Deep Learning.....	359
<i>Christopher Williams and Rosa Filgueira</i>	
CKN: An Edge AI Distributed Framework.....	369
<i>Sachith Withana and Beth Plale</i>	
An Automation Framework for Comparison of Cancer Response Models Across Configurations.....	379
<i>Justin Wozniak, Rajeev Jain, Andreas Wilke, Rylie Weaver, Alexander Partin, Thomas Brettin and Rick Stevens</i>	
frances: Cloud-Based Historical Text Analysis with Deep Learning and Parallel Processing.....	389
<i>Lilin Yu, Ash Charlton, Wilfrid Askins, Melissa Terras and Rosa Filgueira</i>	

Invited Talks

Why we need a Research Data Reference Architecture.....	399
<i>David Abramson, Luc Betbeder-Matibet, Stephen Bird, Jake Carroll, Rhys Francis, Wojtek Goscinski, Ai-Lin Soo, Garry Swan, Carmel Walsh, Glenn Wightwick and Max Wilkinson</i>	
From Workflows to Teamflows in eScience: Integrating Collaborative Science, AI, and Computing for Impact at Scale.....	403
<i>Ilkay Altintas</i>	

e-Science: The Long View.....	405
<i>David De Roure and Pip Willcox</i>	
How Is Artificial Intelligence Changing Science?	409
<i>Ewa Deelman</i>	
Research Software Engineering in 2030.....	413
<i>Daniel S. Katz and Simon Hettrick</i>	
Discovery Testbed: an Observational Instrument for Broadband Research.....	415
<i>Kate Keahey, Nick Feamster, Guilherme Martins, Mark Powers, Marc Richardson, Alexis Schrubbe and Michael Sherman</i>	
Let's Put the Science in eScience.....	419
<i>Carl Kesselman, Robert Schuler and Ian Foster</i>	
APPFLx: Providing Privacy-Preserving Cross-Silo Federated Learning as a Service.....	422
<i>Ravi Madduri, Zilinghan Li, Shilan He, Eliu Huerta, Ryan Chard, Kibaek Kim, Minseok Ryu, Volodymyr Kindratenko, Trung-Hieu Hoang, Pranshu Chaturvedi, Maryellen Giger and Jordan Fuhrman</i>	
Reproducible eScience: The Data Containerization Challenge.....	426
<i>Tanu Malik</i>	
Beyond the Fourth Paradigm – The Rise of AI.....	431
<i>Andreas Marek, Markus Rampp, Klaus Reuter and Erwin Laure</i>	
Toward Democratizing Access to Science Data: Introducing the National Data Platform.....	435
<i>Manish Parashar and Ilkay Altintas</i>	
Democratization of AI: Challenges of AI Cyberinfrastructure and Software Research.....	439
<i>Beth Plale, Sadia Khan and Alfonso Morales</i>	
eScience Serverless Data Storage Services in the Edge-fog-cloud Continuum.....	442
<i>Dante Sánchez-Gallegos, Diana Carrizales-Espinosa, José Luis Gonzalez-Compeán and Jesus Carretero</i>	
From eScience to Impact on the Economy and Society.....	446
<i>Paul Watson and Barry Hodgson</i>	
Tracking Material Reuse across Construction Supply Chains.....	449
<i>Stanly Wilson, Kwabena Adu-Duodu, Yinhao Li, Ringo Sham, Yingli Wang, Ellis Solaiman, Charith Perera, Rajiv Ranjan and Omer Rana</i>	

