

2021 IEEE 17th International Conference on eScience (eScience 2021)

**Virtual Conference
20 – 23 September 2021**



**IEEE Catalog Number: CFP2106A-POD
ISBN: 978-1-6654-4708-9**

**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:	CFP2106A-POD
ISBN (Print-On-Demand):	978-1-6654-4708-9
ISBN (Online):	978-1-6654-0361-0

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 17th International Conference on eScience (eScience) **e-Science 2021**

Table of Contents

Message from the eScience 2021 Chairs	xii
Program Committee	xiv
Steering Committee	xvi
Message from the ReWorDS 2021 Organizers	xvii
Keynotes	xix

Regular Papers

Using Nanopublications to Detect and Explain Contradictory Research Claims	1
<i>Imran Asif (Heriot-Watt University, United Kingdom), Ilaria Tiddi (Vrije Universiteit Amsterdam, The Netherlands), and Alasdair J. G. Gray (Heriot-Watt University, United Kingdom)</i>	
Ensemble Labeling Towards Scientific Information Extraction (ELSIE)—Blob Extraction	11
<i>Erin Murphy (DePaul University, USA), Alexander Rasin (DePaul University, USA), Jacob Furst (DePaul University, USA), Daniela Raicu (DePaul University, USA), and Roselyne Tchoua (DePaul University, USA)</i>	
Extending Defoe for the Efficient Analysis of Historical Texts at Scale	21
<i>Rosa Filgueira (Heriot-Watt University, UK), Claire Grover (University of Edinburgh, UK), Vasiliios Karaiskos (University of Edinburgh, UK), Beatrice Alex (University of Edinburgh, UK), Sarah van Eyndhoven (University of Edinburgh, UK), Lisa Gotthard (University of Edinburgh, UK), and Melissa Terras (University of Edinburgh, UK)</i>	
Context-Aware Execution Migration Tool for Data Science Jupyter Notebooks on Hybrid Clouds	30
<i>Renato L. F. Cunha (IBM Research), Lucas C. Villa Real (IBM Research), Renan Souza (IBM Research), Bruno Silva (IBM Research), and Marco A. S. Netto (IBM Research)</i>	
Serverless Containers - Rising Viable Approach to Scientific Workflows	40
<i>Krzysztof Burkat (AGH University of Science and Technology, Poland), Maciej Pawlik (AGH University of Science and Technology, Poland), Bartosz Balis (AGH University of Science and Technology, Poland), Maciej Malawski (AGH University of Science and Technology, Poland), Karan Vahi (University of Southern California, USA), Mats Rynge (University of Southern California, USA), Rafael Ferreira da Silva (University of Southern California, USA), and Ewa Deelman (University of Southern California, USA)</i>	

Traffic, Earthquakes and Evacuations: A Data Driven Multi-disciplinary Simulation Framework .50.....	
	<i>Mohsen Sichani (Victoria University of Wellington, New Zealand), Kris Bubendorfer (Victoria University of Wellington, New Zealand), and Richard Arnold (Victoria University of Wellington, New Zealand)</i>
Compound Segmentation via Clustering on Mol2Vec-Based Embeddings .60.....	
	<i>Daniyal Kazempour (Ludwig-Maximilians-University Munich, Germany), Anna Beer (Ludwig-Maximilians-University Munich, Germany), Melanie Oelker (Umeå University, Germany), Peer Kröger (Christian-Albrechts-University Kiel, Germany), and Thomas Seidl (Ludwig-Maximilians-University Munich, Germany)</i>
Analyzing Big Environmental Audio with Frequency Preserving Autoencoders .70.....	
	<i>Benjamin Rowe (Queensland University of Technology, Australia), Philip Eichinski (Queensland University of Technology, Australia), Jinglan Zhang (Queensland University of Technology, Australia), and Paul Roe (Queensland University of Technology, Australia)</i>
EPI Framework: Approach for Traffic Redirection Through Containerised Network Functions .80....	
	<i>Jamila Alsayed Kassem (University of Amsterdam), Onno Valkering (University of Amsterdam), Adam Belloum (University of Amsterdam), and Paola Grosso (University of Amsterdam)</i>
Porting Real-World Applications to GPU Clusters: A Celerity and CRONOS Case Study .90.....	
	<i>Philipp Gschwandtner (University of Innsbruck, Austria), Ralf Kissmann (University of Innsbruck, Austria), David Huber (University of Innsbruck, Austria), Philip Salzmann (University of Innsbruck, Austria), Fabian Knorr (University of Innsbruck, Austria), Peter Thoman (University of Innsbruck, Austria), and Thomas Fahringer (University of Innsbruck, Austria)</i>
Spatiotemporal Pattern Mining for Nowcasting Extreme Earthquakes in Southern California .99.....	
	<i>Bo Feng (Indiana University, USA) and Geoffrey C. Fox (Indiana University, USA)</i>
Distributed Spatiotemporal Distance Join for Trajectory Data .108.....	
	<i>Guang Yang (Imperial College London, United Kingdom) and Thomas Heinis (Imperial College London, United Kingdom)</i>
Where to Encode: A Performance Analysis of x86 and Arm-Based Amazon EC2 Instances .118.....	
	<i>Roland Mathá (University of Klagenfurt, Austria), Dragi Kimovski (University of Klagenfurt, Austria), Anatolij Zabrovskiy (University of Klagenfurt, Austria), Christian Timmerer (University of Klagenfurt, Austria), and Radu Prodan (University of Klagenfurt, Austria)</i>
Exploring Learning Approaches for Ancient Greek Character Recognition with Citizen Science Data .128.....	
	<i>Matthew I. Swindall (Middle Tennessee State University, USA), Gregory Croisdale (University of Tennessee, USA), Chase C. Hunter (University of Tennessee, USA), Ben Keener (University of Kentucky, USA), Alex C. Williams (University of Tennessee, USA), James H. Brusuelas (University of Kentucky, USA), Nita Krevans (University of Minnesota, USA), Melissa Sellew (University of Minnesota, USA), Lucy Fortson (University of Minnesota, USA), and John F. Wallin (Middle Tennessee State University, USA)</i>

Edge-to-Cloud Virtualized Cyberinfrastructure for Near Real-Time Water Quality Forecasting in Lakes and Reservoirs .138.....
Vahid Daneshmand (University of Florida, USA), Adrienne Breef-Pilz (Virginia Tech, USA), Cayelan C. Carey (Virginia Tech, USA), Yuqi Jin (University of Florida, USA), Yun-Jung Ku (University of Florida, USA), Kensworth C. Subratie (University of Florida, USA), R. Quinn Thomas (Virginia Tech, USA), and Renato J. Figueiredo (University of Florida, USA)

Leveraging Machine Learning to Detect Data Curation Activities .149.....
Sara Lafia (University of Michigan, USA), Andrea Thomer (University of Michigan, USA), David Bleckley (University of Michigan, USA), Dharma Akmon (University of Michigan, USA), and Libby Hemphill (University of Michigan, USA)

WfChef: Automated Generation of Accurate Scientific Workflow Generators .159.....
Tainã Coleman (University of Southern California, USA), Henri Casanova (University of Hawaii, USA), and Rafael Ferreira da Silva (University of Southern California, USA)

Historic Twitter Mining: A Case Study Aiming to Identify and Capture the Social Media Network Activities of People who Died by Suicide .169.....
Richard O. Sinnott (University of Melbourne, Australia), Philip Law (Monash University, Australia), Jane Pirkis (University of Melbourne, Australia), Tiffany Too (University of Melbourne, Australia), Allan Wu (University of Melbourne, Australia), and Sadia Waleem (University of Melbourne, Australia)

SAMPRA: Scalable Analysis, Management, Protection of Research Artifacts .177.....
Patrick G. Bridges (University of New Mexico), Zeinab Akhavan (University of New Mexico), Jonathan Wheeler (University of New Mexico), Hussein Al-Azzawi (University of New Mexico), Orlando Albillar (University of New Mexico), and Grace Faustino (University of New Mexico)

KnowMeme: A Knowledge-Enriched Graph Neural Network Solution to Offensive Meme Detection .. 186
Lanyu Shang (University of Illinois Urbana-Champaign, USA), Christina Youn (University of Notre Dame, USA), Yuheng Zha (University of Notre Dame, USA), Yang Zhang (University of Notre Dame, USA), and Dong Wang (University of Illinois Urbana-Champaign, USA)

Accelerating Complex Modeling Workflows in CyberWater Using on-Demand HPC/Cloud Resources 196
Feng Li (Indiana University-Purdue University Indianapolis, USA), Ranran Chen (Indiana University-Purdue University Indianapolis, USA), Yuankun Fu (Indiana University-Purdue University Indianapolis, USA), Fengguang Song (Indiana University-Purdue University Indianapolis, USA), Yao Liang (Indiana University-Purdue University Indianapolis, USA), Isuru Ranawaka (Indiana University Bloomington, USA), Sudhakar Pamidighantam (Indiana University Bloomington, USA), Daniel Luna (University of Pittsburgh, USA), and Xu Liang (University of Pittsburgh, USA)

- Extreme Scale Survey Simulation with Python Workflows .206.....
Antonio Villarreal (Argonne National Laboratory, USA), Yadu Babuji (University of Chicago, USA), Tom Uram (Argonne National Laboratory, USA), Daniel S. Katz (University of Illinois, USA), Kyle Chard (University of Chicago, USA), and Katrin Heitmann (Argonne National Laboratory, USA)
- An Empirical Study of Package Dependencies and Lifetimes in Binder Python Containers .215.....
Tim Shaffer (University of Notre Dame, USA), Kyle Chard (University of Chicago; Argonne National Laboratory, USA), and Douglas Thain (University of Notre Dame, USA)