Poster Extended Abstracts

Normality of I-V Measurements Using ML	453
<i>Anees Al-Najjar, Nageswara Rao, Craig Bridges and Sheng Dai</i>	
Enhancing AIS Vessel Trajectories via Trip Detection.....	455
<i>Mirjam Bayer, Sören Dethlefsen, Tabea Fry and Peer Kröger</i>	
LexiFuse+: A Unified One-Class Solution for Imbalanced Short-text Classification.....	457
<i>Saugata Bose and Guoxin Su</i>	
Symbolic Regression Applied to Cosmology: An Approximate Expression for the Density Perturbation Variance.....	459
<i>Ana Carvalho, David Magalhaes Oliveira, Alberto Krone-Martins and Antonio da Silva</i>	
Scalable Infrastructure for Galaxy Image Analysis: I. Measuring Position Angles with Radon transforms.....	461
<i>Neo Chen and Alberto Krone-Martins</i>	
Project Presentation: Recursive Functions and Engineering Probabilistic Ontologies.....	463
<i>Irakli Chिताia, Roland Omanadze and Mikheil Rukhaia</i>	
Towards a Knowledge Graph Enhanced Automation and Collaboration Framework for Digital Twins.....	465
<i>Vasileios Christou, Yuandou Wang and Zhiming Zhao</i>	
Emulating Hydrodynamics from Dark Matter 3D Density Fields.....	467
<i>Miguel Conceição, Alberto Krone-Martins and Antonio da Silva</i>	
Example Evaluations of Plagiarism Cases Using FAIR Metrics and the PDP-DREAM Ontology.....	469
<i>Adam Craig, Anousha Athreya and Carl Taswell</i>	
Visualization and Labeling of Terrestrial LiDAR Data for Three-Dimensional Fuel Classification.....	471
<i>Ivannia Gomez Moreno, Isaac Nealey, Daniel Roten, Mai Nguyen, Daniel Crawl, Kate O'Laughlin, Melissa Floca, Scott Pokswinski and Ilkay Altintas</i>	
Detection of Component Degradation: A Study on Autoencoder-based Approaches.....	473
<i>Dario Guidotti, Laura Pandolfo and Luca Pulina</i>	
Verifying Neural Networks with SMT: An Experimental Evaluation.....	475
<i>Dario Guidotti, Laura Pandolfo and Luca Pulina</i>	
Mobilise-D: Experiences of Processing Large Medical Data Sets Using Cloud Computing resources.....	477
<i>Hugo Hiden and Paul Watson</i>	

CWL-FLOps: A Novel Method for Federated Learning Operations at Scale.....	479
<i>Chronis Kontomaris, Yuandou Wang and Zhiming Zhao</i>	
Integrating R in a Distributed Scientific Workflow via a Jupyter-based Environment.....	481
<i>Mariantonietta La Marra, Diederik Blanson Henkemans, Jessica Titocci, Spiros Koulouzis, Ilaria Rosati and Zhiming Zhao</i>	
A Digital Twin for Climate Extremes using Artificial Intelligence.....	483
<i>Christian Pagé and Anne Durif</i>	
Using Fourier Coefficients and Wasserstein Distances to Estimate Entropy in Time Series.....	485
<i>Scott Perkey, Ana Carvalho and Alberto Krone-Martins</i>	
Implementing a FAIR Information System for Archaeology-Related Interdisciplinary Research.....	487
<i>Steffen Strohm, Hartwig Buening and Matthias Renz</i>	
Accelerating the Design of Scientific Workflows with Simulation-Based Rapid Prototyping.....	489
<i>Frederic Suter</i>	
Semi-Supervising an Anomalous Universe.....	491
<i>David Sweeney, Alberto Krone-Martins, Peter Tuthill and Richard Scalzo</i>	
Automatic Green Land-use Generator for Urban Areas.....	493
<i>Veronica Vidal, Carlos Carrillo, Ana Cortés, Alba Badia and Gara Villalba</i>	
Design by Contract Revisited in the Context of Scientific Data Analysis Workflows.....	495
<i>Anh Duc Vu, Jan Arne Sparka, Ninon de Mecquenem, Timo Kehrer, Ulf Leser and Lars Grunske</i>	
Automating the Evaluation of Datasets for FAIR and CARE Principles.....	497
<i>Rujula Yete, Shweta Purawat, Ismael Perez, Daniel Crawl and Ilkay Altintas</i>	

Workshop: INSTIL 2023

INSTIL 2023 Workshop Foreword and Organisation.....	499
Enhancing Learning about Climate Change Issues among Secondary School Students with Citizen Science tools.....	501
<i>Muhammad Adnan, Luk Knapen, Wim Ectors and Lien Aerts</i>	
A Progressive Web App Template for Citizen Science Projects Involving Spatial Data Collection.....	509
<i>Anudari Batsaikhan and Stephan Hachinger</i>	

Citizen Science for the Sea with Information Technologies: an Open Platform for Gathering Marine Data and Marine Litter Detection from Leisure Boat Instruments.....	515
<i>Ciro Giuseppe De Vita, Gennaro Mellone, Dante Sanchez-Gallegos, Giuseppe Coviello, Diego Romano, Marco Lapegna and Angelo Ciaramella</i>	
EbAcraft: Engaging Local Communities in Learning about Ecosystem-based Adaptation for Coastal Cities in Europe.....	524
<i>Ítalo de Sena, Chiara Cocco, Vojtěch Brůža, Pierre Lenicolais, Saul Crowley and Francesco Pilla</i>	
Informing on Climate-friendly Novel Mobility Measures: Development of the Serious Game "MiniLautern" and Analysis of Player Feedback.....	533
<i>Frank Elberzhager, Patrick Mennig and Yanick Behrendt-Henn</i>	
Pollution Runner: a Serious Game to Promote Awareness towards Air Pollution.....	541
<i>Mattia Fortunati and Antonella Galizia</i>	