Posters

- FPCA Emulation of Cosmological Simulations .225.....
Miguel Conceição (IA, Faculdade de Ciências da Universidade de Lisboa, Portugal), Alberto Krone-Martins (University of California, USA), and Antonio da Silva (IA, Faculdade de Ciências da Universidade de Lisboa, Portugal)
- Expanding IceCube GPU Computing into the Clouds .227.....
Igor Sfiligoi (University of California San Diego, USA), Shava Smallen (University of California San Diego, USA), Frank Würthwein (University of California San Diego, USA), Nicole Wolter (University of California San Diego, USA), David Schultz (University of Wisconsin - Madison, USA), and Benedikt Riedel (University of Wisconsin - Madison, USA)
- Enabling Microbiome Research on Personal Devices .229.....
Igor Sfiligoi (University of California San Diego, USA), Daniel McDonald (University of California San Diego, USA), and Rob Knight (University of California San Diego, USA)
- Extracting Stances on Pandemic Measures from Social Media Data .231.....
Erik Tjong Kim Sang (Netherlands eScience Center, The Netherlands), Shihan Wang (Utrecht University, The Netherlands), Marijn Schraagen (Utrecht University, The Netherlands), and Mehdi Dastani (Utrecht University, The Netherlands)
- Towards Making Fusion Data FAIR .233.....
David Coster (Max Planck Institute for Plasma Physics, Germany), Shaun de Witt (Culham Centre for Fusion Energy, Abingdon, United Kingdom), Iraklis Klampanos (NCSR "Demokritos", Greece), Marcin Plociennik (PSNC, Poland), Andreas Ikonomopoulos (NCSR "Demokritos", Greece), Frederic Imbeaux (CEA, France), Nathan Cummings (Culham Centre for Fusion Energy, United Kingdom), Michal Owsiak (PSNC, Poland), Agata Filipczak (PSNC, Poland), Bartosz Bosak (PSNC, Poland), Stasinou Konstantopoulos (NCSR "Demokritos", Greece), and Pär Strand (Chalmers University of Technology, Sweden)
- Modeling of Interruptions in Computer Network on Control of Non-Zero Minimum of Capacity .235
Marina Yashina (Moscow Automobile and Road Construction State Technical University (MADI), Russia), Alexander Tatashev (Moscow Automobile and Road Construction State Technical University (MADI), Russia), and Ivan Kuteynikov (Moscow Automobile and Road Construction State Technical University (MADI), Russia)

Ultrafast Focus Detection for Automated Microscopy .237.....	
	<i>Maksim Levental (University of Chicago), Ryan Chard (Argonne National Laboratory), Kyle Chard (University of Chicago), Ian Foster (University of Chicago), and Gregg A. Wildenberg (University of Chicago)</i>
HTCondor Data Movement at 100 Gbps .239.....	
	<i>Igor Sfligoi (University of California San Diego, USA), Frank Würthwein (University of California San Diego, USA), Thomas DeFanti (University of California San Diego, USA), and John Graham (University of California San Diego, USA)</i>
Neural Network Based Molecular Dynamics Simulations of Sputtering Processes .241.....	
	<i>Shokirbek Shermukhamedov (University of Innsbruck, Austria), Michael Probst (University of Innsbruck, Austria), and Lei Chen (ITER-Organization, France)</i>
Gold Panning: Automatic Extraction of Scientific Information from Publications .243.....	
	<i>Zhi Hong (University of Chicago), Kyle Chard (University of Chicago), and Ian Foster (University of Chicago)</i>
Software and Data Provenance as a Basis for eScience Workflow .245.....	
	<i>Joseph Conquest (University of Washington Bothell, USA) and Michael Stiber (University of Washington Bothell, USA)</i>
A Roadmap to Robust Science for High-Throughput Applications: The Scientists' Perspective .247..	
	<i>M. Taufer (University of Tennessee Knoxville), E. Deelman (University of Southern California, Los Angeles), R. Ferreira da Silva (University of Southern California, Los Angeles), T. Estrada (University of New Mexico), and M. Hall (University of Utah)</i>
A Case Study in Scientific Reproducibility from the Event Horizon Telescope (EHT) .249.....	
	<i>R. Ketron (University of Tennessee Knoxville), J. Leonard (University of Tennessee Knoxville), B. Roachell (University of Tennessee Knoxville), R. Patel (University of Tennessee Knoxville), R. White (University of Southern California, Los Angeles), S. Caino-Lores (University of Tennessee Knoxville), N. Tan (University of Tennessee Knoxville), P. Milesz (Syracuse University), K. Vahiy (University of Southern California, Los Angeles), E. Deelman (University of Southern California, Los Angeles), D. Brown (University of Southern California, Los Angeles), and M. Taufer (University of Tennessee Knoxville)</i>
Federated Function as a Service for eScience .251.....	
	<i>Yadu Babuji (University of Chicago), Josh Bryan (University of Chicago), Ryan Chard (Argonne National Laboratory), Kyle Chard (University of Chicago and Argonne National Laboratory), Ian Foster (University of Chicago and Argonne National Laboratory), Ben Galewsky (University of Illinois at Urbana-Champaign), Daniel S. Katz (University of Illinois at Urbana-Champaign), and Zhuozhao Li (University of Chicago)</i>
Cloud Clustering Over January 2003 via Scalable Rotationally Invariant Autoencoder .253.....	
	<i>Takuya Kurihana (University of Chicago), Elisabeth Moyer (University of Chicago), Rebecca Willett (University of Chicago), Davis Gilton (University of Wisconsin - Madison), and Ian Foster (University of Chicago)</i>