Workshop: ReWorDS 2023

ReWorDS 2023 Workshop Foreword and Organisation.....	549
From Program Chains to Exploratory Workflows: PopinSnake for Genomic Insertion Detection.....	551
<i>Kedi Cao, Nourhan Elfaramawy, Matthias Weidlich and Birte Kehr</i>	
Preserving File Provenance using Principles of Blockchain to Ensure Scientific Reproducibility.....	558
<i>Rizbanul Hasan, Shweta Purawat, Catherine Olschanowsky and Ilkay Altintas</i>	
Three Pillars of Practical Reproducibility.....	565
<i>Kate Keahey, Jason Anderson, Mark Powers and Adam Cooper</i>	
Building the I (Interoperability) of FAIR for Performance Reproducibility of Large-scale Composable Workflows in RECUP.....	571
<i>Bogdan Nicolae, Tanzima Islam, Robert Ross, Hub Van Dam, Kevin Assogba, Polina Shpilker, Mikhail Titov, Matteo Turilli, Tianle Wang, Ozgur Kilic, Shantenu Jha and Line Pouchard</i>	
Templated Hybrid Reusable Computational Analytics Workflow Management with Cloudmesh, Applied to the Deep Learning MLCommons Cloudmask Application.....	578
<i>Gregor von Laszewski, Jacques P. Fleischer, Geoffrey C. Fox, Juri Papay, Sam Jackson and Jeyan Thiyagalingam</i>	

Workshop: AI4Health 2023

AI4Health 2023 Workshop Foreword and Organisation.....	584
Explainable AI-based Clinical Decision Support System for Obesity Comorbidity Analysis.....	586
<i>Grazia Veronica Aiosa, Maurizio Palesi, Francesca Sapuppo and Maria Gabriella Xibilia</i>	
AI-enabled Solutions, Explainability and Ethical Concerns for Predicting Sepsis in ICUs: a Systematic Review.....	592
<i>Christina-Athanasia Alexandropoulou, Ilias Panagiotopoulos, Styliani Kleanthous, George Dimitrakopoulos, Ioannis Constantinou, Eleni Politi, Dimitrios Ntalaperas, Xanthi Papageorgiou, Charithea Stylianides, Nikos Ioannides, Lakis Palazis, Constantinos Pattichis and Andreas Panayides</i>	
Federated Learning and Genetic Mutation for Multi-Resident Activity Recognition.....	601
<i>Cezar Anicai and Muhammad Zeeshan Shakir</i>	
Automatic Machine Learning applied to Electrical Biosignals: A Selective Review.....	607
<i>João Castelo Branco, Bruno Marques, Lívia Cruz, Regis Magalhães and José Macêdo</i>	
The Tele-Rehabilitaion as a Service (TRaaS) Project: Rationale, Study Design, and Methodology.....	613
<i>Antonio Celesti, Giovanna Sannino, Mario Bochicchio, Maria Fazio, Massimo Villari, Fabrizio Celesti, Mirjam Bonanno and Rocco Salvatore Calabrò</i>	
Better Together: Combining Different Handwriting Input Sources Improves Dementia Screening.....	620
<i>Nina Hosseini Kivanani, Elena Salobrar-García, Lorena Elvira-Hurtado, Inés López- Cuenca, Rosa de Hoz, José M. Ramírez, Pedro Gil, Mario Salas, Christoph Schommer and Luis A. Leiva</i>	
Analyzing distinct Neural Network Models for Oxygen Saturation Prediction towards a Personalized COPD Management.....	627
<i>Pascual Saldana, Xavier Masip Bruin, Albert Alonso and Isabel Blanco</i>	
Client Recruitment for Federated Learning in ICU Length of Stay Prediction.....	635
<i>Vincent Scheltjens, Lyse Naomi Wamba Momo, Wouter Verbeke and Bart De Moor</i>	
A Novel Deep Learning Approach for Colon and Lung Cancer Classification Using Histopathological Images.....	644
<i>Naeem Ullah, Ivanoe De Falco and Giovanna Sannino</i>	
Early Prediction of Sepsis using Time Series Forecasting.....	654
<i>Jinghua Xu, Natalia Minakova, Pablo Ortega Sanchez and Stefan Riezler</i>	