Optimizing Embedded Industrial Safety Systems Based on Time-of-Flight Depth Imaging .255.....	
	<i>Peter Thoman (University of Innsbruck, Austria), Alexander Hirsch (University of Innsbruck, Austria), Markus Wippler (University of Innsbruck, Austria), Thomas Fahringer (University of Innsbruck, Austria), and Robert Hranitzky (tofmotion GmbH, Austria)</i>
Towards System for Knowledge Representation of Campaign Experimentation .257.....	
	<i>Sachith Withana (Indiana University, USA), Kshitij Mehta (Oak Ridge National Laboratory, USA), Matthew Wolf (Oak Ridge National Laboratory, USA), and Beth Plale (Indiana University, USA)</i>
Predicting Flash Floods in the Dallas-Fort Worth Metroplex Using Workflows and Cloud Computing .259.....	
	<i>Eric Lyons (University of Massachusetts at Amherst, USA), Dong-Jun Seo (University of Texas Arlington, USA), Sunghee Kim (University of Texas Arlington, USA), Hamideh Habibi (Walter P. Moore and Associates Inc., USA), George Papadimitriou (University of Southern California, USA), Ryan Tanaka (University of Southern California, USA), Ewa Deelman (University of Southern California, USA), Michael Zink (University of Massachusetts at Amherst, USA), and Anirban Mandal (Renaissance Computing Institute, USA)</i>
Rapid Prototype for Shifting HPC to the Cloud .262.....	
	<i>Sam Weekly (Purdue University - Research Computing, USA), Zoey Mertes (Purdue University - Research Computing, USA), and Alex Younts (Purdue University - Research Computing, USA)</i>
CrisisFlow: Multimodal Representation Learning Workflow for Crisis Computing .264.....	
	<i>Patrycja Krawczuk (University of Southern California, USA), Shubham Nagarkar (University of Southern California, USA), and Ewa Deelman (University of Southern California, USA)</i>
Enhancing Automated FaaS with Cost-Aware Provisioning of Cloud Resources .267.....	
	<i>Matt Baughman (University of Chicago, USA), Ian Foster (University of Chicago and Argonne National Laboratory, USA), and Kyle Chard (University of Chicago and Argonne National Laboratory, USA)</i>
Towards Democratizing Modeling at Scale .269.....	
	<i>Yolanda Gil (University of Southern California, USA), Maximiliano Osorio (University of Southern California, USA), Varun Ratnakar (University of Southern California, USA), Suzanne Pierce (University of Texas at Austin, USA), Je'aime Powell (University of Texas at Austin, USA), Nathaniel Thorne (University of Texas at Austin, USA), and Peter Lubbs (University of Texas at Austin, USA)</i>

Workshop Papers

Incorporating Artificial Intelligence into the Workflow for Calibrating a Numerical Wavetank .271.....	
	<i>Charles Gillan (Queen's University Belfast, United Kingdom), Ciaran Finnegan (Queen's University Belfast, United Kingdom), and Pál Schmitt (Queen's University Belfast, United Kingdom)</i>

Brane: A Framework for Programmable Orchestration of Multi-site Applications .277.....
Onno Valkering (University of Amsterdam, The Netherlands), Reginald Cushing (University of Amsterdam, The Netherlands), and Adam Belloum (University of Amsterdam, The Netherlands)

Capturing and Semantically Describing Provenance to Tell the Story of R Scripts .283.....
Maria Luiza Mondelli (National Laboratory for Scientific Computing, Petrópolis - RJ, Brazil), Sheeba Samuel (Friedrich-Schiller University, Jena, Germany), Birgitta König-Ries (Friedrich-Schiller University, Jena, Germany), and Luiz M. R. Gadelha Jr. (Friedrich-Schiller University, Jena, Germany)

Author Index 289